2020 Annual Conference Proceedings

Decision Sciences in the Age of Connectivity

DECISION SCIENCES INSTITUTE

ISBN: 978-0-578-62648-2
ISSN: 2471-884X

51st Annual Conference

November 21st - 23rd 2020

Table of Contents

DECISION SCIENCES INSTITUTE

The impacts of Environmental Beliefs and Social Capital on Intention to Sign Environmental E-petition on Social Networks Sites

Quang-An Ha
University of Economics Ho Chi Minh City
Email: anhq@ueh.edu.vn

Thi Hong Thu Nguyen
University of Economics Ho Chi Minh City
Email: hthu@ueh.edu.vn

Phuong Nhi Nguyen Pham
University of Economics Ho Chi Minh City
Email: nhinpp@ueh.edu.vn

ABSTRACT

Recently, social network sites (SNS) have become popular in facilitating online petitions related to environmental issues. The social capital embedded on SNS and the users' environmental beliefs could be considered as the main factors that contribute to the success of an online environmental petition. Thus, this study examined the influence of these factors on the attitude toward online petition and the decision to participate in it. The results showed that the environmental beliefs, trust to SNS members, and norms highly influence the attitude toward online petition which leads to signing decisions. Implications of this study are also discussed in this paper.

KEYWORDS: Online petition, Environmental Beliefs, Social Capital, Social Interaction Ties, Trust, Norms

INTRODUCTION

Online petition (or e-petition) is the type of petition that held online to call for the attention of state organizations and to raise public awareness. Initiators of an e-petition, normally a person, groups, or non-profit organizations, perform the position to call for the support of everyone through online platforms (Ha and Chen 2017). Beginning with their original intent of formulating attitudes and viewpoints, these initiators aim to collect as many signatures as possible to make the effects of e-petition and to raise public awareness (Jalali, Ashouri et al. 2016). The signatures are collected through webpages such as Change.org, "We the People", Avaaz.org and GoPetition, which are popular sites for petition initiators to start their campaign. These sites themselves cannot spread the call for a petition, e-petition initiators need the spread the call for petition in other platforms, especially social networks sites. Indeed, e-petition nowadays is rapidly increased in terms of campaign number as well as the number of participants due to the convenience of SNS. SNS helps to transfer e-petition information among its users, thanks to the social capital embeded in the SNS under the form of social interaction ties, trust, and norms. However, the impact of social capital on the attitude toward e-petition and intention to sign e-petition has not been investigated. Besides the social capital embedded in SNS, environmental beliefs are one of the most important

antecedents that lead to user decision to sign an e-petition. Employing the social capital theory, this study attempts to examine the role of social capital factors namely social interaction ties, trust, and norm on the attitude toward e-petition which leads to intention to sign the online petition. Moreover, this study emphasizes the role of environmental beliefs on the attitude toward e-petition and gives more insights into the relationships between these beliefs and social capital factors.

LITERATURE REVIEW

2.1. Environmental activities on Social media and E-petition

As found by Ajzen (1991), attitude toward the behavior is considered as "the degree to which a person has a favorable or unfavorable evaluation of the behavior in question". In the literature, prior studies (Kotchen and Reiling 2000, Ramayah, Lee et al. 2010, Chen and Tung 2014) reported the positive attitude of a person towards a certain behavior will affect his or her intention to conduct that behavior. Especially the certain relationship between attitude and behavioral intention has been confirmed in the context of environmentally friendly products (Paul, Modi et al. 2016).

With the growing development of the utilization of online petition systems, some academic researches (Cruickshank, Edelmann et al. 2010, Ulo, Hidayanto et al. 2019) carried out to investigate how the attitude of internet users toward online petitions impacts their intention to sign the e-petition. The findings discovered that the more positive attitude persons have toward e-petition, the higher degree of intention to participate in e-petition (Ulo, Hidayanto et al. 2019).

Through the expansion of social media, there are many environmental activities organized on social media. E-petitions are held on social media because this platform facilitates in spreading the petition campaign message, communicate with the people concerned, and call for the support of online activism. The most important activities of online activism are to encourage and convince others to participate in the campaign (Milošević-Đorđević and Žeželj 2017). The effectiveness of online activism is based on the ability to appeal to everyone. Engaging as many people as possible is the main goal of these online activities. To engage more people to participate in environmental activities, it is important to establish specific attitudes of participants to the environment (McFarlane and Boxall 2003).

As discussed above, it is reasonable to believe the intention to sign an e-petition of a person will be affected by the attitude towards the e-petition. Hence, we offer the following hypothesis:

Hypothesis 1*: Attitude towards E-petition will have a positive impact on Intention to Sign the e-petition*

2.2. Environmental Beliefs and Attitude towards Signing E-petition

Environmental beliefs are human beliefs that reflect their worldview of natural surroundings (Scott and Willits 1994, Corral-Verdugo, Bechtel et al. 2003). It generally relates to environmental problems such as water scarcity, degradation of the ozone layer, climate change, and global warming (Sharma, Gadenne et al. 2017). When a person interacts with the environment, these beliefs become referential frameworks, determine behavior toward the environment (Corral-Verdugo, Bechtel et al. 2003). Dunlap and Van Liere (1978) developed the New Environment

Paradigm - NEP, “a set of beliefs about human-environment relations” to assess environmental beliefs (Corral-Verdugo, Bechtel et al. 2003). NEP was used in a large number of studies to confirm the impact of these beliefs on pro-environmental behavior (Corral-Verdugo and Armendariz 2000, Dunlap 2008). The stronger pro-environmental beliefs promote the more likely people engage in environmentally friendly buying behaviors (Pickett‐Baker and Ozaki 2008). This conclusion was also supported by several academic researchers (Tanner and Wölfing Kast 2003, Gadenne, Sharma et al. 2011).

According to the Theory of Reasoned Action (Fishbein and Ajzen 1975), the beliefs on a certain behavior are linked to the attitudes toward that behavior. Furthermore, Hartmann and Apaolaza-Ibáñez (2012) suggested consumers who raised environmental concerns would develop a more favorable attitude towards green energy, as well as show their willingness to pay a premium for eco-friendly energy brands.

Based on the literature review and explored in the e-petition context, we believe that there are substantial influences of environmental beliefs through attitudes on behavioral intentions towards online petition. Therefore, we propose the hypothesis:

Hypothesis 2*: Environmental Beliefs has a positive impact on Attitude towards E-petition*

2.3. Social Capital Theory and Social Networks Sites

Nahapiet and Ghoshal (1998) defined social capital as “the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit”. When people connect in communities, social capital exists and lies in their ways to connect and their relationships (Wasko and Faraj 2005, Okoli and Oh 2007). SNS are platforms to help individuals to connect, maintain interpersonal networks (Bernoff and Li 2008, Chen 2012) and it is one of the most important motivations for SNS usage (van Noort, Antheunis et al. 2012). Therefore, the relationship and the ways people connect on SNS also generate social capital and consequently affect and decide human behaviors.

According to Social Capital theory, there are three dimensions: structural dimension, relational dimension, and cognitive dimension. The structural dimension of social capital is referred to the interpersonal relationship between people in a network (Lin and Lu 2011). In SNS context, this dimension is represented by social interaction ties (Chen, Elakhdary et al. 2019), the main bridge connecting information resources among SNS members. The relational dimension refers to the quality and the nature of the relationships between people in a network. Trust and norm are considered as the key aspects of this dimension (Tsai and Ghoshal 1998). While trust can encourage the freedom of information exchange between members, norms can influence members to follow other members' activities, therefore increase the effectiveness of collaborations between members. It is the main reason we employed two dimensions of Social Capital theory as the main theory to explain the use of social media for environmental activities.

2.4. The impact of Environmental Beliefs on Social Capital

As found by Hamid (1989), the higher people believe in the control of social interaction, the more likely they contact comfortably with others. In the environmental context, when people believe in the scarcity of the environment, they tend to have pro-environmental behavior (Pickett-Baker and

Ozaki 2008), and the topic of environmental issues is the easy-to-talk topic that provokes communication with others. Noticeably, communication is a requirement for the creation and preservation of social capital (Kroon, Pierick et al. 2002). In respect of the Social Capital Theory, we believe that communication with other members in the SNS regarding the environmental issues will make the people close to others, which later affects the increasing interaction ties between SNS members. However, the link between social interaction and environmental beliefs, which is the possible driver of environmental communication, is still relatively unidentified. To address this gap, our study also examines how environmental beliefs are related to social interaction. Hence, we offer the following hypothesis:

Hypothesis 3*: Environmental Beliefs will have a positive impact on Social Interaction Ties*

Tsai and Ghoshal (1998) reported that cooperative behaviors, which consist of sharing or exchanging resources, will possibly occur when trust is developed. Furthermore, Bechtel (1999) suggested essential differences in types of trust. In particular, “knowledge-based trust” is a form of trust, in which one party bases on information, anticipate the other party’s behaviors, and believe the other has done well. Another kind of trust is “identification-based trust”, which exists when the parties successfully identify and understand each other's desires and intentions. Van der Kroon (2002) also evaluated that these types of trust rely on experiences and beliefs in the ethical integrity of other parties.

In terms of the earlier discussion derived from past studies, we predict that the trust of members is positively influenced by environmental beliefs. Hence, we make the subsequent prediction:

Hypothesis 4*: Environmental Beliefs will have a positive impact on Trust to members*

A norm occurs in a society, in which persons often perform in a certain way and may face punishments if they do not act in the same way (Axelrod, 1997). The norms can be created via socialization in childhood or provoked through life experiences (Van der Kroon 2002). In the research of energy-saving practice, Gadenne, Sharma et al. (2011) showed that the norm of environmental behavior could be positively affected by environmental beliefs. As individuals are aware of ecological problems and believe these matters possibly occur, then normative responses emerge, many people adopt pro-environmental behaviors and shape a norm. For example, persons believe climate change impacts every human life, and this belief drives them to engage in conventional and normative actions to protect the environment (Ozaki 2011). Furthermore, Yoon (2018) also found that environmental beliefs have a positive impact on an individual's norms for using Green IT which is “a future-oriented and pro-environmental information technology”. Therefore, we propose the hypothesis:

Hypothesis *5: Environmental Beliefs will have a positive impact on E-petition Norms*

2.5. Social Capital and Attitude towards E-petition

Friedkin and Johnsen (2011) pointed out the “social influence network theory” model which is applied to interpret and predict individuals’ attitude change in small group dynamics. Especially, the social network of interpersonal influences is involved in this mechanism. The persons may

evaluate and associate their own and others' positions through social interaction (Friedkin and Johnsen 2011).

Besides, Tindall and Piggot (2015) discovered the reasons why people are supportive or reluctant about climate change actions. Climate change attitudes are not only based on personal characteristics but also influenced by interpersonal relationships or social networks. Humans tend to establish a social relationship and shape their values and attitudes that are similar to others in their network.

According to Social capital theory, social interaction ties could be considered as the central bridge to connect information resources (Tsai and Ghoshal 1998). SNS provides a platform for their users to connect to others, therefore they could spread environment-related information easily and rapidly. Moreover, the social interaction on SNS can lead to a better understanding of environmental issues as well as directions of environmental activities.

As discussed, it is reasonable to believe that social interaction ties cause individuals to have favorable attitudes towards e-petition. Hence, we offer the following hypothesis:

Hypothesis 6: *Social Interaction Ties will have a positive impact on Attitude towards E-petition*

In the literature, the research identifies and investigates the influence of trust on voting for change. Some have argued that trust is positively connected with the voting turnout. Citizens who are more trusting towards institutionalized participation are more likely to vote (Li 2008). Additionally, Belanger and Nadeau (2002) mentioned that political trust can establish favorable attitudes that possibly make people accept a certain degree of inadequate performance of a particular political system. Moreover, trust in government may enhance individuals to cooperate with authorities and actively take part in e-participation (Scherer and Wimmer 2014).

On SNS, the trust allows persons to exchange information freely, which could enhance the effectiveness of collaboration (Robert Jr, Dennis et al. 2009). Trust motivates the user to engage in cooperative interaction (Chen, Elakhdary et al. 2019) and e-petition is a kind of activity that needs a lot of collaborations of users. A previous study showed that the final decision to take part in a petition is heavily influenced by the trust to the e-petition message (Ha & Chen, 2017). Ha and Chen (2017) also showed out that the antecedent of trust to the e-petition message are sources of credibility and argument quality. On SNS, when people trust other members, each member could become a source of credibility and enhance the trust of the call of petition. Thus, the trust to members could change the attitude toward the e-petition. We propose that:

Hypothesis 7: *Trust to members will have a positive impact on Attitude towards E-petition*

According to the Theory of Reasoned Action, the norm is considered an important factor that affects an individual attitude (Fishbein and Ajzen 1975). This finding is consistent with other past studies (Vallerand, Deshaies et al. 1992, Chang 1998) that proved the important link between subjective norms and attitudes. This significant relationship might be derived from the effect of the social environment on shaping a persons' attitude. Tarkiainen and Sundqvist (2005) also highlighted the subjective norms' effect on attitudes is related to the moral decision-making process which reflects environmental concern. For example, buying organic food could be considered as ethical behavior, and favorable or unfavorable attitudes towards green purchasing

"pass on" among individuals (Tarkiainen and Sundqvist 2005). Recent research indicates the strong positive relationship between environmental social norms and attitudes. Thus, we make the following prediction:

Hypothesis 8: E-petition norms on SNS will have a positive impact on Attitude towards E-petition

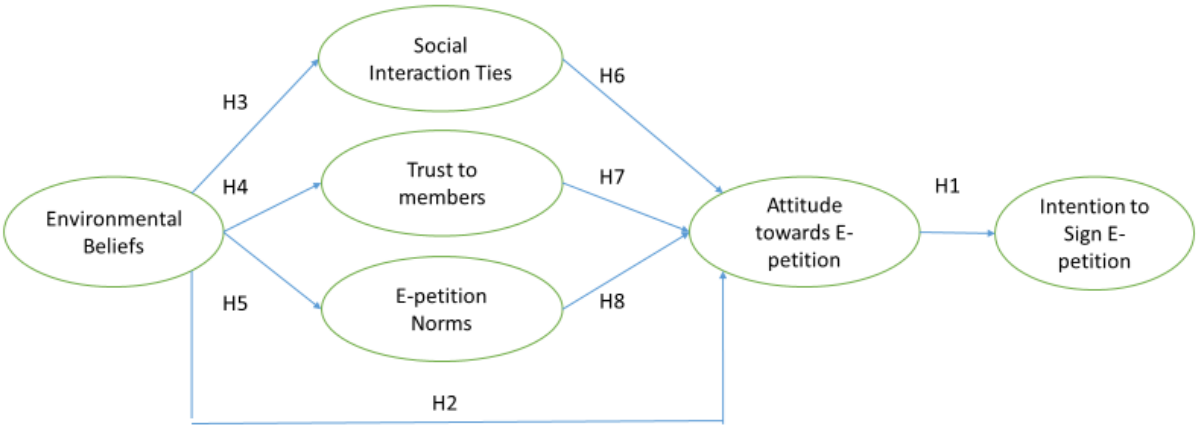


Figure 1: The framework of this study

RESEARCH METHODOLOGY

3.1 Sampling

Because most e-petition platforms are popular in the US, we used Mturk to collect data. We collected 152 respondents in total. The characteristic of the respondents is shown in Table 1. It is shown that 57.2% of respondents are male and most of them are in the age of 31-40 (32.2%), have a bachelor's degree (52.6%), and have 1-5 years of working experience.

Table 1: Characteristics of Respondent

Characteristics	Frequency (N=152)	Percent (100%)
Gender		
Male	87	57.2
Female	65	42.8
Age		
21–30	46	30.3
31-40	49	32.2
41-50	22	14.5
50-60	23	15.1
>60	12	7.9
Education		
High school or below	40	26.3
Bachelor's degree	80	52.6
Master's degree	27	17.8
Doctoral degree	5	3.3
Working background		
Education	22	14.5

Characteristics	Frequency (N=152)	Percent (100%)
Industry	32	21.1
IT	38	25.0
Other	51	33.6
Retireme	2	1.3
Unemploy	7	4.6
Working experience (years)		
1-5	63	30.0
5-10	54	25.7
10-20	42	20.0
> 20	50	23.8
Time on Social media per day		
0-1 hours	62	40.8
1-2 hours	49	32.2
2-3 hours	29	19.1
3-4 hours	5	3.3
Above 4 hours	7	4.6
Number of times to sign e-petition		
0	26	17.1
1-2	58	38.2
3-4	19	12.5
>4	49	32.2

3.2 Measurements

This study adapted the measurement from previous studies and modified them to the current context. All items used 7 point Likert scales. The detailed measurement is shown in Table 2.

Table 2:

Construct	Items	Questions	Source
Environmental Beliefs (EB)	eb1	We are approaching the limit of the number of people the earth can support	(Dunlap and Van Liere 1978)
	eb2	The balance of nature is very delicate and easily upset	
	eb5	When humans interfere with nature it often has disastrous consequences	
	eb7	To maintain a healthy economy, we will have to control industrial growth	
	eb8	Humans must live in harmony with nature in order to survive	
	eb9	The earth is like a spaceship with only limited room and resources	
	eb11	There are limits to which our industrialized society can expand	
	eb12	Mankind is severely abusing the environment	
Social Interaction Ties (SI)	si1	I engage in a high level of interaction with other members in the SNS to discuss environmental issues.	(Chiu, Hsu et al. 2006)
	si2	I spend a lot of time interacting with some members of the SNS regarding environmental issues.	
	si3	I have frequent communication with some members of the SNS regarding environmental issues.	
Trust to members (TRUST)	tru1	Members who are using SNS for environmental activities are good and honest.	(Chiu, Hsu et al. 2006)

Construct	Items	Questions	Source
	tru2	Members who are using SNS for environmental activities meet my expectations in terms of honesty.	
Intention to Sign E-petition (INTE)	int1	I will sign this petition	(Ha and Chen 2017)
	int2	I think people should sign this petition	
	int3	I will take part in this petition	
	int4	I will tell others to sign this petition	
Attitude towards E-petition (ATT)	atti1	I like the idea of signing e petition	(Paul, Modi et al. 2016)
	atti2	I have a favorable attitude toward signing e-petition for environment	
	atti3	Signing the e-petition is a good idea to protect the environment.	
E-petition Norms (NORM)	nor1	People on social networks think I should sign the e-petition	(Paul, Modi et al. 2016)
	nor2	My friends on social networks think I should sign the e-petition	
	nor3	People whose opinions I value think I should sign the e-petition	
	nor4	People who are important to me think I should sign the e-petition	

RESULTS

4.1 Descriptive Statistic

First, we run Factor Analysis on SmartPLS 3.0 (Ringle, Sven et al. 2015) to test reliability, convergent, and discriminant validity of measurements. After running CFA, we have to delete 4 items of Environmental Beliefs (eb3,4,6,10). In Table 3, the standardized loadings of all remained items are bigger than 0.7 which is higher than the rule of thumb suggested by Hair et al. (2011). The measurement also has high reliability when 2 reliability indexes (composite reliability – C.R. and Cronbach's alpha) are greater than 0.7 (Fornell and Larcker 1981, Bagozzi 2011). The AVE of all constructs is bigger than 0.5 indicate the measurements meet the requirement for convergent validity (Chin 1998). According to the Fornell-Larcker criterion, constructs' square root of AVE should be bigger than other inter-construct correlations indicates the high discriminant validity. Table 4 shows that the measurements met the requirement of discriminant validity.

Table 3. Descriptive Statistics and CFA Results

Construct	Items	Mean	SD	Factor Loading	Cronbach's Alpha	C.R.	AVE
Environmental Beliefs (EB)	eb1	4.763	1.953	0.706	0.882	0.907	0.550
	eb2	5.447	1.580	0.844			
	eb5	5.474	1.478	0.778			
	eb7	5.421	1.440	0.658			
	eb8	5.967	1.205	0.735			
	eb9	5.112	1.734	0.756			
	eb11	5.776	1.171	0.672			
	eb12	5.737	1.413	0.768			
Social Interaction Ties (SI)	si1	4.500	1.630	0.941	0.933	0.957	0.881
	si2	4.421	1.644	0.930			
	si3	4.454	1.681	0.944			
Trust to members (TRUST)	tru1	4.605	1.531	0.931	0.864	0.936	0.880
	tru2	4.717	1.536	0.945			
E-petition Norms (NORM)	int1	4.717	1.532	0.957	0.961	0.971	0.894
	int2	4.757	1.589	0.901			

Construct	Items	Mean	SD	Factor Loading	Cronbach's Alpha	C.R.	AVE
	int3	4.789	1.584	0.963			
	int4	4.730	1.564	0.892			
Attitude towards E-petition (ATT)	atti1	4.592	1.914	0.962	0.956	0.972	0.920
	atti2	4.664	1.871	0.966			
	atti3	4.704	1.799	0.949			
Intention to Sign E-petition (INTE)	nor1	4.362	2.082	0.933	0.947	0.962	0.863
	nor2	4.849	1.708	0.928			
	nor3	4.395	2.081	0.958			
	nor4	4.026	2.093	0.962			

Table 4: Correlations between research constructs

	ATT	EB	INTE	NORM	SI	TRUST
ATT	0.959					
EB	0.488	0.742				
INTE	0.872	0.508	0.929			
NORM	0.760	0.434	0.687	0.946		
SI	0.419	0.115	0.351	0.404	0.939	
TRUST	0.523	0.236	0.481	0.413	0.304	0.938

Common method bias

We use VIF values generated by SmartPLS for all constructs to asses the commance method bias (Kock 2015). The results showed that all VIF of constructs in a model were smaller than the 3.3, which indicates the model was free from CMB.

The results of PLS Analysis

Using SmartPLS 3.0 (Ringle, Sven et al. 2015), PLS algorithm is run to estimate the coefficients of proposed hypotheses. Then, following Hair et al.'s (2011) recommendation, we run bootstrap with 5,000 samples to assess the significance of coefficients. The results of PLS analysis is shown in Table 5.

Table 5: *The PLS results*

	Hypothesis	beta	t	p	Results
H1	Attitude -> Intention to Sign E-petition	0.872	33.563	0.000	Supported
H2	Environmental Beliefs -> Attitude towards E-petition	0.188	3.013	0.003	Supported
H3	Environmental Beliefs -> Social Interaction Ties	0.115	1.248	0.212	*Unsupported*
H4	Environmental Beliefs -> Trust to members	0.236	2.954	0.003	Supported
H5	Environmental Beliefs -> E-petition Norms	0.434	6.826	0.000	Supported
H6	Social Interaction -> Attitude towards E-petition	0.111	1.772	0.076	*Unsupported*
H7	Trust to members -> Attitude towards E-petition	0.221	3.084	0.002	Supported

H8	E-petition Norms -> Attitude towards E-petition	0.543	8.176	0.000	Supported

From Table 5, we can see that Attitude towards E-petition highly affects the Intention to sign e-petition (β=0.872, p<0.001) which means H1 is supported. H2, H4, and H5 are supported since Environmental beliefs influence Attitude towards E-petition, Trust to members, and E-petition Norms (β_{H2}=0.188, p<0.01; β_{H4}=0.236, p<0.05 and β_{H5}=0.434, p<0.001 respectively). However, the insignificant of H3 showed that these beliefs don't affect social interaction ties. Besides, Social interaction ties don't have a significant impact on Attitude towards E-petition while both Trust and Norms have a positive influence on this attitude (β_{H6}=0.111, p>0.05; β_{H7}=0.221, p<0.05 and β_{H5}=0.543, p<0.001 respectively).

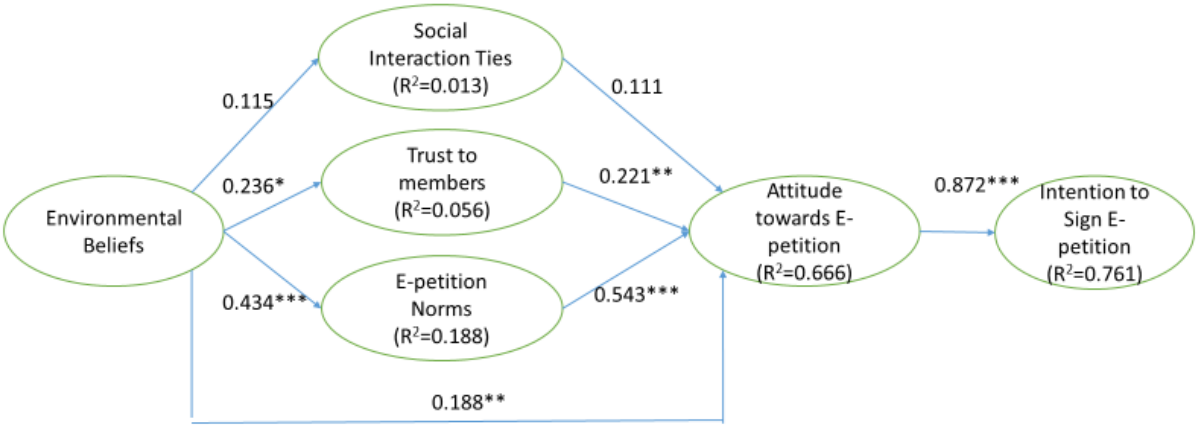


Figure 2: The results of PLS

DISCUSSION AND CONCLUSIONS

This study provides contributions to the literature of social capital on SNS by utilizing its concepts into a new context – the e-petition context. Besides, in this study, the important role of environmental beliefs are confirmed when it has significant impacts on trust to members, e-petition norm, and especially the Attitude toward e-petition. However, environmental beliefs don't have a significant impact on social interaction ties. One of the possible reasons for this phenomenon is that environmental beliefs exist relatively popular in the community and it is not the main reason for promoting interactions on social networks. Besides, social interaction ties also don't influence attitude towards e-petition which indicates that the interaction between people on social media is not the cause of attitude change toward the e-petition. The main effect of this study emphasized that trust to members could significantly impact on attitude toward e-petition, which leads to the intention to sign the petition. In addition to that result, the study also shows that e-petition norms have the highest impact on attitude towards E-petition, which reveals its important role in e-petition campaigns. Indeed, according to previous studies, people's norms play a key role in an individual's pro-environment activities (Yoon 2018).
For practical implications, this study suggested that e-petition initiators should pay attention to educate SNS users about environmental issues. The education of environmental-related issues could enhance the environmental beliefs and therefore increase the attitude toward e-petition.

Moreover, e-petition initiators who want to run an e-petition campaign on SNS effectively also need to choose reputable and highly trusted people to spread the call for petition.

REFERENCES

Ajzen, I. (1991). "The theory of planned behavior." Organizational behavior and human decision processes **50**(2): 179-211.

Bagozzi, R. P. (2011). "Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations." MIS Quarterly: 261-292.

Bechtel, C. (1999). "An investigation into the antecedents of trust in cooperative buyer-supplier relationships."

Belanger, E. and R. Nadeau (2002). "Political trust and the vote in multiparty elections." Trabajo Presentado en la reunión anual de la American Political Science Association, Boston **29**.

Bernoff, J. and C. Li (2008). "Groundswell: Winning in a world transformed by social technologies." Boston: Harvard Business School Publishing.

Chang, M. K. (1998). "Predicting unethical behavior: a comparison of the theory of reasoned action and the theory of planned behavior." Journal of business ethics **17**(16): 1825-1834.

Chen, J. V., et al. (2019). "The Continuance Use of Social Network Sites for Political Participation: Evidences from Arab Countries." Journal of Global Information Technology Management **22**(3): 156-178.

Chen, M.-F. and P.-J. Tung (2014). "Developing an extended theory of planned behavior model to predict consumers' intention to visit green hotels." International Journal of Hospitality Management **36**: 221-230.

Chen, R. (2012). "Living a private life in public social networks: An exploration of member self-disclosure." Decision Support Systems **55**(3): 661– 668.

Chin, W. W. (1998). "Issues and opinions on structural equation modeling." MIS Quarterly **22**(1): 7-10.

Chiu, C. M., et al. (2006). "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories." Decision Support Systems **42**(3): 1872-1888.

Corral-Verdugo, V. and L. I. Armendariz (2000). "The "new environmental paradigm" in a Mexican community." The journal of environmental education **31**(3): 25-31.

Corral-Verdugo, V. c., et al. (2003). "Environmental beliefs and water conservation: An empirical study." Journal of Environmental Psychology **23**(3): 247-257.

Cruickshank, P., et al. (2010). Signing an e-petition as a transition from lurking to participation, na.

Dunlap, R. E. (2008). "The New Environmental Paradigm Scale: From Marginality to Worldwide Use." The journal of environmental education **40**(1): 3-18.

Dunlap, R. E. and K. D. Van Liere (1978). "The "new environmental paradigm"." The journal of environmental education **9**(4): 10-19.

Fishbein, M. and I. Ajzen (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA, Addison-Wesley.

Fornell, C. and D. Larcker (1981). "Evaluating structural equation models with unobservable variables and measurement error." Journal of Marketing Research **18**(3): 39-50.

Friedkin, N. E. and E. C. Johnsen (2011). Social influence network theory: A sociological examination of small group dynamics, Cambridge University Press.

Gadenne, D., et al. (2011). "The influence of consumers' environmental beliefs and attitudes on energy saving behaviours." Energy policy **39**(12): 7684-7694.

Ha, Q.-A. and V. J.-C. Chen (2017). The impact of Argument Quality and Source Credibility on Decision to Participate in Online Petition: The Moderator effect of Emotions. Decision Sciences Institute.

Hair, J. F., et al. (2011). "PLS-SEM: Indeed a silver bullet." Journal of Marketing Theory and Practice **19**(2): 139-151.

Hamid, P. N. (1989). "Self-monitoring and locus of control as determinants of social interaction: A preliminary investigation." Social Behavior and Personality: an international journal **17**(2): 125-134.

Hartmann, P. and V. Apaolaza-Ibáñez (2012). "Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern." Journal of Business Research **65**(9): 1254-1263.

Jalali, M. S., et al. (2016). "Information diffusion through social networks: The case of an online petition." Expert Systems with Applications **44**: 187-197.

Kock, N. (2015). "Common Method Bias in PLS-SEM." International Journal of e-Collaboration **11**(4): 1-10.

Kotchen, M. J. and S. D. Reiling (2000). "Environmental attitudes, motivations, and contingent valuation of nonuse values: a case study involving endangered species." Ecological Economics **32**(1): 93-107.

Kroon, S. M. A. v. d., et al. (2002). Social capital and communication. The Hague, Agricultural Economics Research Institute.

Li, L. (2008). "Political trust and petitioning in the Chinese countryside." Comparative Politics **40**(2): 209-226.

Lin, K. Y. and H. P. Lu (2011). "Intention to continue using Facebook fan pages from the perspective of social capital theory." CyberPsychology, Behavior & Social Networking **14**(10): 565-570.

McFarlane, B. L. and P. C. Boxall (2003). "The role of social psychological and social structural variables in environmental activism: an example of the forest sector." Journal of Environmental Psychology **23**(1): 79-87.

Milošević-Đorđević, J. S. and I. L. Žeželj (2017). "Civic activism online: Making young people dormant or more active in real life?" Computers in Human Behavior **70**: 113-118.

Nahapiet, J. and S. Ghoshal (1998). "Social capital, intellectual capital, and the organizational advantage." Academy of Management Review **23**(2): 242-266.

Okoli, C. and W. Oh (2007). "Investigating recognition-based performance in an open content community: A social capital perspective." Information & Management **44**(3): 240-252.

Ozaki, R. (2011). "Adopting sustainable innovation: what makes consumers sign up to green electricity?" Business strategy and the environment **20**(1): 1-17.

Paul, J., et al. (2016). "Predicting green product consumption using theory of planned behavior and reasoned action." Journal of retailing and consumer services **29**: 123-134.

Pickett‐Baker, J. and R. Ozaki (2008). "Pro‐environmental products: marketing influence on consumer purchase decision." Journal of Consumer Marketing **25**(5): 281-293.

Ramayah, T., et al. (2010). "Green product purchase intention: Some insights from a developing country." Resources, conservation and recycling **54**(12): 1419-1427.

Ringle, C. M., et al. (2015). SmartPLS. Bönningstedt, SmartPLS.

Robert Jr, L. P., et al. (2009). "Individual swift trust and knowledge-based trust in face-to-face and virtual team members." Journal of Management Information Systems **26**(2): 241-279.

Scherer, S. and M. A. Wimmer (2014). Conceptualising trust in E-participation contexts. International Conference on Electronic Participation, Springer.

Scott, D. and F. K. Willits (1994). "Environmental attitudes and behavior: A Pennsylvania survey." Environment and Behavior **26**(2): 239-260.

Sharma, B., et al. (2017). "Environmental beliefs, norms and behaviours: An investigation of their relationships using data from green consumers." The Journal of New Business Ideas & Trends **15**(1): 1.

Tanner, C. and S. Wölfing Kast (2003). "Promoting sustainable consumption: Determinants of green purchases by Swiss consumers." Psychology & Marketing **20**(10): 883-902.

Tarkiainen, A. and S. Sundqvist (2005). "Subjective norms, attitudes and intentions of Finnish consumers in buying organic food." British food journal.

Tindall, D. B. and G. Piggot (2015). "Influence of social ties to environmentalists on public climate change perceptions." Nature Climate Change **5**(6): 546-549.

Tsai, W. and S. Ghoshal (1998). "Social capital and value creation: The role of intrafirm networks." Academy of management Journal **41**(4): 464-476.

Ulo, K. L. M., et al. (2019). "Factors influencing internet users' intention to sign e-petitions." Transforming Government: People, Process and Policy.

Vallerand, R. J., et al. (1992). "Ajzen and Fishbein's theory of reasoned action as applied to moral behavior: A confirmatory analysis." Journal of personality and Social Psychology **62**(1): 98.

Van der Kroon, S. (2002). Social capital and communication, LEI.

van Noort, G., et al. (2012). "Social connections and the persuasiveness of viral campaigns in social network sites: Persuasive intent as the underlying mechanism." Journal of Marketing Communications **18**(1): 39-53.

Wasko, M. M. and S. Faraj (2005). "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice." MIS Quarterly **29**(1): 35-57.

Yoon, C. (2018). "Extending the TAM for Green IT: A normative perspective." Computers in Human Behavior **83**: 129-139.

DECISION SCIENCES INSTITUTE

Mass Layoffs and State Business Climates: What About the Non-Manufacturing Sector?

Christopher J. Surfield
Dobson DaVanzo and Associates
450 Maple Ave E #303
Vienna, VA 22180
Phone 703-260-1768
Email: Christopher.surfield@dobsondavanzo.com

C. Surender Reddy,
Department of Management and Marketing,
College of Business and Management,
Saginaw Valley State University,
7400 Bay Road
University Center, MI 48710
Phone 989-964-4099
Email: reddy@svsu.edu

ABSTRACT

Prior work on state-level public policy has focused on the manufacturing sector. This is the first study to address any relationship between state-level business policies and mass job loss in the non-manufacturing sector. Spatial model estimation is also used to identify any effects that policies have beyond a state's border. Unlike the case of manufacturing employment, we find very limited evidence that state-level public policy has much of a relationship with losses outside of that sector. (J50, J60, M50) . . .

KEYWORDS: Business Climates, Layoffs, Government Policies, Non-Manufacturing, Spatial Modeling

INTRODUCTION

States tout their rankings on the various 'best states for' lists to demonstrate their attractiveness to businesses. Prior empirical evidence suggests that 'pro-business' policies, such as those which foster more flexible labor markets or lower tax rates, seem to have a relationship with various labor market outcomes within a state. Namely, there is a suggestion that 'pro-business' states have both higher growth in their overall employment levels, as well as decreased losses within their manufacturing sector, when compared to their more poorly-ranked counterparts.

States face clear incentives to craft 'pro-business' policies. Such policies may not only attract new employers but may also prevent the poaching of their existing employment base by rival states. Firms evaluate where to locate themselves based upon the ability to maximize their profits (Waslyenko and McGuire, 1985; Greenhut, 1956). Apart from sales taxes, public policy changes are unlikely to have much of an effect on the revenue side of this profit-maximizing equation. Where state legislation might make a difference is on the costs of doing business within that state.

State-level policies that lower operating costs, perhaps through lower regulatory compliance costs or lower corporate tax rates, might help to increase firm profits, ceteris paribus.

Policies which may help firms maximize their profits through lower labor force adjustment costs, however, might also make it easier to dismiss workers or relocate their facilities to rival state. For example, if a state adopted policies designed to increase labor market flexibility or lowered unemployment insurance premiums, they might also find that any jobs gained might be at risk. If establishments are truly 'footloose' in the sense that they can be easily attracted to states with more attractive business climates, then they might also be the same firms which could be subsequently poached by competing states offering even more beneficial business climates. [Further, policies designed to relax business regulations and decrease unemployment insurance premiums themselves might directly lead to less stable employment (Autor, 2003; Autor, Kugler and Kerr, 2007; Surfield, 2014)]

One concern regarding the impact that policy differences might have on labor market outcomes is attached to the focus of the prior work. Much of the focus of the prior empirical work has been on manufacturing jobs. However, there may be structural differences between manufacturing and non-manufacturing establishments with these differences leading to business climates having differential effects on employment levels.

In the case of long-term investment, the manufacturing sector faces large, and typically sunk, costs. These investments might cause manufacturers to respond differently to business climate changes, particularly when such polices affect the cost of their (variable) worker input. Manufacturing firms may be more apt to maintain employment in states which make the labor input cheaper, ceteris paribus. It also seems likely the case that, in the manufacturing sector, policy differences might play significant role when deciding whether to maintain or increase employment once a facility is established, but a more marginal role in whether to relocate a facility or not. [Further, policies designed to relax business regulations and decrease unemployment insurance premiums themselves might directly lead to less stable employment (Autor, 2003; Autor, Kugler and Kerr, 2007; Surfield, 2014)] These large facility investments would suggest that manufacturing establishments, once established, are less likely to be lured to other states offering a better business climate.

The structure of this paper is as follows. After a review of the representative studies regarding state business climates and employment outcomes, we discuss our data and the construction of our dependent variable. We merge data culled from the Mass Layoff Statistics with those contained in the Quarterly Census of Employment and Wages to construct our primary variable of interest. In quantifying what relationship, if any, state policy differences have with mass job loss, we also estimate a spatial model to take into consideration the possibility that such policy effects may spill over to surrounding states. We then present our findings regarding various state public policies and jobs lost outside of the manufacturing sector due to mass layoff events. Concluding remarks follow. . .

EXISTING WORK

State officials have particularly acute incentives to craft "pro-business" policies to maintain their employment bases, especially with respect to large employers. If a layoff event occurs at a large employer, states see a decline in both corporate and personal income tax revenues as well as increases in the number of unemployment insurance (UI) claims. Losses are further compounded by the spillover effects that these layoff events have on the local area's remaining economic activity. States could garner benefits being viewed as fostering a more pro-business environment if such policies help them to at least minimize the severity and persistence of earnings losses during economic downturns. At the very least, the retention of jobs helps states avoid the inherent tax revenue losses and the increased demands placed upon unemployment insurance programs.

To get at the effects of state-level public policy competitions on employment outcomes, [The 'definitiveness' of the various rankings themselves in their ability to accurately measure a state's business climate is not without controversy. See, for example, Fisher (2005) for a criticism and evaluation of the rankings.] prior investigations have deployed the use indices to broadly measure a state's public policy portfolio. One such example is Kolko, Neumark, and Mejia (2013), who explored whether business climate ratings have predictive value for new job creation, gross state product, wages, and total employment. The authors generally concluded that state-level climate ratings are not significantly related to these economic outcomes. Of the eleven rankings included in their analyses, they found only two had significant predictive value: the Fraser Institute's Economic Freedom of North America (EFNA) index and the Tax Foundation's State Business Tax Climate (SBTC) index. Kolko et al. report that states scoring higher on these two indexes enjoy faster employment growth.

The authors noted that the indexes used in their primary analyses are the result of an aggregation of several policy subcomponents. They explored these subcomponents to determine if policy areas differed in their predictive power for a state's employment picture. The size of a state's government expenditures seemed to have significant, positive implications for a state's employment picture. The authors speculated that expenditures might be associated with infrastructure improvements or educational expenditures to upgrade the state's workforce skillsets.

Surfield and Reddy (2016) take up the question of whether the manufacturing job gains found in the literature are increases in the number of new jobs or just a change in the distribution of where they are located. The authors examine the possibility that states adopting more flexible labor markets or more relaxed regulatory policies might also see differences in the number of manufacturing jobs lost. The same two indexes found to have predictive power in Kolko et al. were used to measure the broader business climate within a state.

If policies designed to attract jobs into a state simply change the distribution of where manufacturing establishments are located, then states with 'better' business climates may very well see lower numbers of manufacturing job losses. More poorly-ranked states, however, are likely see an *increase* in manufacturing job losses when such jobs move to more 'pro-business' neighboring states. If, however, these policies foster the creation of new manufacturing jobs, then one might expect to see 'pro-business states' having a diminished number of job losses, with little to no effect on the number of jobs lost by surrounding states.

The authors found evidence that more flexible labor market policies, be it right-to-work statutes or broader pieces of legislation, seem to be associated with a reduced number of UI claims within the manufacturing sector. Further, not only did the states adopting these policies see the benefit of fewer UI claims, but so did neighboring states as well. Part of the explanation as to why such a spillover effect exists might lie in worker movements. Unemployed workers may be moving away from poorly-ranked states to the better-ranked neighboring states to take advantage of more favorable (manufacturing) labor markets.

Given that state business climates have been found to have significant effects on economic outcomes within the manufacturing sector, we seek to see if these effects are found outside of manufacturing. For example, differences in the capital structures of the two employment sectors might lead to differences in how employers respond to changes in state business climates. We use the same business indexes deployed in Kolko et al. (2013) and Surfield and Reddy (2016). Given the legislative costs, in terms of both time and regulatory enforcement, of changing state public policies, we focus on job losses observed from large employers. We turn next to the data used to facilitate this investigation. . .

DATA

Our primary source of mass job loss comes from the Mass Layoff Statistics (MLS). The MLS data are then combined with those contained in the Quarterly Census of Employment and Wages (QCEW). We first discuss the construction of our dependent variable. We then turn to our primary explanatory variables of interest, the business climate indexes. Finally, we discuss our other state-level explanatory variables.

A. Job Loss Measure

The Mass Layoffs Statistics collected data on worker separations due to a mass layoff event (MLE). For a job loss to be associated with a MLE, it must have occurred as part of a layoff event in which an establishment had at least fifty unemployment insurance claims being lodged against it within a five-week window. [The establishment of this lower bound of fifty unemployment insurance claims does narrow the universe of establishments which would be covered by the MLS. Tabulations of Census County Business Pattern Data suggest that approximately five percent of all establishments employ fifty or more workers, which would be of sufficient size as to encounter a MLE. Further details on the scope of the MLS, as well as the decision to limit a MLE to fifty or more workers, are contained in Brown and Siegel (2005).] Analysts contact the affected employers once a state's unemployment insurance (UI) agency has identified a potential MLE. Data are then collected regarding the number of workers impacted, [The number of separations reported by an employer may differ substantially from the number of unemployment insurance claims. Workers who separate from an employer, but fail to apply for unemployment benefits, would lead to MLE-induced job losses being undercounted. Alternatively, an employer could report fewer separations relative to UI claims if a firm believed that the layoffs were temporary, such as for a product line changeover or due to delays in an anticipated start of a new contract/project.] the cause(s) of the MLE, [We considered the possibility that not all job losses would be responsive to a state's set of public policy tools. For example, establishment shut-downs due to extreme financial exigencies would unlikely be completely prevented with changes to a state's business climate. While identifiable in the aggregate totals, the data provided at the industrial level, however, do not allow us to identify the number of jobs lost due to such instances. This becomes particularly problematic when we attempt to net out manufacturing job losses.]and select demographic details of the separated workers. These data are then aggregated and reported to the Bureau of Labor Statistics (BLS) for publication in a standardized format.

The BLS does exercise some discretion in terms of publicly reporting data on MLEs. If the public release of MLE data could uniquely identify an establishment, or a limited set of establishments, then data on these MLEs were suppressed. The data are then recoded to note that they are not being reported. In these events, our results might be viewed as a lower-bound to any job losses incurred by way of a MLE.

Finally, the BLS provides a rough measure of the severity of MLE-induced job losses by making a distinction in terms of an UI claim's duration. The BLS considers UI claims that last at least thirty days to be 'extended' claims. These extended job loss tallies are released quarterly, with aggregated annual tallies being reported at the end of each year.

Our dependent variable is constructed from the annual tallies of extended UI claims. Extended unemployment durations are not only negatively related to the likelihood of being extended another offer of employment (Meyer, 1990; Kletzer, 1998), but might also signal that the job losses were permanent in nature. Those workers claiming unemployment benefits for less than thirty days, or reporting to have been separated by their employer but opting not to file for UI benefits, might have done so because they had a recall expectation, already lined up a subsequent offer of employment in anticipation of the job loss, or moved to seek employment

elsewhere. In any case, these shorter-term UI claim durations might suggest that the unemployment induced by the MLE is more likely to be a transitory feature of a worker's history relative to those filing an extended claim.

To construct our dependent variable, we take the number of total extended UI claims reported in the MLS and net out those UI claims observed in the manufacturing sector. This tally is then standardized. We report job loss as the rate of incidence at which jobs were shed.

The standardization process requires the use of our second dataset: the Quarterly Census of Employment and Wages. The QCEW reports the average number of jobs within a given state in any given year. The number of non-manufacturing workers reported in the QCEW (per ten thousand workers) were used to divide the total number of non-manufacturing mass job loss tallies. We used the prior year's QCEW employment tallies as our divisor as it seems more likely that any jobs lost were those reported as being filled the year prior.

Collection of MLS data began in the second quarter of 1995. The collection of MLE-related job losses ended in the first quarter of 2013. Given the incompleteness of the annual tallies for these two years, we excluded them from our analyses. For all usable waves of MLS data, we have available the corresponding data required in the QCEW. Thus, our analysis covers the 1996-2012 MLS waves.

B. State Business Climate Measures

Our primary focus is on mass job loss and its relationship with the business climate fostered by various public policy decisions. We employ two indexes to measure the various dimensions of the business climate. Both were found by Kolko et al. (2013) and Surfield and Reddy (2016) to have a relationship with employment outcomes.

We deploy both the Fraser Institute's Economic Freedom of North America (EFNA) index [Stansel et al. (2014).]and the Tax Foundation's State Business Tax Climate (SBTC) [Stansel et al. (2014).] index. [We note that two states were excluded from our analyses. Alaska and Hawaii were excluded from our analyses given that their distance from other states makes it unlikely that employers would move operations to those states on the basis of their business climate alone. We also exclude the District of Columbia from our analyses given the fairly small number of MLEs observed, as well as its exclusion in our business climate rankings.] The EFNA index is an overall ranking of a state's business climate. A state's overall score is derived from an assessment made by the Fraser Institute across three public policy areas: government and welfare expenditures size, labor market freedom, and takings and discriminatory taxation policies. The overall index and each sub-index are all scored from 0 to 10, where 10 always implies a more pro-business climate. For example, then, a *higher* government expenditure score actually means that government expenditures are *lower*. (Likewise, higher labor market regulation and takings scores imply fewer regulations and takings.)

Given that the EFNA started in 1981, we can completely match this index to the entire time period covered by our MLS data. The Fraser Institute reports each state's overall score, as well scores in each of the three sub-areas. [The EFNA has been used to explain differences in other economic outcomes, such as the unemployment rate, labor-force participation rate and the employment-population ratio (Heller and Stephenson, 2014).] As was the case in Kolko et al. (2013), we convert state-level scores into z-scores by subtracting off the year's mean across states and dividing by the year's cross-state standard deviation. [This was done out of concern that the criteria which determines the score might change slightly across the years. These standardized scores thus provide a state's ranking relative to its peers while controlling for this potential inconsistency.]

The second index that we employ is the Tax Foundation's State Business Tax Climate (SBTC) index. [This was done out of concern that the criteria which determines the score might

change slightly across the years. These standardized scores thus provide a state's ranking relative to its peers while controlling for this potential inconsistency.] As was the case with the EFNA index, a state's overall score in the SBTC index is derived from the scores it garnered across a number of sub-measures. While the EFNA focuses on a state's overall business climate, the Tax Foundation focuses narrowly on tax policy. The SBTC scores states on their corporate tax rates, property tax rates, sales tax rates, and unemployment insurance rate premiums.

The SBTC index and subcomponents are all scored from 0 to 10, where a 10 indicates the most favorable tax climate. For example, in the case of sales taxes, a 10 would indicate that a state does not levy any sales taxes. As with EFNA scores, we also standardize the SBTC scores into z-scores. Scores for personal tax rates were excluded on the basis that such taxes likely do not influence the decision of an employer to terminate jobs, even if they might influence the likelihood of a worker moving to/from that particular state. [This was done out of concern that the criteria which determines the score might change slightly across the years. These standardized scores thus provide a state's ranking relative to its peers while controlling for this potential inconsistency.] The SBTC was first conducted in 2003 and we are able to update the index until the end of our MLS data in 2012.

If more favorable business climates help states to prevent job losses, then we would expect to see negative coefficient estimates attached to these indexes. Conversely, if states are weakening their labor market protections and other regulations to encourage employers to relocate facilities to their states, we then may expect to see positive coefficient estimates. In this case, positively-signed coefficient estimates would imply that 'lower' ranked states, e.g. those with less-friendly business climates, see lower rates of UI claims relative to 'higher' ranked pro-business states. States with increased labor market flexibility might have also made it easier for establishments to shed workers. Further, those states which are less business-friendly might also have imposed more cumbersome labor force adjustment costs, such as unemployment insurance rates or other penalties, thereby giving firms an incentive to avoid laying off large numbers of workers (even if such policies do not necessarily attract new firms).

C. Other Explanatory Variables

Our final set of explanatory variables measure other state-level characteristics that may be correlated with MLE-induced job losses. The first is the average weekly wage paid to workers reported in the QCEW for each state annually. This variable controls for any influence that labor costs might have on a state's incidence of mass job loss.

Older establishments may be more likely to downsize than younger establishments. We construct a variable from the County Business Patterns (CBP) data that measures the percentage of a state's non-manufacturing establishments employing 250 or more workers. This variable was constructed as a potential proxy for the relative age of a state's employment base. [This was done out of concern that the criteria which determines the score might change slightly across the years. These standardized scores thus provide a state's ranking relative to its peers while controlling for this potential inconsistency.]

The general labor market conditions within each state are taken into consideration with the inclusion of its unemployment rate and the size of its workforce. The average annual unemployment rates for the states were obtained from the BLS' Local Area Unemployment Statistics program. The total employment levels, measured in tens of thousands of workers, were obtained from the QCEW.

Finally, we include controls for age, educational, and ethnic characteristics for each state. These variables are from the January waves of the basic survey of the Current Population Survey. We use the fraction of a state's population aged sixteen to twenty years, as well as the fraction aged sixty-five and older; the fraction of a state's population which is African-American as well as its fraction of non-African-American, non-Caucasian residents; and the fraction of a state's

population possessing a bachelor's degree as well as the fraction holding a post-bachelor's (graduate) degree.

ESTIMATION STRATEGY

Two econometric models are estimated to help us quantify the relationship between state public policy differences and the number of jobs lost due to an MLE. The first approach is panel data ordinary least squares estimation. The results from these estimations provide our initial benchmark results. We estimate both a random effect and, to take into consideration any time-invariant state-level unobserved heterogeneity, a fixed effects estimator. [The use of fixed effects estimators required the exclusion of a commonly used measure of business climates: right-to-work status. Over the period covered by this study, only one state saw a change in their status.]

Our second estimation strategy is to deploy a spatial model to allow for the possibility that a state's legislation regarding business climates might also have indirect effects on neighboring states. For example, the job gains pointed to in the existing literature might not solely be the product of new job creation. Some of the jobs gained by pro-business states might be coming at the expense of jobs poached from rival (neighboring) states. In this case, some of the jobs gained by pro-business states might just simply be a redistribution of an existing job base as firms relocate their facilities away from less-favorable business climates to more favorable ones.

Accordingly, we rely on spatial estimation techniques outlined in LeSage and Fischer (2008) and Elhorst (2010). [The use of fixed effects estimators required the exclusion of a commonly used measure of business climates: right-to-work status. Over the period covered by this study, only one state saw a change in their status.] In particular, we estimate the Spatial Durbin Model (SDM) for panel data or

$$y_{i,t} = \alpha_i + \rho W_{i,j} y_{t-1} + X_{i,t}\beta + W_{i,j} X_{j,t}\theta + \varepsilon_{i,t}, \quad [1]$$

$$\text{where } \varepsilon_{i,t} = \lambda W \varepsilon_{i,t} + \nu_{i,t}. \quad [2]$$

The first two terms of Equation 1 are the standard intercept and an autoregressive lag in the dependent variable. The key terms in this model are the third and fourth terms.

The third term, $X_{i.t}\beta$, provides us with the direct effects associated with changes in that state's characteristics on the dependent variable. In our case, the coefficient estimates produced by this term represent an estimate of what relationship, if any, that changes in a state *i*'s business climate may have on the number of extended UI claims in its non-manufacturing sector. Noted by LeSage and Fischer (2008), β also includes any feedback effects of state policies resulting from the autocorrelation passing through neighboring states and back into the original state.

The fourth term, $W_{i,j}X_{j,t}\theta$, is what allows us to identify any possible effects that state *i*'s policies may have on neighboring states *j*. In this term, the parameter θ measures the spatial effects on neighboring states and, again, any feedback effects resulting from ρ. In our case, the coefficients obtained from this term represent the indirect effects that a state's business climate may hold for the number of extended UI claims in the non-manufacturing sector faced by neighboring states. We should note that the parameter θ is the total (cumulative) effect on neighboring states and should not be interpreted as a 'per-neighbor' effect.

To facilitate the estimation of the SDM, a weighting matrix, $W_{i,j}$, is required to identify the spatial relationships between observations. A number of ways, such as employing distance bands or nearest neighbors, have been adopted to define 'neighboring' observations. In our case, we defined only contiguous states as 'neighbors.' This was done given the dissipation of any effects related to different state climates associated with locations that are further from the border between states (Holmes, 1998). In this case, other business factors, such as transportation costs

or network effects, seemed to dominate any cost differences associated with state policy decisions.

Using Stata 11, we estimated the SDM model using our panel data through the procedure outlined in Belotti, Hughes and Mortari (2013). [If certain assumptions regarding the parameters θ, λ and ρ, are met, then the SDM model could be condensed into either a Spatial Autoregressive Model (SAR) ($\theta = \lambda = 0$) or a Spatial Error Model (SEM) ($\theta = \rho = 0$). The results of these tests, presented in the last rows of Tables 6 and 7, suggest that, on average, such models are inappropriate.] The use of the MLS panel data does provide, however, one key challenge. In specifying the spatial weight matrix, the panel dataset must be strongly balanced. Recall that not all MLE-related data are publicly reported given confidentiality considerations. This restriction leaves us with an unbalanced panel, as only thirty-five states have complete data on their MLEs being reported between 1996 and 2012. The remaining fourteen states, and their incomplete set of observations, can be accommodated in the standard panel data ordinary least squares estimations, but are unable to be used in the estimation of the SDM. [If certain assumptions regarding the parameters θ, λ and ρ, are met, then the SDM model could be condensed into either a Spatial Autoregressive Model (SAR) ($\theta = \lambda = 0$) or a Spatial Error Model (SEM) ($\theta = \rho = 0$). The results of these tests, presented in the last rows of Tables 6 and 7, suggest that, on average, such models are inappropriate.]

Recall that our dependent variable is constructed by netting out any reported manufacturing job losses from the total losses reported within a given state. To recover at least a portion of these fourteen states, we evaluated the degree to which the data suppression for these two components occurred across the years covered by our two indexes. For those state-year observations for which neither the total MLE losses nor the manufacturing-associated MLE losses are reported, we were unable to recover any of the observations. For those state-year observations, however, for which we have total MLE losses, but not the corresponding manufacturing MLE losses due to data suppression, we adopted an imputation strategy. [If certain assumptions regarding the parameters θ, λ and ρ, are met, then the SDM model could be condensed into either a Spatial Autoregressive Model (SAR) ($\theta = \lambda = 0$) or a Spatial Error Model (SEM) ($\theta = \rho = 0$). The results of these tests, presented in the last rows of Tables 6 and 7, suggest that, on average, such models are inappropriate.]

We first averaged the ratio of manufacturing job losses to total losses across the actual observed years of data for an affected state. We then used that ratio to estimate the missing year's manufacturing MLE losses. A dummy variable indicating whether the data were imputed (zero otherwise) was also constructed. This dummy variable effectively takes the observation back out of the sample as the dependent variable would be completely explained by the dummy variable for the imputed observation while leaving a fully balanced panel to facilitate the estimation of the spatial model. This allowed us to recover four states, and their sixty (four) reported (imputed) observations. [This strategy allowed us to recover four states as well in our analyses including the SBTC index. This gave us a total of thirty-seven states over the 2002-2012-time frame.]

RESULTS

Cross-tabulations of our data are presented in Table 1. The states are differentiated by whether they are a right to work (RTW) state or not. As we can see in the first row, states that adopted RTW legislation had significantly fewer number of MLE-induced UI claims relative to non-RTW states. On average, states which adopted right-to-work legislation appear to have twenty-eight fewer extended UI claims (per ten thousand workers), or nearly half of the claims observed in non-RTW states. This negative differential observed within the non-manufacturing sector is consistent with findings contained in prior work focusing on manufacturing.

Table 1: Descriptive Statistics for Extended Unemployment Duration Claims in the Non-Manufacturing Sector and Control Variables (1996-2012)

	National Average	*n*	Right-to-Work State	*n*	Non-Right-to-Work State	*n*
a. Mass Layoffs						
Standardized Incidence of Extended UI Claims	51.72 (47.16)	738	36.25 (38.98)	325	63.89** (49.45)	413
b. State Business Climate Measures						
Overall Economic Freedom Index Score	[a]		0.42 (0.77)	370	-0.35** (1.02)	446
Overall State Business Tax Climate Score	[a]		0.31 (1.03)	221	-0.26** (0.88)	269
c. Additional Control Variables						
Average Weekly Wage (in 1982-84 dollars)	366.44 (69.54)	833	332.72 (37.70)	370	393.38** (77.04)	463
Fraction of Population, Age 16-20 (in percent)	8.92 (1.36)	833	9.26 (1.45)	370	8.64** (1.21)	463
Fraction of Population, Age 65 or Older (in percent)	15.62 (2.42)	833	15.49 (2.61)	370	15.72 (2.25)	463
Fraction of Population, African-American (in percent)	10.96 (11.46)	833	12.63 (11.14)	370	9.63** (11.55)	463
Fraction of Population, Non-African American, Non-Caucasian (in percent)	4.62 (3.40)	833	4.32 (3.07)	370	4.86** (3.62)	463
Fraction of Population, Bachelor's Degree (in percent)	15.38 (3.50)	833	14.36 (2.79)	370	16.19** (3.78)	463
Fraction of Population, Post-Bachelor's Degree (in percent)	7.63 (2.97)	833	6.28 (1.74)	370	8.72** (3.28)	463
Fraction of Large Establishments (in percent)	0.52 (0.15)	833	0.51 (0.13)	370	0.53** (0.16)	463
Unemployment Rate (in percent)	5.57 (2.03)	833	5.33 (2.07)	370	5.76** (1.99)	463

Notes: Results reported as mean (standard deviation), [a] Not reported at the national level. ** denotes significance in the mean between the two right-to-work state groups at the 5% level, manufacturing employment control not reported for simplicity purposes

The second panel of Table 1 presents the average of the standardized business climate indexes used to control for the broader set of public policies. Not surprisingly, states which adopted RTW legislation are adjudicated as being more business-friendly in both the Economic Freedom Index of North America and the State Business Tax Climate Index relative to their non-RTW counterparts. Those adopting this piece of legislation garnered significantly higher, and positive, scores on both standardized indexes.

Finally, panel *c* of Table 1 suggests that there are other differences between the two state groups which help to explain the observed differences in the extended UI claim rates. On average,

states which are RTW states have lower weekly wages. RTW states are generally younger, have significantly more (less) African-American (non-African-American/Non-Caucasian) residents as a proportion of their total population, and generally less educated. RTW states also have fewer establishments that employ two hundred and fifty workers or more and see lower average unemployment rates relative to non-RTW states.

Given the competing implications that state workforce characteristics and their business climates may have on the number of extended UI claims, we turn next to our first set of multivariate results presented in Table 2. [This strategy allowed us to recover four states as well in our analyses including the SBTC index. This gave us a total of thirty-seven states over the 2002-2012-time frame.] We use the full, unbalanced, sample to obtain our baseline estimates. The first column of the table presents the coefficient estimated for a state's overall EFNA score. We use these scores to control for broader aspects of a state's business climate.

Table 2: Random and Fixed Effects OLS Estimates of Extended Job Loss Claims in the Non-Manufacturing Sector (1996-2012)

	(1)	(2)	(3)	(4)
Overall EFNA Score	-3.28 (3.51)	-5.67 (5.23)		
Size of Government Expenditures Score			0.90 (2.67)	0.24 (3.30)
Labor Market Freedom Score			-3.00 (4.16)	-8.31 (5.92)
Takings and Discriminatory Taxation Score			-2.04 (3.14)	-0.47 (5.32)
Specification	RE	FE	RE	FE
Hausman Test	χ^2=33.77 (p=0.17)		χ^2=35.04 (p=0.20)	

n = 728

Notes: Results reported as coefficient estimate (standard errors clustered at the state level). Each specification includes the additional controls listed in Table 1, total annual employment (in 10,000 workers), and sixteen-year dummies (omitted year was 1996). ***, **, * denote significance at the 0.01, 0.05 and 0.10 levels, respectively.

In column 1, we obtain a negatively-signed coefficient estimate from our random effect specification. The negatively-signed coefficient estimate has also been found in prior work. Unlike the case of manufacturing, however, the coefficient estimate attached to the EFNA scores fails to achieve significance at conventional levels.

The three subcomponents used to derive a state's overall score are used to obtain the results highlighted in column 3. We still fail to uncover any evidence that states' business climates have a relationship to the number of extended UI claims outside of their manufacturing sectors. In sum, it would appear differences in state business climates have no relationship with the number of extended UI claims in the non-manufacturing sector.

Perhaps unobserved heterogeneity might be biasing our results. We take this into consideration with the results contained in Table 2. Columns 2 and 4 present the results obtained from our fixed effects estimator. Again, we find that a state's overall EFNA score is not significantly related to the number of UI claims in its non-manufacturing sector (column 3).

Table 2's fourth column corresponds to the set of results presented in the second column, albeit now including state-level fixed effects. We again find that none of the sub-areas of a state's EFNA sub-component scores correlate with the number of extended UI claims being made outside of the manufacturing sector. In short, we fail to find any relationship between the UI claims outside of the manufacturing sector and public policies which was the case in the prior literature's focus on business climates and manufacturing employment. States with 'better' business climates see no more, and no fewer, UI claims outside of the manufacturing sector than do their more poorly ranked counterparts.

When we use SBTC scores as our proxy for state business climates, we again fail to uncover any evidence that doing better on this index correlates with a change in MLE-induced extended UI claims in the non-manufacturing sector. While the time period covered by this index is shorter than that covered by the EFNA, the same general findings hold. The results presented in the first column of Table 3 suggests that, while of the expected negative sign given the prior empirical literature, coefficient estimate attached to the standardized SBTC scores are statistically no different from zero. This result is obtained from both our random and fixed estimators.

Table 3: Random and Fixed Effects OLS Estimates of Extended Job Loss Claims in the Non-Manufacturing Sector (2003-2012)

	(1)	(2)	(3)	(4)
Overall State Business Tax Climate Score	-4.82	-8.94		
	(5.71)	(9.93)		
Corporate Tax Score			7.78	12.34
			(6.38)	(8.96)
Property Tax Score			-1.80	-4.32
			(3.39)	(5.93)
Sales Tax Score			0.25	2.16
			(3.60)	(5.54)
Unemployment Insurance Tax Score			2.52	1.93
			(3.45)	(4.13)
Specification	RE	FE	RE	FE
Hausman Test	χ^2=31.95 (p=0.044)		χ^2=32.00 (p=0.100)	

n = 437

Notes: Results reported as coefficient estimate (standard errors clustered at the state-level). Each specification includes the additional controls listed in Table 1, total employment (in 10,000 workers),
and nine-year dummies (omitted year was 2003). ***, **, * denote significance at the 0.01, 0.05 and 0.10 levels, respectively.

Finally, the last two columns present the results obtained when we use the four SBTC subcomponents which may potentially have a relationship with the number of MLE-induced extended UI claims. None of the decisions made in any of these tax areas seem to have a relationship to MLE-induced UI claims in the non-manufacturing sector.

We next estimate our spatial models. A state's RTW status is again excluded to help facilitate the estimate both a random and fixed effects specification. Table 6 presents the results obtained from our spatial models. Note that there are three effects per model. The direct effect provides the estimates of the relationship between a given state's characteristics, in this case both its overall and subcomponent EFNA scores, and the number of extended non-manufacturing UI claims observed within that state. The coefficient estimates presented in the indirect effect columns represent any spillover effects that a given state's characteristics may have in neighboring states. Recall that these are cumulative effects across all neighboring states and not a 'per neighbor' effect. Finally, the total effect is the total of both the direct and indirect effects of a given state's characteristics.

Table 4: Spatial Durbin Model Estimates of Extended Unemployment Insurance Claims in the Non-Manufacturing Sector(1996-2012)

	Direct Effect		Indirect Effect		Total Effect		Direct Effect		Indirect Effect		Total Effect	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Overall EFNA Score	-4.09 (2.99)	-5.80 (4.34)	-4.54 (7.50)	0.28 (9.50)	-8.63 (8.25)	-5.52 (10.44)						
Size of Government Score							1.07 (2.35)	2.71 (2.59)	1.02 (5.43)	0.44 (5.20)	2.08 (6.18)	3.15 (6.21)
Labor Market Freedom Score							6.33* (3.82)	5.40 (4.68)	-27.52*** (6.50)	-26.01*** (9.09)	-21.18*** (7.37)	-20.61** (10.03)
Takings and Discriminatory Taxation Score							-9.98*** (3.80)	-14.38*** (4.87)	11.06 (7.43)	20.88** (10.31)	1.08 (9.12)	6.49 (10.39)
Specification	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE
n = 595												
Hypotheses Tests												
	RE Model		FE Model		RE Model		FE Model					
ρ	ρ = 0.19 (p=0.000)		ρ=0.20 (p=0.000)		ρ=0.19 (p=0.000)		ρ=0.20 (p=0.000)					
SAR [$\lambda = \theta = 0$]	χ^2=0.70 (p=0.404)		χ^2=0.01 (p=0.919)		χ^2=15.93 (p=0.001)		χ^2=17.37 (p=0.001)					
SEM [$\rho = \theta = 0$]	χ^2=0.50 (p=0.478)		χ^2=0.05 (p=0.816)		χ^2=17.58 (p=0.001)		χ^2=15.11 (p=0.002)					
Hausman Test	χ^2=37.68 (p=0.158)				χ^2=4,022.24 (p=0.000)							

Notes: See notes to Table 2.

The results presented in the first column of Table 4 suggests that changes in a state's overall EFNA score do not seem to have much of a correlation with the number of MLE-induced UI claim rates outside of manufacturing. Further, differences in state business policies do not seem to have any effect on neighboring states (Column 3). The prior work found that EFNA

scores not only were associated with lower UI claim rates within the manufacturing sector for a given state, but also for neighboring states. The suggestion for the manufacturing sector was that states with better business climates saw lower claim rates themselves and helped to lower rates in neighboring states as workers moved to the better ranked states to take advantage of the better labor market opportunities. We fail to uncover a similar effect when we look outside of the manufacturing sector.

Our fixed effects estimation results correspond to those obtained using the random effects specification. Column 2 presents the coefficient estimate attached to EFNA scores for the states themselves after we take into consideration unobserved heterogeneity. It fails to achieve significance at conventional levels. Column 4 provides the associated effects on neighboring states and the coefficient estimate is again statistically no different from zero. In sum, when we proxy for a state's attractiveness as a place to do business using EFNA scores, we fail to uncover any evidence that there is a relationship with the number of non-manufacturing MLE-induced extended UI claims.

When we use the subcomponents of states' EFNA scores, we obtain a slightly different set of results. In this case, some elements of a state's business climate might be related to the number of UI claims in the non-manufacturing sector, but these results are sensitive to the specification. Column 7's results would suggest that states with more favorable labor market scores see an *increase* in their extended UI claims rate. This finding does dissipate when we adopt the fixed effects estimator (Column 8). For a state's takings and discriminatory tax score, we generally see that being better ranked on this aspect generally serves to reduces the number of UI claims within a state's non-manufacturing sector. This reduction is within the range of ten to fourteen fewer claims (per ten thousand workers).

When we look at the possible spill-over effects (Columns 9 and 10), we see that states which have more flexible labor markets seem to help reduce the number of UI claims in neighboring states, if not necessarily within their own state. This might suggest that the employment growth associated with better rankings pointed to in Kolko et al. is the result of new jobs being created and not necessarily a redistribution of existing jobs. While states cannot seemingly reduce the number of claims within their own states, it may be that the growth in (new) jobs might help to bring new workers into their states. This finding is generally consistent across both the random and fixed effects estimator.

There is a weak suggestion that states' tax policies have deleterious effects on neighboring states. States bordering better ranked states, at least as measured by the EFNA takings and discriminatory taxation scores, might see an increase in the number of UI claims within their non-manufacturing sectors (column 10).

The size of a state's government score, which we use as a proxy for how well a state can enforce its regulations, generally seems to have no relationship with MLE-induced UI claims both within a state and across its neighboring states.

Shifting next to the use of SBTC scores as a proxy for how attractive a state is to business in our spatial model, we again uncover few significant coefficient estimates. These findings are presented in Table 5. The first two columns would suggest that a state's overall SBTC score might be related to fewer UI claims within a state's non-manufacturing sector. The first column suggests that better ranked states see a significant reduction in the number of extended UI claims attributable to a MLE within the non-manufacturing sector. This finding is also obtained from the use of our fixed effects estimator (Column 2). There do not appear to be any spillover effects.

Table 5: Spatial Durbin Model Estimates of Extended Unemployment Insurance Claims in the Non-Manufacturing Sector (2003-2012)

	Direct Effect		Indirect Effect		Total Effect		Direct Effect		Indirect Effect		Total Effect	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Overall State Business Tax Climate Score	-8.50** (3.63)	-8.12* (4.58)	-2.55 (9.49)	-1.83 (10.05)	-11.04 (10.18)	-6.29 (11.76)						
Corporate Tax Score							6.84 (6.29)	8.62 (7.97)	-35.33** (15.79)	-13.55 (20.39)	-28.49 (17.94)	4.92 (23.48)
Property Tax Score							-6.57 (4.64)	-3.13 (5.46)	16.94* (8.79)	27.01** (11.21)	10.36 (9.97)	23.88 (12.95)
Sales Tax Score							0.50 (5.60)	-3.15 (6.73)	-14.82 (11.26)	-22.53 (13.75)	-14.32 (14.58)	19.38 (17.10)
Unemployment Insurance Score							0.95 (3.28)	1.90 (3.94)	3.55 (6.55)	3.02 (8.49)	4.51 (8.11)	4.92 (9.89)
Specification	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE
n = 360												

Hypotheses Tests

	RE Model	FE Model	RE Model	FE Model
ρ	ρ = 0.20 (p=0.003)	ρ=0.19 (p=0.005)	ρ=0.20 (p=0.003)	ρ=0.19 (p=0.004)
SAR [$\lambda = \theta = 0$]	χ^2=0.14 (p=0.705)	χ^2=0.09 (p=0.762)	χ^2=10.17 (p=0.038)	χ^2=9.30 (p=0.054)
SEM [$\rho = \theta = 0$]	χ^2=0.03 (p=0.858)	χ^2=0.02 (p=0.885)	χ^2=11.32 (p=0.023)	χ^2=8.64 (p=0.071)
Hausman Test	χ^2=30.74 (p=0.129)		χ^2=92.19 (p=0.000)	

Notes: See notes to Table

Unlike when EFNA scores are used, we generally find that the subcomponents used to construct the overall SBTC score do not seem to have any relationship to MLE-induced UI claim rates. The exception seems to be with regards to how a state's policy towards taxes might influence neighboring states (Columns 9 and 10). We generally see that states which neighbor better ranked states in terms of their property taxes see an increase in the number of MLE-induced UI claims. There is also a suggestion that states with better ranked neighbors in terms of corporate tax policy might see fewer UI claims themselves (Column 9). This effect dissipates, however, when we take unobserved heterogeneity into consideration with the results presented in Column 10.

Scores for either sales tax or their unemployment insurance tax policies seem to portend little in the way of UI claims within the non-manufacturing sector. This holds not just within a given state, but also for neighboring states. In general, we fail to find state business climates, as proxied by the Tax Foundations' SBTC index, is associated with a reduction in extended UI claims in non-manufacturing sectors.

Across the majority of our multivariate estimate results, we see little evidence that state policy differences are associated with differences in the UI claims rates for the non-manufacturing sectors. This stands in contrast with prior findings examining the manufacturing sector. This gives

us the suggestion that it is the manufacturing sector which seems to respond to policy differences. Outside of this sector, however, we fail to uncover much evidence that state public policy actions are related to the number claims they see being placed on their UI systems.

CONCLUDING COMMENTS

Results contained within the prior literature suggest that state business climates are correlated with a number of labor market outcomes within a state's manufacturing sector. This relationship, however, does not seem to hold for the non-manufacturing sector.

We fail to find much evidence that a state's overall EFNA score has a relationship with the number of UI claims being made outside of its manufacturing sector. The same holds when we use the SBTC scores as our proxy for a state's business climate.

Where we see some points of contact between the two sectors and their relationship with business climates comes in our spatial models and some of the EFNA subcomponent scores. These results seem limited to a state's labor market flexibility and taxation policies. In the case of its labor market regulations, it would appear only neighboring states see much benefit. This may be because the employment growth pointed to in the prior work is the result of new job creation and workers from surrounding states are moving to take advantage of the better labor market opportunities. Taxation policy differences, at least within the EFNA scores, seem to hold benefits only for the adopting states themselves.

Finally, the balance of the spatial results obtained when we use the SBTC Index seems to suggest that the non-manufacturing sector is much less responsive to policy differences than is the manufacturing sector. We find very limited evidence that scoring better on this index holds much in the way of a relationship with extended UI claims being made. Scoring better on the overall SBTC score does seem to have a relationship with the number of non-manufacturing job losses within a state, while the subcomponents portend that it is neighboring states which are affected by policy changes.

References

Autor, David H. "Outsourcing at Will: The Contribution of Unjust Dismissal to the Growth of Employment Outsourcing." *Journal of Labor Economics,* 21(1), 2003, 1-42.

Autor, David H., Adriana R. Kugler, and William D. Kerr. "Do Employment Protections Reduce Productivity? Evidence from U. S. States." *The Economic Journal,* 117(521), 2007, 189-217.

Bartik, Timothy J. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1991.

Belotti, Federico, Gordon Hughes and Andrea Piano Mortari. "XSMLE – A Command to Estimate Spatial Panel Models in Stata." German Users Group Meeting, 2013.

Bologna, Jamie, Donald J. Lacombe and Andrew T. Young. "A Spatial Analysis of Incomes and Institutional Quality: Evidence from US Metropolitan Areas." Department of Economics, West Virginia University Working Paper 14-35, 2014.

Brown, Sharon P. and Lewis B. Siegel. "Mass Layoff Data Indicate Outsourcing and Offshoring Work." *Monthly Labor Review,* 128(8), 2005, 3-10.

Drenkard, Scott and Joseph Henchman. "2013 State Business Tax Climate." The Tax Foundation Background Paper #64, 2012.

Elhorst, J. Paul. “Applied Spatial Econometrics: Raising the Bar.” *Spatial Economic Analysis,* 5(1), 2010, 9-28.

Evans, David S. “The Relationship Between Firm Growth, Size and Age: Estimates from 100 Manufacturing Industries.” *Journal of Industrial Economics,* 35(4), 1987, 567-81.

Fisher, Peter. *Grading Places*. Washington, D. C.: Economic Policy Institute, 2005.

Greenhut, Melvin L. Plant Location in Theory and in Practice: The Economics of Space. Chapel Hill, NC: University of North Carolina Press. 1956.

Heller, Lauren R. and E. Frank Stephenson. "Economic Freedom and Labor Market Conditions: Evidence From the States." *Contemporary Economic Policy*, 32(1), 2014, 56-66.

Holmes, Thomas. “The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders.” *Journal of Political Economy,* 106(4), 1998, 667-705.

Kletzer, Lori G. “Job Displacement.” *Journal of Economic Perspectives,* 12(1), 1998, 115-36.

Kolko, Jed, David Neumark and Marisol Cuellar Mejia. “What Do Business Climate Indexes Teach Us About State Policy and Economic Growth?” *Journal of Regional Science,* 53(2), 2013, 220-55.

LeSage, James P. and Manfred M. Fischer. “Spatial Growth Regressions: Model Specification, Estimation and Interpretation.” *Spatial Economic Analysis,* 3(3), 2008, 275-304.

LeSage, James P. and Robert Kelly Pace. Introduction to Spatial Econometrics. Boca Raton, FL: Taylor and Francis CRC Press. 2008.

Meyer, Bruce D. “Unemployment Insurance and Unemployment Spells.” *Econometrica,* 58(4), 1990, 757-82.

Neumark, David and Jennifer Muz. "The "Business Climate" and Economic Inequality." National Bureau of Economic Research Working Paper No. 20260, 2014.

Stansel, Dean, Jose' Torra and Fred McMahan. Economic Freedom of North America. Vancouver: The Fraser Institute. 2014.

Surfield, Christopher J. “Government Mandates and Atypical Work: An Investigation of Right-to-Work States.” *Eastern Economic Journal,* 40(1), 2014, 26-55.

Surfield, Christopher J. and C. Surender Reddy. “Mass Layoff, Manufacturing and State Business Climates: Does State Policy Matter?” *Contemporary Economic Policy*, 34(4), 2016, 630-4

DECISION SCIENCES INSTITUTE

Intelligent Workforce Scheduling with Seasonal ARIMA Model and Column Generation: A Case of Airport Ground Staff Scheduling

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Pan Zou
Georgia Institute of Technology
Email: zoupzp123@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Wenchong Chen
Georgia Institute of Technology
Email: wchen630@gatech.edu

Stefan Quaadgras
Georgia Institute of Technology
Email: stefanquaadgras@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

Workforce scheduling is a process to estimate the staffing demands and design staff work schedules to meet current and future workforce demands. This paper proposes an intelligent workforce scheduling system to assist industries with workforce management. Workforce scheduling is divided into three stages: demand forecasting, shift planning, and rostering. The seasonal ARIMA model is applied for demand forecasting. A set covering model is formulated for shift planning. The column generation is applied as to deal with combinatorial explosion of rostering. Finally, a case of airport ground staff scheduling is presented to illustrate the feasibility and potential of the proposed system.

KEYWORDS: Workforce scheduling, Mixed integer programming, Column generation, Time series analysis

INTRODUCTION

For service industries, workforce scheduling is an important technique to help companies manage employees and tasks. It involves the forecast of future workforce demands, shift planning for each

day in the future, and rostering that connects shifts to form different shift lines based on local labor laws and employment rules and to cover the shift demands (Lin et al., 2015). Well-designed workforce scheduling can facilitate the employee utilization rate, reduce the labor cost and the operation cost, and improve service satisfaction. Therefore, there are growing demands for an automatic and efficient workforce scheduling system to assist large-scale companies with workforce management, especially when there are a large number of employees to manage.

Developing mathematical models and algorithms for workforce scheduling is an automatic planning approach and requires less domain knowledge than manual planning, and its solution can serve as a reference for practitioners. However, three technical challenges exist in the implementation. The first challenge is to accurately forecast workforce demands. For the work without clearly scheduled tasks, like nursing and airport security check, future workforce demands should be estimated for further scheduling problems. The second challenge is the modeling of the trade-off between understaffing and overstaffing during shift planning. And the third challenge is the combinatorial explosion during rostering, in which a subset of feasible roster lines should be selected in exponentially growing sets to cover the shift demands.

To overcome these challenges, this paper proposes an intelligent workforce scheduling system. The seasonal autoregressive integrated moving average (ARIMA) model is developed for demand forecasting. A set covering model is formulated with the consideration of understaffing and overstaffing factors for shift planning. And column generation is applied in the branch and price framework to solve rostering problems. The rest of this paper is organized as follows. The literature review discusses the background of workforce scheduling, including demand forecasting, shift planning, and rostering. System analysis for intelligent workforce scheduling is presented. Demand forecasting with the seasonal ARIMA model is introduced. Problem definition and formulation of shift planning and rostering are discussed. The column generation approach is discussed as the solution to combinatorial explosion. Finally, a case of airport ground staff scheduling is reported to illustrate the feasibility and effectiveness of the proposed system.

LITERATURE REVIEW

Workforce scheduling is a challenging problem faced by various service industries like airports, hospitals, and toll collection (Herbers & Hromkovic, 2005). It has received large attention since the integer programming techniques are applied to real-world problems. Multiple subproblems are included in this field: demand forecasting, shift planning, and rostering. Much research is done on these problems.

Demand forecasting is to predict the future workforce demands based on historical data and is the first step of workforce scheduling. There are both linear and nonlinear methods for such quantitative demand forecasting problem, but not a single model is ideal for all situations (Petropoulos et al., 2014). Among these methods, the Holt-Winters method and the seasonal ARIMA model are two commonly used models. The former expresses the trend and seasonality component in the model and is appropriate for forecasting seasonal data (Holt, 2004). The ARIMA model is the combination of the autoregressive model and the moving average model (Ho & Xie, 1998). It can accurately predict future data when the data is stationary and directly related to the timestamps (Ediger & Akar, 2007).

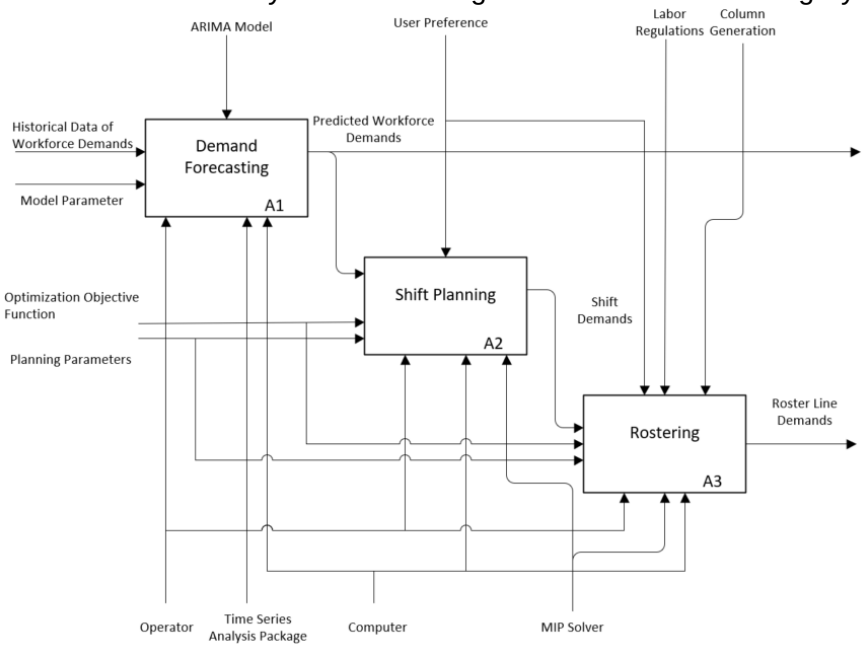
Shift planning is to design a set of shifts with different start time and duration to cover the workforce demands within the planning horizon (Gong et al., 2019). Most recent studies emphasize how to construct effective shifts. Variations of these problems are found to be NP-hard and NP-complete (Ernst et al., 2004). The shift planning problem is mathematically

formulated as a set covering problem, solved by summing a sequence of shifts from the shift set based on the shift design requirement (Musliu et al., 2004). Because of the advance in information management software systems, it requires smaller time granularity, larger personnel scale, and longer planning horizon (Van den Bergh et al., 2013). Meanwhile, from the modern management perspective, designed shifts are desired to be length and placement adjustable with specific break rules (Bechtold et al., 1990). Thus, more efficient shift planning algorithms need to be developed to cope with the increasing computational complexity. In accordance with Danzig's set covering formulations, the mixed integer programming (MIP) methods have been widely applied to solve the shift planning problems (Herbers & Hromkovic, 2005, Di Gaspero et al., 2007). Metaheuristics are suggested to be critical methods for coping with combinatorial optimization problems by incorporating modular rules to implement shift planning efficiently (Clausen, 2010, Lagodimos, 2000).

The research on rostering is widely studied for the critical meaning of the solution quality. A detailed review of recent research on personnel scheduling and rostering is presented (Ernst et al., 2004). Among various service industries, the nurse rostering problem (NRP) is a typical problem to study for the multi-qualification and multi-time-period shift type characteristics. Overviews of the models and methodologies are presented (Cheang et al., 2003, Burke et al., 2004). Rostering with work patterns is also a commonly studied problem. For example, a two-phase algorithm is proposed to address the rostering with a 5&2 work pattern (Alfares, 1998). Another typical example is rostering with the 14&7 work pattern considering cyclic weekly demands (Alfares, 2002).

SYSTEM ANALYSIS OF INTELLIGENT WORKFORCE SCHEDULING

Figure 1: Functional Analysis of the Intelligent Workforce Scheduling System



The proposed intelligent workforce scheduling system implements demand forecasting, automatic shift design, and rostering with time series analysis and deterministic optimization techniques. The system functional analysis is shown in Figure 1. Firstly, historical data of workforce demands are collected and used to train the seasonal ARIMA model for the prediction of workforce

demands in the time of the planning horizon. Then the predictions are used for shift planning. During this step, the planning preferences, like the shift duration, the shift starting time window, as well as the preference towards understaffing and overstaffing, are set by the user. The shift planning problem is modeled as a set covering problem and solved with the MIP solver. The calculated shift demands are utilized as the inputs of rostering. Labor regulations are introduced as the constraints of shift connection to form a feasible roster line. Column generation is applied to solve the rostering problem, and the roster lines are obtained as the results for shift demand covering, which are also the solution to this workforce scheduling problem.

SEASONAL ARIMA MODEL-BASED DEMAND FORECASTING

Demand forecasting is the initial step of workforce scheduling. In this study, the ARIMA model is selected, which has two variations: the non-seasonal ARIMA model and the seasonal ARIMA model. In time series analysis, seasonality is defined as a time period over which a pattern of changes will repeat. Seasonality is a typical characteristic of workforce demand data in service industries. Thus, the forecasting model is developed based on the seasonal ARIMA model.

The non-ARIMA model consists of three parts: the autoregressive model, the integrated model, and the moving average model. The autoregressive model AR(p) implements forecasting with a linear combination of past values or variables. The integrated model I(d) refers to doing d-order differencing to make the time series data stationary. And the moving average model MA(q) indicates the regression error is a linear combination of error terms at various time points in the past. The ARIMA model ARIMA(p,d,q) can be formulated as below:

$$\phi(B)\,(1-B)\,(x_t-\mu)=\theta(B)\,e_t \quad (1)$$

$$\phi(B)=1-\phi_1 B-\ldots-\phi_p B^p \quad (2)$$

$$\theta(B)=1+\theta_1 B+\ldots+\theta_q B^q \quad (3)$$

After introducing the seasonality, the seasonal ARIMA model ARIMA(p,d,q)×(P,D,Q)S is formulated as below:

$$\Phi(B^S)\,\phi(B)\,(1-B)\,(1-B^S)\,(x_t-\mu)=\Theta(B^S)\,\theta(B)\,e_t \quad (4)$$

$$\Phi(B^S)=1-\Phi_1 B^S-\ldots-\Phi_p B^{PS} \quad (5)$$

$$\Theta(B^S)=1+\Theta_1 B^S+\ldots+\Theta_q B^{QS} \quad (6)$$

LARGE-SCALE SHIFT PLANNING WITH A SET COVERING MODEL FORMULATION

Shift planning aims to design a set of shifts each day with different start time and duration to cover the workforce demands within the planning horizon. The solution to the shift planning problem is to explore the combination of shifts and to cover the demands based on trade-off rules while avoiding the violation of constraints from the regulatory and operational perspectives (Dantzig, 1954). The objective should be set to minimize the operation cost while avoiding insufficient and superfluous workforce.

The formulation of optimal shift design problems starts with the modeling of the timeline. The time of a planning horizon is represented by N, whose unit is a day. This planning period N is discretized as T equally sized time slots, t=1,2,…,T. The length of each time slot is named as time granularity a_g, whose unit is minute. The workforce demand is represented in a 2D matrix $D\in N^{(|\Phi|\times T)}$, where the row φ represents different types of demands, while the column is the timeline

in the planning period. The element of the demand matrix is natural numbers and describes the number of the required human resource units. The total amount of the required category is defined by $|\Phi|$. The shift is represented by s, while all s is collected in the shift set $\boldsymbol{S}$. The shifts are distinguished by the start time p_s and end time q_s, so a shift is in the time interval $[p_s,q_s]$.

The mixed-integer programming model of the shift planning can be formulated as follows:

$$z^*=\min \alpha_c\sum_{s\in S} c_s x_s+\alpha_u \sum_{t=1}^{T}\sum_{\varphi\in\Phi} u_{\varphi t} +\alpha_o\sum_{t=1}^{T}\sum_{\varphi\in\Phi} o_{\varphi t} \quad (7)$$

$$\text{s.t. } \sum_{s\in S} a_{st} x_s m_{s\varphi}+\sum_{\varphi\in\Phi} u_{\varphi t} -\sum_{\varphi\in\Phi} o_{\varphi t}=D_{\varphi t} \quad \forall\varphi,\ 1\le t\le T \quad (8)$$

$$\sum_{s\in S} v_{is} x_s \le V_i,\ \forall V_i\in V \quad (9)$$

$$u_{\varphi t}, o_{\varphi t}, x_s\in N \quad (10)$$

$$a_{st}\in\{0,1\} \quad (11)$$

The objective function (7) is a weighted sum of shift operation cost, understaffing cost, and overstaffing cost, using weight factor α_c, α_u, and α_o, respectively. The shift operation cost is the sum of the cost of the specific shift c_s multiplies the number of this shift x_s. The overstaffing cost and understaffing cost are calculated by summing up the number of workforce deficiency and workforce excess of all demands along the timeline, respectively. Constraint (8) describes that the staff equilibrium of all demands must be satisfied at any time along the timeline. This staff equilibrium requires the sum of available staff and workforce deficiency must be equal to demands after subtracting the workforce excess. The operational variable $a_{st}=1$ describes shift s is active at time t and $a_{st}=0$ describes shift s is inactive at time t. Similarly, operational variable $m_{s\varphi}=1$ describes shift s can fulfill demand φ, while $m_{s\varphi}=0$ describes shift s cannot fulfill demand φ. Constraint (9) is the number constraints of the specified shift category, where $v_{is}=1$ means shift s belongs to the category i. Shift category limit V_i specifies the maximum number of a particular shift category, and all V_i are collected in V. Constraint (10) states that the $u_{\varphi t}$, $o_{\varphi t}$, and x_s are nonnegative integers. The above-formulated MIP problem can then be solved by the integer programming solver.

ROSTERING WITH SPECIFIC WORK PATTERNS USING COLUMN GENERATION

Rostering aims at selecting lines of specified shifts with day-off information based on the labor law and employment rules to cover the shift demands obtained from shift planning. The objective is to minimize the operation cost while satisfying the demands for different shifts on the planning horizon. Because rostering is to form a roster line with different combinations of shifts, in the result, one more shift than needed can bring large workforce excess. Thus, the quality of rostering operations has a great impact on workforce scheduling results.

The rostering problem can be formulated as a set covering problem, which is solved by forming rostering lines using the shifts existing in the shift planning result to cover the shift demands while satisfying specific work patterns and avoiding violating the labor law and the employment rules (Lin et al., 2015). Some definitions are given in Table 1 to better illustrate the rostering problem.

Table 1: Rostering Term Explanation	
Term	Explanation
Shift demand	Number of different required shifts from the shift planning results
Work pattern	The minimum period that specifies the sequence of the days of work and required days of rest, sometimes certain shift types are specified on some workday, which is often seen in the nurse scheduling problem. (For example, M-A-O means working on a morning shift on the first day, an afternoon shift on the second, and off on the third day.)
Roster line	A line of specified shifts on the planning horizon
Roster	A set of roster lines for needed employees on the planning horizon
Rostering	The process of selecting roster lines to cover the shift demands based on the labor law and employment rules

The challenge of the rostering problem lies in two aspects. Firstly, to find feasible roster lines with specified work patterns while satisfying constraints of labor law and employment rules can be formulated as a shortest path problem with resource constraints, which is NP-hard in the strong sense (Boland et al., 2006). Secondly, as the planning horizon grows linearly, the number of feasible roster lines grows exponentially, making it impractical to enumerate every combination to solve it directly, especially when the number of everyday shifts is large. To overcome the challenge, the branch and price algorithm using column generation is often applied.

The integer programming model for generalized rostering problems can be formulated as a set covering model described in the following:

$$\min \sum_{r \in R} c_r x_r \tag{12}$$

$$\text{s.t. } \sum_{r \in R} a_{sr} x_r \geq q_s, \ \forall s \in S \tag{13}$$

$$\sum_{r \in R} a_{gr} x_r \leq m_g, \ \forall g \in G \tag{14}$$

$$\sum_{r \in R} x_r \leq n \tag{15}$$

$$x_r \in N \tag{16}$$

$$a_{sr}, a_{gr} \in \{0,1\} \tag{17}$$

Since the rostering process is after shift planning and companies usually have a high understaffing cost while a low labor cost, the objective function (12) is set to be the sum of labor cost. The shift planning result is identified by the shift set S that contains every shift used s. The roster line set is represented by R, and each roster line r in the set is a $|S|\times 1$ binary matrix, in which each element suggests whether the corresponding shift is chosen in that roster line. The labor cost of a specific roster line is represented by c_r, which is calculated according to the total working hours of that roster line. The shift demand requirement is described in Constraint (13). a_{sr} indicates whether a shift s is chosen in a roster line r, with $a_{sr}=1$ meaning the shift belongs to this roster line and 0 meaning the shift is not chosen. x_r refers to the number of a roster line in the solution. The corresponding requirement for the number of a shift is q_s, $\forall q_s \in Q_s$. The limitation on the work pattern groups is represented in Constraint (14). A pattern group g is one variation of the specified work pattern, and $\forall g \in G$. a_{gr} indicates the pattern group that the roster line belongs to. The maximum number of a pattern group is represented by m_g. This constraint is often used to better manage employees working on the same work pattern. For example, a roster line M-O-M belongs to a pattern group W-O-W (work-off-work) and a_{gr} equals the number that represents the pattern group, and m_g is the maximum number of roster lines on W-O-W. Constraint (15)

indicates the limitation on the number of employees, and Constraint (16) guarantees the integrality of the solution.

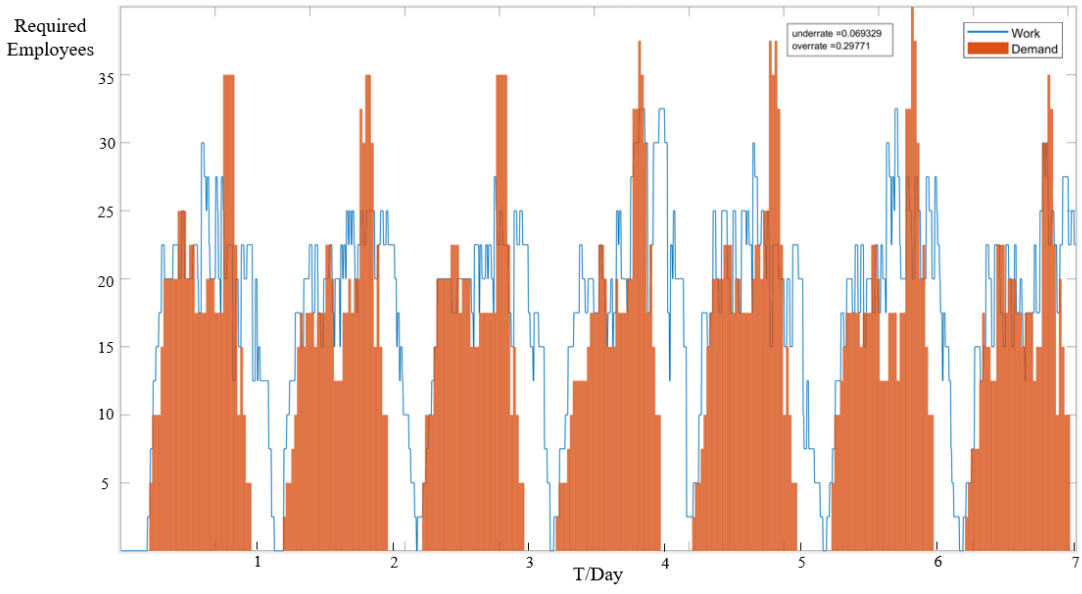
For such combinatorial optimization problems, it is nearly impossible to explicitly enumerate all feasible combinations. These problems are often solved by the column generation approach, which is embedded into the branch and bound algorithm. The rationale of the column generation technique is similar to the simplex method. Because the non-basic variables of the solution are in the majority, only the variables that have the potential to make the solution performance better should be considered. Thus, this approach decomposes the problem into two related optimization problems: the master problem that measures the performance of the current solution and the pricing problem (or the subproblem) that continuously generates high-quality and feasible combinations for the master problem. In this study, the master problem is formulated as in Equation (12) – (17), while the pricing problem is formulated as a shortest path problem with resource constraints, and it is solved by the label correcting algorithm based on the concept of dynamic programming.

A CASE OF AIRPORT GROUND STAFF SCHEDULING

To illustrate the feasibility and effectiveness of the proposed system, a case based on a real-world workforce scheduling problem at a large airport is reported. Because the future flight information is given, demand forecasting does not require the use of the seasonal ARIMA model but can be calculated based on the flight tasks within the planning horizon. The case is implemented by MATLAB 2019a in Win 10 pro operation system. The CPU is Intel Xeon CPU E3-1505M v5 @ 2.90GHz.

To validate the shift planning model, a 7-day shift planning problem is solved. The result of the calculated workforce versus workforce demands is shown in Figure 2. The result shows that the combined shifts cover most of the demands: the understaffing rate is 6.93%, while the overstaffing rate is 29.77%. The results are acceptable according to the industry requirements.

Figure 2: The result of Shift Planning - Calculated Workforce versus Workforce Demands



The other experiment is conducted to validate column generation-based rostering. The objective is to generate a 7-day rostering, and the input shifts are shown in Table 2, where the shift date, start time, duration (represented in minutes), and the required number are given for each type of the shift. The work pattern is set as W-W-W-W-W-O-O, and the interval between two shifts in one roster line must be larger than 10 hours. The minimum weekly working hour is 40, and the maximum is 45. The rostering result is shown in Table 3. Figure 3 shows the shift coverage on each day of the rostering solution: all shifts are covered, and there are three superfluous shifts.

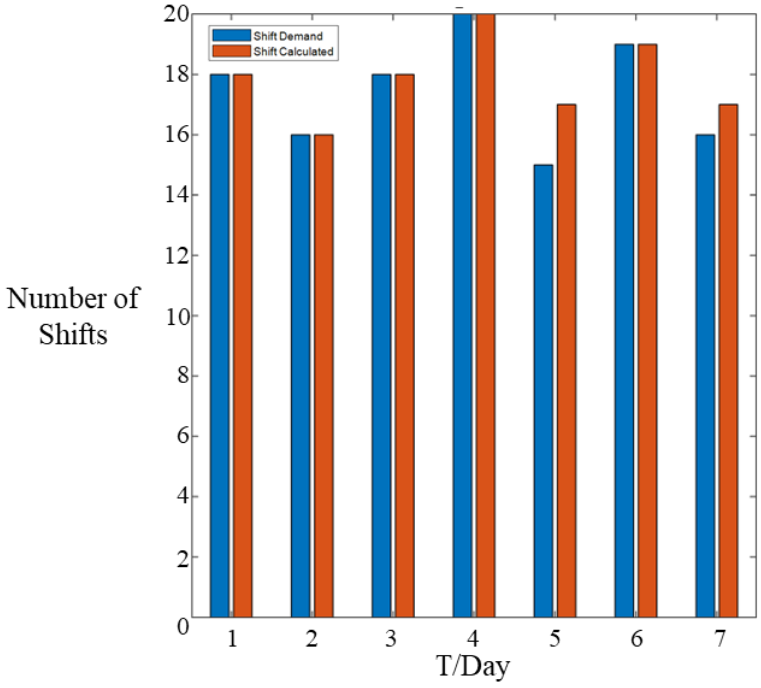
Table 2: Inputs from shift planning

Day	Start (min)	Duration (min)	Num	Day	Start (min)	Duration (min)	Num	Day	Start (min)	Duration (min)	Num
1	210	510	1	3	450	480	2	5	900	480	2
1	240	480	2	3	810	480	3	5	930	480	1
1	330	480	1	3	840	480	1	6	240	480	2
1	360	480	3	3	870	480	2	6	330	480	1
1	390	480	1	3	930	480	2	6	360	480	3
1	450	480	1	3	1050	480	1	6	420	480	2
1	810	480	2	4	240	480	1	6	450	480	1
1	840	480	4	4	330	480	2	6	810	480	4
1	870	480	1	4	390	480	5	6	900	480	3
1	900	480	2	4	420	480	2	6	930	480	2
2	210	600	1	4	810	480	1	6	1050	480	1
2	240	600	1	4	840	480	3	7	240	510	1
2	330	480	1	4	870	480	4	7	330	510	2
2	360	480	1	4	1020	480	2	7	360	510	1
2	390	480	1	5	240	540	1	7	450	480	2
2	450	480	2	5	240	600	1	7	480	480	1
2	810	480	2	5	330	510	1	7	510	510	1
2	870	480	4	5	360	480	1	7	870	480	2
2	930	480	3	5	390	480	1	7	900	480	1
3	210	510	1	5	420	480	1	7	930	480	2
3	240	480	1	5	450	480	1	7	1050	510	3
3	330	510	2	5	810	480	3				
3	360	480	3	5	870	480	2				

Table 3: The Rostering Result

Day	Start (min)	Duration (min)	Day	Start (min)	Duration (min)	Day	Start (min)	Duration (min)	Day	Start (min)	Duration (min)	Day	Start (min)	Duration (min)
1	240	480	4	390	480	5	390	480	6	900	480	7	1050	510
1	840	480	4	390	480	5	810	480	6	900	480	7	900	480
1	900	480	4	1020	480	5	810	480	6	810	480	7	900	480
1	330	480	2	330	480	3	330	510	4	840	480	7	1050	510
1	360	480	2	240	600	3	360	480	4	870	480	5	810	480
2	210	600	3	360	480	4	240	480	5	330	510	6	240	480
3	360	480	4	330	480	5	360	480	6	360	480	7	450	480
1	810	480	2	930	480	3	930	480	4	870	480	7	930	480
3	810	480	4	870	480	5	870	480	6	810	480	7	1050	510
1	810	480	2	870	480	3	870	480	6	240	480	7	360	510
1	360	480	4	390	480	5	930	480	6	810	480	7	510	510
2	870	480	3	840	480	4	840	480	5	900	480	6	900	480
2	810	480	3	450	480	4	810	480	5	450	480	6	450	480
2	390	480	3	330	510	4	390	480	5	420	480	6	930	480
1	210	510	2	450	480	3	450	480	6	360	480	7	330	510
1	870	480	2	930	480	3	810	480	4	1020	480	7	450	480
1	840	480	2	870	480	3	1050	480	6	420	480	7	930	480
1	840	480	4	330	480	5	450	480	6	930	480	7	870	480
1	840	480	4	420	480	5	240	600	6	420	480	7	240	510
1	360	480	2	810	480	3	810	480	6	360	480	7	480	480
1	900	480	2	870	480	3	870	480	4	840	480	5	900	480
2	930	480	3	930	480	4	870	480	5	870	480	6	1050	480
1	390	480	4	390	480	5	360	480	6	330	480	7	330	510
1	240	480	2	360	480	3	210	510	4	420	480	5	240	540
1	450	480	2	450	480	3	240	480	6	810	480	7	870	480

Figure 3: The shift coverage of a 7-day Rostering with a 5&2 Work Pattern



Two performance measures are used to evaluate the rostering algorithm performance: the computing time and the number of superfluous shifts. In the experiment above, the computing time is 28.48 seconds, in which most of the time is used on column generation. The required number of shifts to fully cover the demands from the solution is 125, which is 2.5% more than the actual required number. The result is acceptable for the industry application.

CONCLUSIONS

This paper proposes an intelligent workforce scheduling system to assist modern service industries with workforce management. The seasonal ARIMA model is used to predict workforce demands using historical data. Shift planning is modeled as a set covering problem with understaffing and overstaffing components to help the trade-off between labor costs and service satisfaction. The column generation is applied in rostering to address combinatorial explosion and improve the computing efficiency. A real-world case of airport ground staff scheduling is reported to illustrate the feasibility and effectiveness of the proposed system. Future work of this research is to develop a more adaptive model for demand forecasting and to accelerate the label correcting algorithm for faster column generation.

REFERENCES

Alfares, H. K. (1998). An efficient two-phase algorithm for cyclic days-off scheduling. Computers & Operations Research, 25(11), 913-923.

Alfares, H. K. (2002). Optimum workforce scheduling under the (14, 21) days-off timetable. Advances in Decision Sciences, 6(3), 191-199.

Bechtold, S. E., & Jacobs, L. W. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. Management Science, 36(11), 1339-1351.

Burke, E. K., De Causmaecker, P., Berghe, G. V., & Van Landeghem, H. (2004). The state of the art of nurse rostering. Journal of scheduling, 7(6), 441-499.

Cheang, B., Li, H., Lim, A., & Rodrigues, B. (2003). Nurse rostering problems—a bibliographic survey. European Journal of Operational Research, 151(3), 447-460. Clausen, T. (2010). A Rule-Based Local Search Algorithm for General Shift Design Problems in Airport Ground Handling.

Dantzig, G. B. (1954). Letter to the editor—A comment on Edie's “Traffic delays at toll booths”. Journal of the Operations Research Society of America, 2(3), 339-341.

Di Gaspero, L., Gärtner, J., Kortsarz, G., Musliu, N., Schaerf, A., & Slany, W. (2007). The minimum shift design problem. Annals of operations research, 155(1), 79-105.

Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. Energy policy, 35(3), 1701-1708.

Ernst, A. T., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. European journal of operational research, 153(1), 3-27.

Ernst, A. T., Jiang, H., Krishnamoorthy, M., Owens, B., & Sier, D. (2004). An annotated bibliography of personnel scheduling and rostering. Annals of Operations Research, 127(1-4), 21-144.

Gong, X., Wang, S., & Jiao, R. (2019, December). An Efficient 2D Genetic Algorithm for Optimal Shift Planning Considering Daily-Wise Shift Formats: A Case of Airport Ground Staff Scheduling. In 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1440-1444). IEEE.

Herbers, J., & Hromkovic, J. (2005). Models and algorithms for ground staff scheduling on airports (No. RWTH-CONV-121335). Fakultät für Mathematik, Informatik und Naturwissenschaften.

Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. Computers & industrial engineering, 35(1-2), 213-216.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, 20(1), 5-10.

Lagodimos, A. G., & Leopoulos, V. (2000). Greedy heuristic algorithms for manpower shift planning. International Journal of Production Economics, 68(1), 95-106.

Lin, D., Xin, Z., & Huang, Y. (2015, December). Ground crew rostering for the airport check-in counter. In 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1462-1466). IEEE.

Musliu, N., Schaerf, A., & Slany, W. (2004). Local search for shift design. European journal of operational research, 153(1), 51-64.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for Courses' in demand forecasting. European Journal of Operational Research, 237(1), 152-163.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. (2013). Personnel scheduling: A literature review. European journal of operational research, 226(3), 367-385.

DECISION SCIENCES INSTITUTE

Automated Predictive Accident Alert System for Vulnerable Road Users

Dothang Truong
Embry-Riddle Aeronautical University, Daytona Beach, Florida
truongd@erau.edu

ABSTRACT

Vulnerable road users (VRUs) are most at risk in traffic as they sustain a higher risk of injury in any collision against a vehicle. As the number of VRU fatalities continues to rise, it is critical to have a predictive tool to mitigate and minimize the injuries for VRUs. The paper proposes a method and framework to develop and test an automated predictive accident alert system (APAAS) for VRUs, including pedestrians and cyclists, using the Artificial Neural Network algorithm. This system can automatically predict and detect a crash accident involving VRUs before it happens.

KEYWORDS: vulnerable road users, automate predictive accident alert system, artificial neural network, deep learning

INTRODUCTION

Vulnerable road users (VRUs) refers to "those unprotected by an outside shield, namely pedestrians and two-wheelers, as they sustain a greater risk of injury in any collision against a vehicle and are therefore highly in need of protection against such collisions." Among these VRUs, pedestrians and cyclists are "most unlikely to inflict injury on any other road user, while motorized two-wheelers, with heavier machines and higher speeds, may present a danger to others" (OECD, 1998). As the traffic demand continues to grow, the number of crash accidents involving VRUs increases. Figures 1 shows the crash fatality statistics for pedestrians and cyclists, respectively, in the US from 1994 to 2017. While the number is fluctuating, it appears that the fatalities have been increasing since 2009 and reached a peak number in 2016. In 2016, 6,080 pedestrians and 852 cyclists were killed in crash accidents, respectively. The statistics shows concerning safety issues to VRUs as they transport on the road.

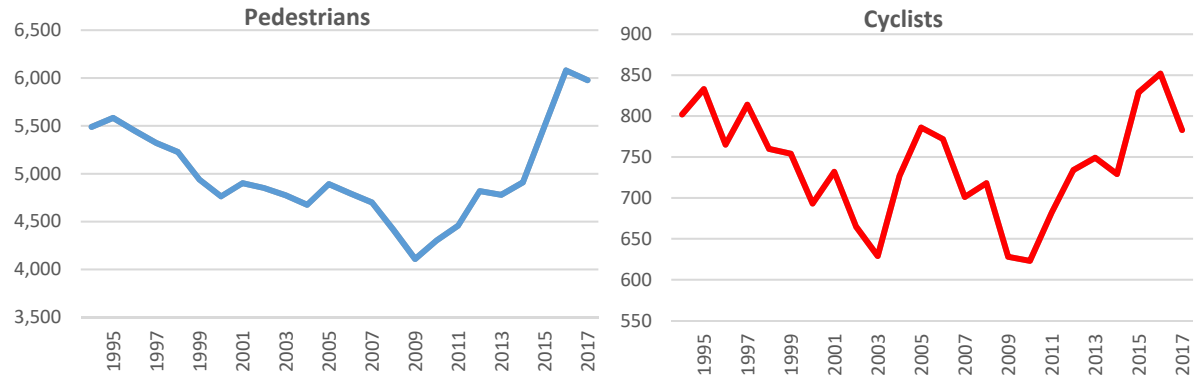


Figure 1: Crash statistics for VRUs
Source: NHTSA Fatality Analysis Reporting System

When a collision occurs between a vehicle and a VRU (pedestrian or cyclist), the VRU tends to sustain a higher severity of the injury and needs to be attended immediately by the Emergency Medical Technician (EMT). The VRU's likelihood to survive the crash will be higher if the EMT can arrive at the crash site in a timely manner (Nishimoto et al., 2019). However, emergency centers usually receive a very high number of calls for crash accidents in the area with varieties of crash types at different levels of injuries. Given the limited EMT resources and uncertainty with crash accidents, it is very challenging for the EMT to get to a VRU-related crash site in a short amount of time. If provided with a prediction of potential crashes involving VRUs, along with the location and the injury severity, the emergency center can be ready in advance and have the EMT available to go to the crash site promptly (Large et al., 2019; Lee et al., 2017). Such prediction can be provided by a predictive system that automatically predicts the possibility of a VRU-related crash incident and location. This system proves to be beneficial to the authorities in mitigating and minimizing the fatality risks to VRUs. However, such a system has yet to be developed.

Several research studies in VRU safety seem to focus on determining causes for accidents and injuries for VRUs (Constant and Lagarde, 2010; Otte et al., 2012). Additionally, other studies developed risk assessment models that aim at determining impact factors and how they affect the injury severity for VRUs (Mohamed et al., 2013; Ghomi et al., 2016). These studies are retrospective, so they tend to be more reactive than predictive. Recently, a rising trend in transportation research is centered on developing automated detection systems through image recognition at a specific location, such as sidewalks, corners, intersections, and railroads (Hartmann and Watzenig, 2019). These systems integrated deep learning to artificial intelligence applications, and they mainly focus on providing a tool to automatically warn the vehicle drivers of pedestrians in the area of interest or warn the railroad authority of a trespasser in the railroad area (Zaman et al., 2019). While these systems are very useful to avoid collisions or accidents involving VRUs in a specific area of interest, they do not provide an automated predictive warning to the emergency center at a larger scale, such as a county or state. These tools are still in the testing stage and are not available for all vehicles. Additionally, it also depends on how the driver takes the warning and react to avoid the collision. Accidents can still happen, and VRUs are still likely to be severely injured. An automated predictive system is critical to the safety of VRUs as it can predict and alert the state emergency centers of possible VRU-related crashes and locations. Such a system would allow the authorities to be proactive in mitigating and minimizing the fatality risks to VRUs.

RESEARCH OBJECTIVES

The purpose of this paper is to propose a method and framework to develop and test an automated predictive accident alert system (APAAS) for vulnerable road users (VRUs), including pedestrians and cyclists. Specifically, this research has two primary objectives

1) Develop and test a prediction model using Artificial Neural Network, or deep learning, to predict the risk of accidents involving VRUs (pedestrians and cyclists) in US states using NHTSA FARS crash data
2) Develop and test an automated predictive accident alert system (APAAS), using the predictive algorithm in step 1, to automatically detect the location of the accident, probability of the accident occurrence, and injury level for VRUs

The APAAS can predict and detect an accident involving VRUs before it happens (including location, occurrence probability, and injury level) and allows state authorities to take mitigation actions to prevent or minimize the injuries for VRUs. The system will be developed and tested for three US states, North Carolina, Virginia, and Florida. Figure 2 shows an example demonstrating how the APAAS system would function to predict an accident involving VRUs before it happens. A prediction of the VRU-related accident at a particular time in the future, along with the location and injury level, would allow state authorities to mitigate proactively and to minimize the fatality risk to the VRUs.

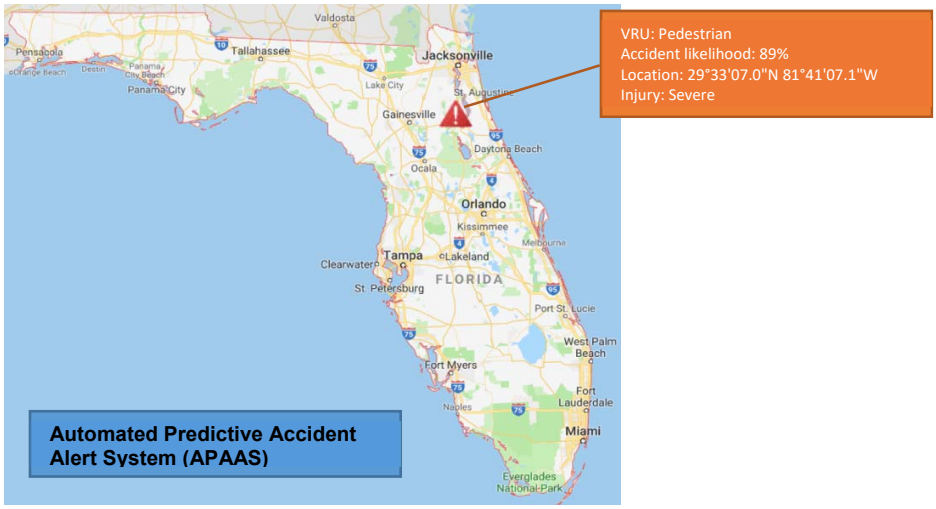


Figure 2: Example of an APAAS system (Source: Google Map)

RESEARCH METHODS

In order to meet the objectives, it is proposed to use Artificial Neural Network (ANN), a deep learning method, to develop and train a predictive model for accidents involves VRUs using Fatality Analysis Reporting System (FARS) data. Then, the tested ANN algorithm will be used to develop and test an Automated Predictive Accident Alert System (APAAS) using the Google Earth overlay process.

Data collection

In this research, the crash data from the NHTSA Fatality Analysis Reporting System (FARS) will be used. FARS is a nationwide census providing NHTSA, Congress, and the American public yearly data regarding fatal injuries suffered in motor vehicle traffic crashes (NHTSA, 2019). Crash data from the years 2008-2017 in the US will be collected, prepared, and merged. The APAAS system will be tested for three states: Florida, Virginia, and North Carolina. The sample sizes for these states are 24,976, 7122, and 12,421, respectively. The same process can be used to develop APAAS for other states.

Artificial Neural Network (ANN)

Artificial Neural Network (ANN), also called deep learning, is a neural network method that uses multiple hidden layers. An ANN mimics the human brain's biological neural network. As shown in Figure 3, the ANN has an input layer, a set of hidden layers, and an output layer. In this figure, for each pair of nodes, *x* represents an input node, *y* represents an output node, and θ is node bias. In each hidden and output layer, there are artificial neurons interconnected via adaptive weights (*w*). These weights are calibrated through a training process with input-output data (Haykin, 2009). Equation 1 presents a general formula for calculating the value of each node

$$y_j = f\left(\theta_j + \sum_i w_i x_i\right) \tag{1}$$

In this research, it is proposed to use the multilayer perceptron (MLP) method, a feedforward artificial neural network, which is useful for its ability to solve problems stochastically and often allows approximate solutions for complex problems. For each artificial neuron, there is an activation function, which can be any function with range [-1; 1]. The most common activation functions are the tangent sigmoid and the logarithmic sigmoid, as shown in Equations 2 and 3, respectively (Haykin, 2009; Vieira et al., 2017).

$$\sigma(y) = \frac{1}{(1 + e^{-y})} \tag{2}$$

$$\sigma(y) = -1 + \frac{2}{1 + e^{-2y}} \tag{3}$$

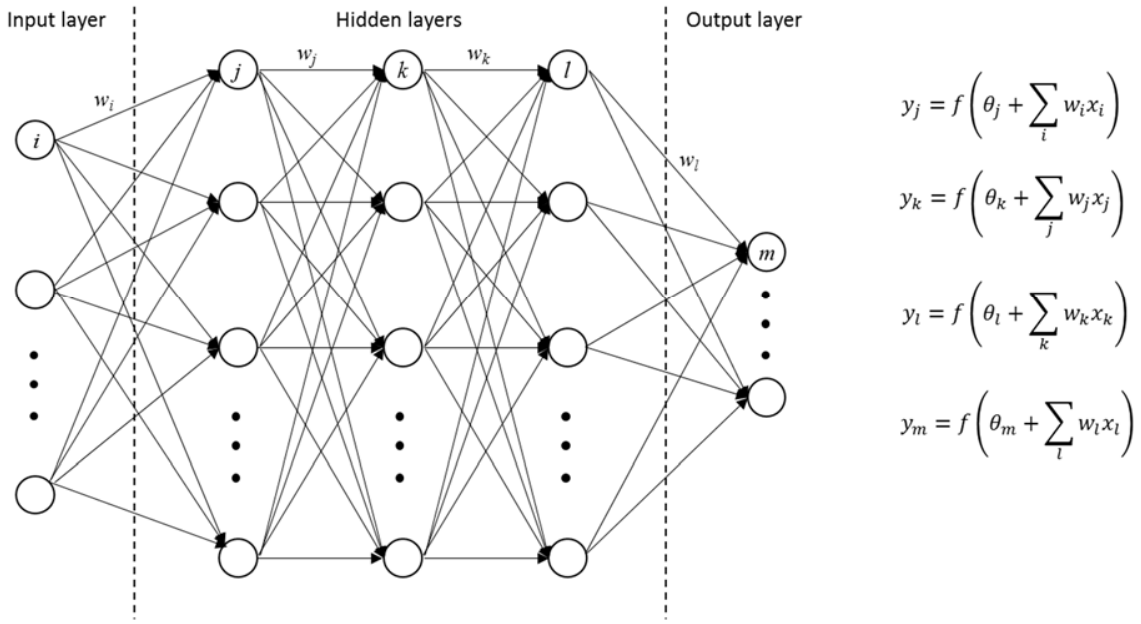


Figure 3: Artificial Neural Network (ANN) architecture

For the APAAS, the output layer includes a target variable, which is a vector consisting of the occurrence of VRU crash accident, location (longitude and latitude), and injury severity. The input layer includes predictor variables from the FARS datasets. Table 2 in the Appendix presents preliminary variables that can be used as predictors in this model. The paper will follow the SEMMA (Sample-Explore-Modify-Model-Assess) process to train and test the ANN model. SEMMA is a widely-used data mining process created by SAS Institute (Sarma, 2013). First, a sample of data is selected from the database and partitioned into training and validation samples. The training sample is used to train the model, while the validation sample is used to validate the model. Second, the data is explored to understand the variables and any unanticipated trends. Third, the data is treated and transformed as needed. Fourth, an ANN model is developed and trained based on the identified target variables and predictors. Fifth, the model is evaluated based on the prediction accuracy using average squared error (ASE) and predictive power using Lift chart, ROC chart, and Gini index (Tuffery, 2011). Adjustments to the activation function will be made to improve the model performance.

Automated Predictive Accident Alert System (APAAS)

The tested ANN algorithm from the previous step will be used to develop an Automated Predictive Accident Alert System (APAAS) for VRUs. An application using Java language will incorporate the predictive algorithm to allow feeding data to the system, which will produce a prediction of a VRU-related accident with a specific location and potential injury severity. Then, coordinates of the location will be overlaid on Google Map using Google Earth Pro. The Google Earth Pro coordinate overlay process includes several steps, including importing data, styling the map to customize the points on the map, editing map data, labeling data, and sharing and embedding the map (Google Earth, 2019). The APAAS will be developed as a web-based predictive tool.

CONCLUSIONS

Vulnerable road users (VRUs) are most at risk in traffic, and ensuring their safety is an important task for state authorities. The result of this research will be an automated predictive accident alert system (APAAS) that automatically predicts an accident involving VRUs before it happens (including location, occurrence probability, and injury level). This research fills the gap in the VRU safety literature, which lacks an automatic predictive system for crash accidents on the state level. A robust predictive algorithm for VRU-related accidents will be trained and tested using Artificial Neural Network (ANN), or deep learning, method. Using this algorithm, the APAAS can produce an accurate prediction of the VRU-related accident at a particular time in the future, along with the location and injury level, which would allow state authorities to proactively mitigate and to minimize the fatality risk to the VRUs. In this research, the system will be developed and tested for three US states, North Carolina, Virginia, and Florida. The algorithm can be used to develop a similar system for all states in the US, using the same ANN method and process.

REFERENCES

Constant, A., & Lagarde, E. (2010). Protecting vulnerable road users from injury. *PLoS Med* 7(3): e1000228. doi:10.1371/journal.pmed.1000228

OECD (1998). Safety of Vulnerable Road Users. *DSTI/DOT/RTR/RS7(98)1/Final*. Retrieved from https://safety.fhwa.dot.gov/ped_bike/docs/oecd_safety.pdf

Ghomi, H., Bagheri, M., Fu, L., & Miranda-Moreno, L. F. (2016). Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: A comparative study. *Traffic Injury Prevention*, 17(8), pp. 833-841, doi: 10.1080/15389588.2016.1151011

Google Earth Pro (2019). Google Earth Pro coordinates overlay. Retrieved from https://www.google.com/earth/outreach/learn/visualize-your-data-on-a-custom-map-using-google-my-maps/

Hartmann, M., Watzenig, D. (2019). Optimal motion planning with reachable sets of vulnerable road users. *IEEE Intelligent Vehicles Symposium*. doi:10.1109/IVS.2019.8814121

Haykin, S. S. (2009). *Neural networks and learning machines* (Third ed.). New York: Prentice Hall/Pearson.

Large, D. R., Kim, H., Merenda, C., Leong, S., Harvey, C., Burnett, G., Gabbard, J. (2019). Investigating the effect of urgency and modality of pedestrian alert warnings on driver acceptance and performance. *Transportation Research Part F: Psychology and Behaviour*. 60, pp. 11-24. doi:10.1016/j.trf.2018.09.028

Lee, E., Wu, J., Kang, T., & Craig, M. (2017). Estimate of mortality reduction with implementation of advanced automatic collision notification. *Traffic Injury Prevention.* 18(sup1). pp. S24 - S30. doi:10.1080/15389588.2017.1317090

Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, 54, pp. 27-37. doi:10.1016/j.ssci.2012.11.001

Nishimoto, T., Kubota, K., Ponte, G. (2019). A pedestrian serious injury risk prediction method based on posted speed limit. *Accident Analysis and Prevention*. 129, pp. 84-93. doi:10.1016/j.aap.2019.04.021

Otte, D., Jänsch, M., Haasper, C. (2012). Injury protection and accident causation parameters for vulnerable road users based on German In-Depth Accident Study GIDAS. *Accident Analysis and Prevention*. 44(1), pp. 149-153. doi:10.1016/j.aap.2010.12.006

Sarma, K.S. (2013). *Predictive Modeling with SAS Enterprise Miner – Practical Solutions for Business Applications (*2nd ed.)*.* Cary, NC: SAS Press.

Tuffery, S. (2011). *Data mining and statistics for decision making*. Hoboken, NJ: Chichester, West Sussex: Wiley.

Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience and Biobehavioral Reviews*, 74(Pt A), pp. 58-75. Retrieved from https://doi.org/10.1016/j.neubiorev.2017.01.002

Zaman, A., Ren, B., Liu, X. (2019). Artificial intelligence-aided automated detection of railroad trespassing. *Transportation Research Record*, 2673(7), pp. 25–37. doi:10.1177/0361198119846468

DECISION SCIENCES INSTITUTE

Understanding Leaveism: Mapping Factors Through Workplace Behaviour

Bohra Sarrah
Ahmedabad University
Email; sarrah.b@ahduni.edu.in

Modi Hasti
Ahmedabad University
Email; hasti.m@ahduni.edu.in

Shah Garima
Ahmedabad University
Email; garima.s@ahduni.edu.in

Shah Ronak
Ahmedabad University
Email; ronak.s@ahduni.edu.in

Shah Ruchita
Ahmedabad University
Email; ruchita.s@ahduni.edu.in

ABSTRACT

People at workplace face many psychological, sociological, and physical issues which can be caused due to multiple factors like working unpaid overtime, taking work home or working on vacation/leave. Leaveism includes practicing these three factors. This research aims to check the existence of leaveism globally and explore more factors affecting it. The hypotheses are tested using analytical tools. The research presents theoretical model covering the factors and consequences of leaveism obtained. The research shows that leaveism is not only about quality of work life but also about other consequences. The research provides managerial implications to address the problem of leaveism.

KEYWORDS: Presenteeism, Unpaid overtime, Workplace behaviour, Productivity, and Quality of work life

INTRODUCTION

People working in any space, may it be in corporate houses or in educational institutions, face many psychological, sociological, and physical issues like stress and fatigue, work- life imbalance, or ergonomic injuries. For instance, when an employee feels a kind of belongingness to his/her workplace, he/she tends to do more voluntary work after the pre- defined working hours, even if no extra benefits are

provided (Health assured, 2018). That employee would not mind working overtime or doing some work from home even after completing their daily tasks; but this can result in low productivity or can affect his/her work life balance negatively. Another example can be, if all the team members are finishing their tasks before time, it creates an indirect peer pressure on the employee to finish his/her work quickly as well. To make this possible, that employee might work overtime or while on leave. Though it implies that the work shall be done on time, this might affect the employee's or the organisation's efficiency negatively. Organisation's structural problems might also lead employees to work late or do unpaid overtime. For example, if the task allocated by the higher authority is not clear, or if short deadlines are given, then the employees may have to work for long hours to complete the task on hand before the deadline. This can induce stress and if this happens repeatedly, then it would result in poor quality of work which directly affects the organisation's performance.

The scenarios mentioned above have shown a series of event which is termed as presenteeism- a term used to describe employees who come to work even when they are sick (Aronsonn, Gustafsson, &Dallner, 2000). Many other scenarios were also considered a reason for presenteeism until Dr. Ian Hesketh, a researcher at University of Lancaster in the UK, coined the term "leaveism" in the year 2013 (Hesketh, Cooper, & Ivy, 2014). According to his definition, leaveism is a practice in which- employees use their official period off like annual leaves, paid leaves, accumulated flexi hours, etc. to take days off when they are sick or unhealthy. The definition extends to include the work done by employees at home because they were not able to complete it in their working hours and the work done when they are on vacation to meet their deadlines and complete the targets.

The term leaveism is coined in recent past and therefore not much has been explored as of now. So, the aim of this research is to study the factors that affect leaveism, check if it is prevalent or not and whether it affects the organisational setup.

LITERATURE REVIEW

The term "leaveism" was coined in 2013 by Dr. Ian Hesketh who is a researcher at the University of Lancaster in the UK (Hesketh, Cooper, & Ivy, 2014). This topic is under- researched but leaveism has an impact on organisation's effectiveness as well as on employees. The reason or motivation behind this phenomenon is unclear as it can be a result of loyalty, enjoyment, compulsion or any other factor. Thus, it becomes imperative to know which factors are leading to leaveism. Currently, employees utilize allocated time off, like annual leave entitlements, flexi-hours banked, re-rostered rest days, etc to take leaves when they are unwell/unhealthy; employees take work home that could not be completed in routine working hours; employees work while on leave or public holiday to complete the deadlines and targets (Hesketh, Cooper, & Ivy, 2014).

Presenteeism is one of the factors which helps in defining the concept of leaveism. It means showing up for work even when an individual is ill (Aronsonn, Gustafsson, &Dallner, 2000). A study by (Johns, 2009) shows that attending work while being ill leads to a greater overall loss of productivity than purposely shirking of work. A research by (Palo &Pati, 2013) has described an individual choosing to come to work when he/she is ill, by testing the following factors: 1) Work-related factors: employees may not have flexi working hours policy in the organisation leading to presenteeism 2) Personal factors: employees use work as an escape from family or social gatherings, more relevant in the Indian context. 3) Personality Factors: employees have intrinsic belief in themselves and their abilities to complete the tasks which make them think that they can perform equally good even when they are unwell (Bandura & Locke, 2003). Other factors that promote presenteeism include work teams that expect regular presence on the job, limited sick leave entitlements, high job demand which leads to burnout, fear of dismissal and fear about promotion opportunities.

Presenteeism and absenteeism have an indirect relationship. A study has shown that any factor that restricts the opportunity to remain absent could stimulate presenteeism (Koopmanschap, Burdorf A., K., W. J., & Severens, 2005). Organisational policies such as sick pay, attendance control, downsizing and permanency of employment have all been considered responsible to instigate presenteeism (Johns, 2009). This helps in explaining the relationship between presenteeism and leaveism. Leaveism is using allocated time off to work, like coming to work while being sick and not utilising the dedicated sick leaves for the same. This has a direct impact on employee wellbeing and organisation's productivity.

To test presenteeism in Sweden, a study was conducted on the Swedish workforce of approximately 3801 employees, to find out whether the employees prefer to go to work even if they are sick. A lot of changes have been made in their sickness compensation system, for example, sickness benefit of 90% of the income was reduced to 75%, for the first 14 days of sickness. A qualifying day was introduced, which meant a person had to wait to be eligible for the sick leave. Also, the sickness benefit was increased to 30 days instead of 14 days (Aronsonn, Gustafsson, &Dallner, 2000). Such changes in the sickness policy pressurized the conditions at workplace. Thus, problems of personal finances and a psychological change in the behaviour of employees towards taking sick leave were observed. They preferred to come to work rather than taking a leave. Though, there might be other reasons too for why an employee would come to work even if he/she is sick, like, their loyalty and unwillingness to burden the organisation, job satisfaction or fear of being dismissed or demoted, etc.

Presenteeism is not limited to the corporate sector. The highest presenteeism is found in the education, care, and welfare sector. And it is also observed that in these fields, a greater number of women work as compared to men. So, it is concluded that women tend to work when sick rather than taking a leave (Aronsonn, Gustafsson, &Dallner, 2000).

This suggests that sickness presenteeism is more in women than in men. From the sample size of 3801 employees who were interviewed, a third of the employees mentioned that they had gone to work, given their ill state of health, for two or more times in the preceding year of the survey even when they felt that they should have taken a sick leave. The employees have reported that they had to re-do their work when they had remained absent because of sickness, and the odds ratio of the same is 95%. The highest sickness presenteeism is observed to be due to fatigue and depression. This means that even if employees are depressed or stressed, they do not take leave and come to work. The relation between the level of pay and sickness presenteeism using Spearman's rank correlation coefficient showed that high sickness presenteeism is related to lower than or equal to the median income level. From this entire study on Swedish employees, it was observed that sickness presenteeism is prevalent in various sectors resulting in leaveism.

A study using Quality of work-life module and General Social Survey (Casner-Lotto, 2003) on 1787 employed people found out that 26% of the employees reported having overtime work which is two hours more per week and on an average, two days more in a month as compared to their counterparts.Working overtime has a direct social impact of getting less time to spend with the families and for their personal matters (Casner-Lotto, 2003). These findings suggest that individual and social wellbeing come at a trade-off of higher earnings or economic rewards which comes at the trade-off of more fatigue and stress. Apart from the social impacts, it is also observed that working overtime for a continuous period also affects the employee’s on-the-job performance. (Golden, 2006)

The paper by (Hart, 1999) states the reasons why an employee would work without getting paid for it- 1) Uncertainty over task completion time:Whenthe employer might not be very conversant with the detailed work requirements to complete the task, the estimated period for work completion is based on the forecast and not the actual result, so the employees need to work more at the later stages to complete the task undertaken. 2) Task allocation: The employees might work more and without pay if they are assigned tasks based on past performances and dedication. So, an employee would want to put more effort so that he/she receives the next task. 3) Team performance: If the team is to be paid based on their value addition to the group, the low productive employees will force themselves to work more to be at par, even though it is without pay. Another reason involves unanticipated layoffs or unscheduled absences due to illness. In such cases, the team leaders, to maintain the team's performance, reputation, and remuneration, might increase their working time. 4) Bullying: Employees can be bullied in a way that affects their performance like, for instance, blaming one wrongfully, sabotaging their work or continuously interfering in their work can affect employee’s performance and efficiency. There are psychological effects of bullying such as constantly thinking, worrying about work even when one is on leave or dreading to go to work and wanting to stay at home instead. And all these contribute to unpaid

overwork where one must work more to complete the task at hand (Crystal Raypole, 2019).

The other reasons based on the psychological construct of the employees are- using work as an escape, organisation citizenship behaviour, high expectations for an increase in compensation, doing meaningful work and deep connection with the workplace. Finally, the paper by (Hart, 1999) concludes that working unpaid hours is positively related to low productivity within the team, status of managers, lack of union, non-involvement in ethnic groups, high hourly wages, marital status, and age; which implies leaveism due to unpaid overtime.

Apart from working unpaid overtime, there are various instances observed that employees take work home, in excess to the daily work, which goes unaccounted for. This creates a discrepancy in considering the extra work done by these employees. According to the study by (Lucy P., Wulff, & Sabrina, 2010). About 8 percent of corporate employees bring work home, to catch up or to finish. The employees bringing work home end up working more than those who only work at their workplace. The extra work carried out in one's home is usually unpaid. This unpaid work leads to imbalances in work life and family life and creates dissatisfaction. This increase in working from home is due to advances in information technology.

In addition to unpaid overtime and taking work home, the people at workplace have developed a ‘working vacation culture’ with devices that are engineered to work from anywhere and hard to switch off entirely. People who work on vacations find a trade-off between spending vacation relaxing and come back with overflowing inboxes or taking work on vacation. Research has found that up to 66% of the total workforce of America work while they are on vacation. (Timely Blog, 2018)

A study was conducted by Project Time Off (Leondhart,2019) which shows that the work- culture plays a role in working while on vacation. According to that study, people who work remotely or work-from-home tend to put pressure on them to stay connected to their workplace as compared to people who do not work remotely or from home. It also states that 78% of the people who were interviewed were comfortable with taking their work on vacation as they preferred taking the stress of work on vacation rather than the stress of cutting off the connectivity entirely. Also, many people who are self-employed and are the head of their businesses felt that they cannot lose connectivity entirely as there is no one to manage their business. The 2019’s Priceline study on Work-Life Balance showed the results that 47% of Gen Z and 40% of the millennials who were on vacation felt the pressure of checking their phones (Vaynshteyn, 2019). The psychological aspects which make people work while being on vacation include- fear of the mess after his/her return, fear of falling behind, proving dedication, to stay engaged, fear of a break affecting the future of the career. The study has identified the above- mentioned factors and concluded them to be unhealthy. Such factors are considered a standard mark of a work culture that is toxic as there are no policies stating anything about working while on vacation.

This is substantiated by the example of Harvard Business Review where they term e-mails sent by the employees while being on vacation as “tiny cultural erosion”, which indicates that employees do not completely switch off from their work leading to leaveism.

Another aspect of this is that employee’s donot go out on vacations but instead keep working. Research by the U.S. Travel Association reported that at the end of 2017, 52% of the U.S. employees had their vacation days unused; whether paid or unpaid. This summed up to approximately 212 million vacation days in a year alone. While finding the answer to why the employees are not going on vacation even when they know the mental and physical benefits of it; they arrived at the term ‘Vacation guilt’ (Boogaard, 2019). Vacation Guilt arises when an employee gets some paid time off but feels that they have not earned it or deserved it. Studies have shown the results that those employees who have the un-limited vacation periods in-fact take less time off. Doug Vargas, a senior application developer at Google stated that the company did not offer unlimited vacation days as it made the employees feel that they did not earn that time off (Vice, Michaela Cross , 2018).

The Priceline study showed that new joiners feel that it is not appropriate to ask for vacation immediately; 6 out of the 10 respondents stated that they would wait for at least 6 months to ask for long leaves and 21% of the respondents would wait almost a year to feel comfortable to ask for long break. (Vaynshteyn, 2019)

Another factor leading to leaveism can be ‘work martyrs’. According to the study by (Vice, Michaela Cross, 2018) millennials are considered as ‘work martyrs.’ A work martyr is someone who likes the idea that no one else can do their job and that they are irreplaceable. They try to show complete dedication and feel guilty about taking any time off. These workers are inefficient. By studying a group of successful business consultants; it was found that they faked their working hours. This would also make sense because the study states that the most productive workers work 40 hours a week. Exhausted over-workers may produce more in a shorter run but in the longer run, they are inefficient (Vice, Michaela Cross, 2018).

As per the book ‘Unbending Gender’, by law professor Joan Williams, it is described that workplace is built around those workers who are continuously available and are present full time as they do not have other household and childcare responsibilities. These workers are ‘ideal workers’ (Heppner, 2013). The ideal worker at the executive-level is often expected to work long hours in addition to the regular working hours. According to (McClintock-Comeaux, 2013), ideal worker is the one who is committed to the workplace and does not allow any distractions, including personal life and family, outside the work environment to affect his/her job. This would also mean working overtime for the job. Hence, this leads to leaveism as an ideal worker is expected to work long hours without getting paid for it.

The paper by (Chaitanya & Tripathi, 2001) explains the dimensions of Organisation Citizenship Behaviour (OCB) based on a sample of 100

scientific officers of a public sector organisation situated in the Southern part of India. The roots of OCB lie in the social exchange theory which identifies the conditions under which people feel obligated to revert when they are benefitting from a person's or an entity's actions. This implies that if they feel they are paid well or are getting many over- all benefits, then people would be more inclined towards doing extra work at that place. OCB is also known as extra-role performance (Linn, Jill W., & Richard, 1994) pro-social behaviour, (Richard G., James, Daryl O., & Robert, 1997) organisational spontaneity, spontaneous organisational behaviour (George & Brief, 1992) and even counter-role behaviour (Staw & Boettger, 1990). The formal definition of OCB can be- a discretionary work behaviour of the employees which is not recognized through any formal reward or sanction system of their organisation yet is very important for organisational effectiveness. The paper by (Bolino, Klotz, Turnley, & Harvey, 2013) aims at investigating reasons when OCB might be a negative trait in an employee- due to self-serving motives, mundane motives or other negative factors; OCB can be of low quality or might be performed in place of in-role task which makes it harmful to organisational performance and employees exhibiting OCB might be overstressed or overloaded with work and might have family conflict. Thus, it is to be noted that the dark side of OCB shall also impact the dark side of leaveism, directly or indirectly.

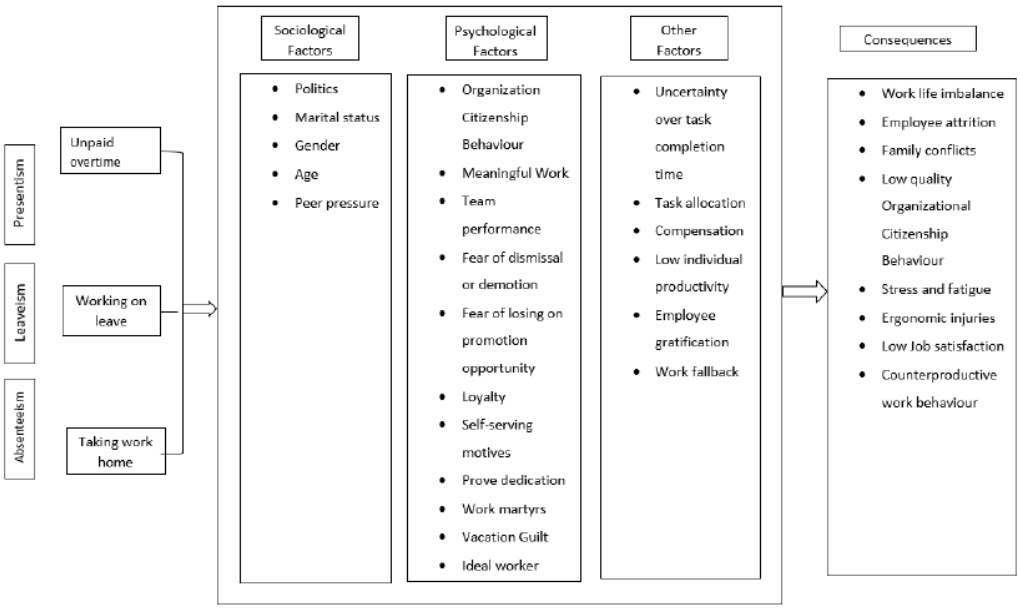
RESEARCH GAP

Three major research gapsidentified are as follows:

1) Non-discovery of leaveism due to the dominating prevalence of presenteeism and absenteeism at workplaces.
2) It is observed that people assume that the ill-effects of overtime lead only to psychological or physical harm and hence the sociological factors leading to leaveism remain unexplored.
3) As only individual factors have been researched to study leaveism, the possibility that multiple factors in different contexts might be together leading to leaveism, is overlooked.

Therefore, it becomes imperative to study the topic of leaveism holistically so that this problem can be addressed accordingly at various workplaces.

THEORETICAL MODEL



The above model, based on the literature review, studies various factors that affect leaveism. The factors for unpaid overtime, working on leave and taking work home, which means leaveism, are divided into 3 categories- sociological, psychological, and other factors. This is done to differentiate between macro (social) and micro (individual) factors. As a result of these factors, potential consequences are stated, which signal the urgency to address the matter of leaveism.

The term leaveism has been derived from the terms - absenteeism (Kenton, 2019) and presenteeism, which were initially used to define remaining absent due to sickness and being present even while one is sick, respectively. But after the term leaveism was coined in the year 2013, it covered- unpaid overtime, working on leave and taking work home. The factors for the three subparts of leaveism definition are divided into 3 broad categories- Sociological factors mean the variables that arises out of the community, culture, family, religion, ideology, etc. which influences one's individual thinking and actions (Rose, 1967). Psychological factors define those that states an individual's perceptions, learnings, attitude, and beliefs (Yvette Brazier, 20018). The other factors are those that fall under the structural set up of an organisation or those which cannot be labelled specifically. In accordance with this bifurcation, sociological factors include all societal ones as politics, gender, age, etc. and how do these sociological factors lead to leaveism. In a similar manner, psychology of an individual showing his/her loyalty, fears, guilt, happiness of doing work, etc. is described as a part of psychological factors. Lastly, other factors include task allocation, work fallback, centralization, etc. which sum up to the structural issues of an organisation, while low individual productivity or work fallback cannot be labelled as an individual factor that leads to

leaveism.

Summing up all the factors, the potential consequences of those are stated as potential outcomes. For instance, worrying too much about missing a promotion or fearing a deadline can lead one to work endlessly and eventually result into stress and fatigue. Thus, this model also shows the series of events that lead to the consequences mentioned.

HYPOTHESES

The hypotheses are derived from factors mentioned in the literature review and theoretical model.

This research intends to test the existence of leaveism; peer pressure; organisational citizenship behaviour; fear of dismissal; self-serving motives; working more to prove dedication; using free time to work; low individual productivity; compensation; employee gratification; work pile up; organisational structure; allocation of new task/project; technology; mundane work; and stress and fatigue as a consequence of leaveism.

RESEARCH METHODS

Primary Research Method

Primary Data is a method of collecting data on field which is original, unaltered, and more reliable. This data collection method ensures that the data collected is relevant to our research and out of context variables are not addressed which saves the energy and resources of the data collectors making it easier for analysing the original data (Syed Muhammad, 2016). This research includes primary data collection through questionnaire method which is as follows:

Questionnaire

Questionnaire is a set of written questions with either close ended answers or open-ended answers.The questions framed for our survey are based on psychological, sociological & other factors identified in the theoretical model. It is a structured questionnaire which considers all the factors affected by leaveism to derive conclusions over what role it plays and its impact on the factors. The questionnaire includes scenario-based questions to further analyse the factors and map people's behaviour in response to those factors.

The questionnaire is designed in 4 different sets based on the respondent's occupation- student, faculty, corporate employee or self-employed. All the sets have different questions with situations which relate to the respondent's occupation. Each situation defines one or more factors. The questions include both open-ended and closed-ended questions. The responses will help us map those factors through tools like SPSS and NVIVO.

Secondary Research Method

The secondary source is a source of information that is not collected first-hand. It is the source of information that is obtained from someone's primary data. The most recently published sources are the most relevant while conducting study through secondary sources (Solberg, 2000). Examples of secondary sources include websites, journal articles and reports.

TOOLS AND TECHNIQUES

SPSS (Statistical Package for the Social Sciences)

After the coding of the responses collected through questionnaire, the data is run through the statistical tests to enable the user to build and validate the predictive models. The data is analysed and statistical models will be derived through SPSS which will help in the final analysis and to test our hypotheses (Statistics Solutions, 2019). All the coded, closed ended questions from the research questionnaire are run through SPSS and analysed based on one-way ANOVA and T test.

NVIVO

NVIVO is used for analysis of qualitative data which could otherwise not be done, except from Atlanti and Reader from SPSS. It is a software package developed by the QSR international. It is used to analyse unstructured data, image data, audio, video, interviews, surveys, articles, reports,and focus group discussion. In this research, NVIVO is used for analysing the open-ended responses of the research questionnaire, by conducting theme analysis and sentiment analysis in order to know the main content or areas of concern and the sentiment (positive, negative, neutral or mixed) attached to it.

SAMPLE

As reaching out to the masses is a quite difficult task, sampling methods are used to identify representative groups which give relatively closer results to that of the population. Sample is a defined set of people with good representation of population with specifications like age, gender, nationality etc. To test a heterogeneous population, samples are divided into small subgroups (known as strata) to have some homogeneity which can be used for further data analysis.

Following are the details of our sample for the questionnaire:

Target audience: Students, corporate employees, self-employed, faculties (this research includes only white-collared positions)

Type of sample: Heterogeneous

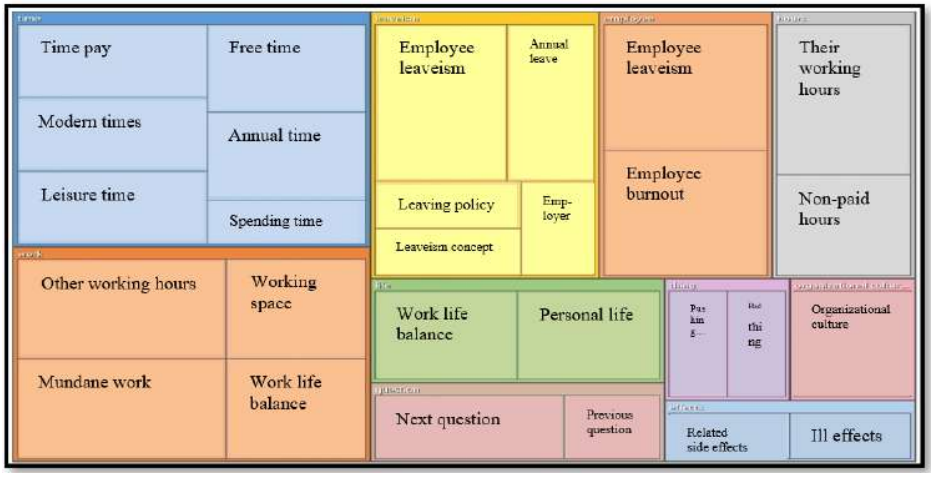
Sample size: 253

Geographic location: Cross-border

QUALITATIVE ANALYSIS THROUGH NVIVO

Theme Analysis

NVIVO uses a language pack to identify the most important themes, analyse the content from the file and present it in the form of parent nodes and child nodes. Following is the diagram that is obtained based on the open-ended question which was to know the idea of leaveism from the perspective of the respondents:



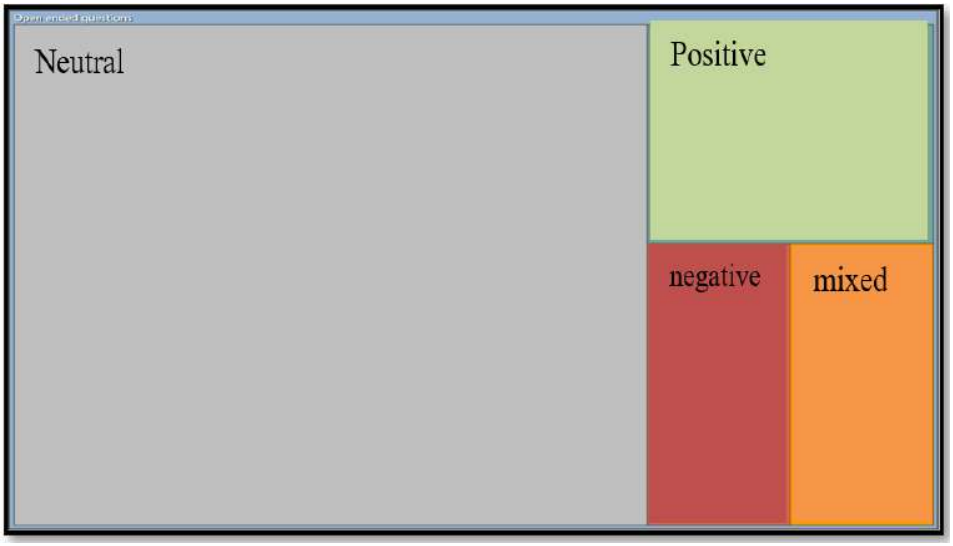
From the above diagram, it is observed that time, work and leaveism are the main themes identified. This means people have most repeatedly associated time, work and leaveism to the concept of leaveism. On the other hand, it is seen that thing, organisation culture and effects have been slightly associated. This signifies that most of the respondents perceive leaveism from the viewpoint of a factor and not a consequence.

The parent node ‘time’ shows 6 different child nodes that the respondents have majorly mentioned while sharing their perspective on leaveism. The parent node ‘work’ shows 4 different child nodes which have been traced by NVIVO. By working hours, it can be inferred that the respondents have associated the work they do during that period and how work is often assigned to them at odd hours too. This can be linked to their idea of leaveism as working overtime too. The parent node ‘leaveism’ includes 5 child nodes.

While analysing the other child nodes in total, it is observed that only a few consequences like employee burnout, side effects and ill effects have been considered.

Sentiment Analysis

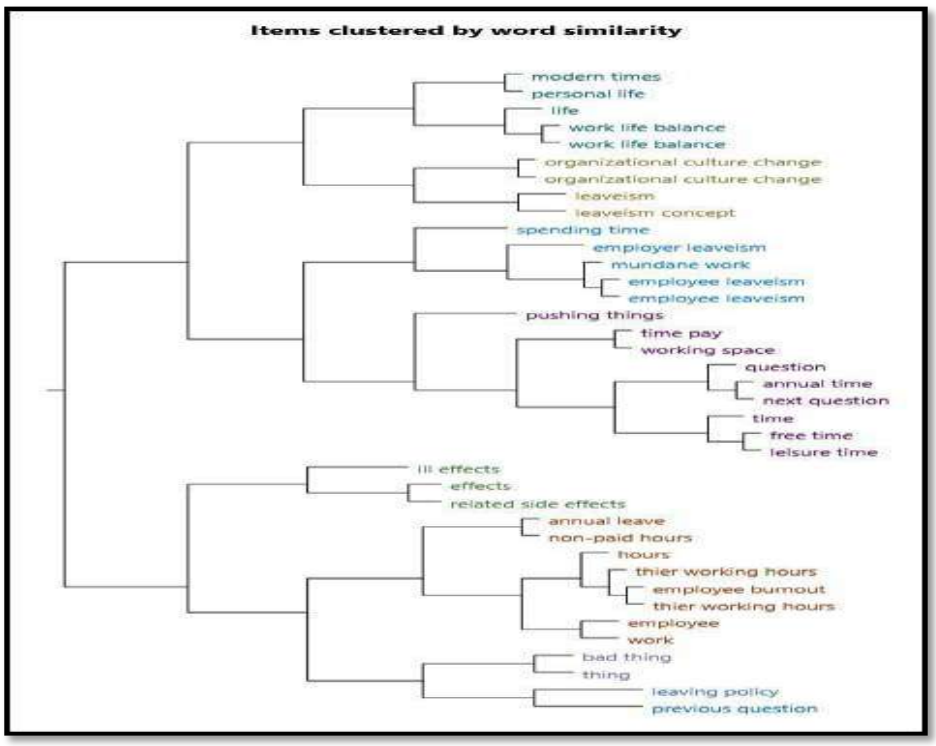
The sentiment analysis highlights the tone and nature of sentiments in which the responses are written and how it conveys. The following figure shows the sentiment analysis for the open-ended questions in the questionnaire.



The main sentiment of most of the responses is neutral along with positive to some extent. This conveyed that the responses had a neutral tone. This shows that the responses barely consisted of any negating words. The neutral nature of responses could be due to various reasons like, the respondents not understanding the meaning of the term leaveism properly, due to misinterpretation of certain questions, lack of clarity or awareness. Although there were responses with positive, negative, and mixed tones, their proportion was very less. So, we can understand from this sentiment analysis that the overall tone or nature of responses was neutral.

Items Clustered By Word Similarity

The nodes in the cluster diagram have been clustered by word similarity. The words here, are clustered based on the theme analysis of the responses generated from the questionnaire.

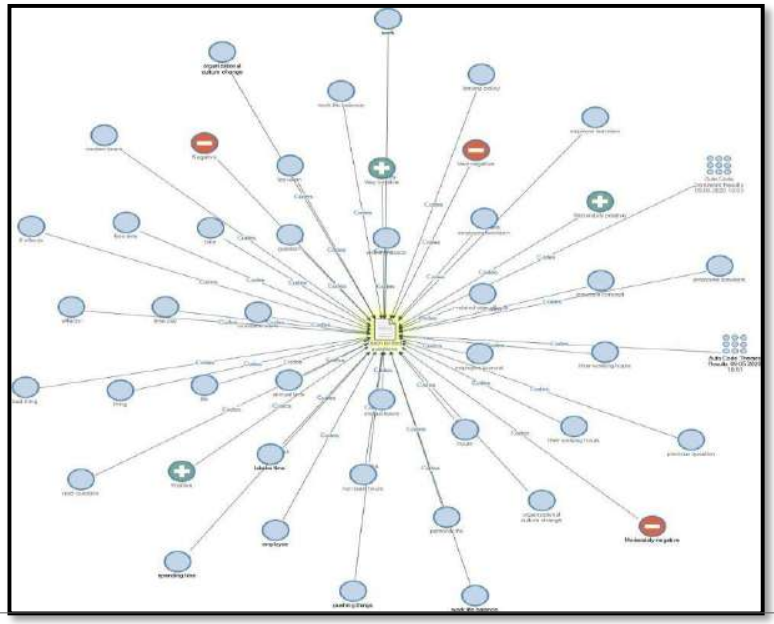


The first cluster based on word similarity comprises of modern times, personal life, lifeand work life balance. It means that people believe that in modern times, because of leaveism, work life balance has become an issue and because of this they are not able to give time to their personal life. This is because they work overtime, be it due to psychological, sociological or organisational factors. The second cluster talks about leaveism due to the organisational culture. It means that people feel that leaveism is being practised because of the organisation's culture like their values, vision, mission, beliefs, working style, workspace, and policies. The next cluster reflects that the employees also practice leaveism to get a break from their mundane work and time spent or required to be spent for completion of the work. The other clusters reflect that leaveism might also be due to the working hours, leisure time or free time. This shows that leaveism can be compelled or it can also be practiced willingly.

The second half of the cluster talks about the consequences of leaveism. People believe that there can be ill effects and related side effects because of working overtime, taking work home or working while on leave/holiday. People might be practicing leaveism, if their working hours are taken into consideration for payment or have non-paid hours for not working. This can be because of the leave policies of the organisation. Thus, people think that leaveism is a bad thing, as reflected in the cluster, and lead to employee burnout. So, addressing it is utmost important for the well-being of students, faculties, employees, self-employed and the organisation.

Sociogram Analysis

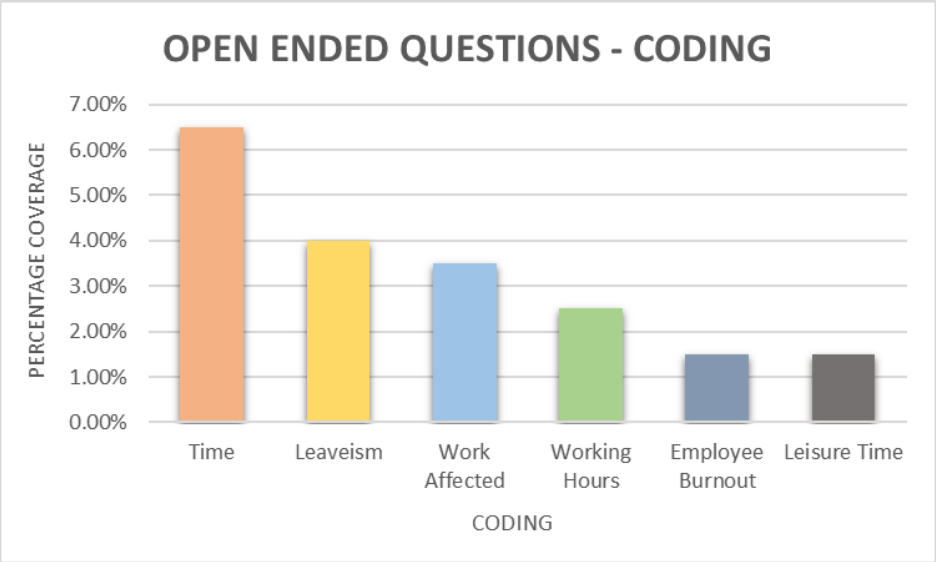
Sociogram in NVIVO helps in analysing the social networks and their visual connections. Groups, people, and organizations may be joined by different types to relationships, but not all are directly joint, there can be multiple relationships. Sociogram analysis deals with the pattern and structure of those relationships and identifies the causes and consequences. It can also be used to identify the flow of information and the main source of information. In this research, the sociogram analysis is done on the open-ended responses and so there in-degree as the information is coming in.



The information that has been flowed in from the responses reflect the respondents' views onleaveism and the sentiment attached to it. There's a mix of positive and negative responses. i.e. some people think that leaveism has positive impacts, while some believe that it has negative and moderately negative impacts. The reason as to why leaveism might be practiced is because of free time, leisure time or the working time/hours and the work requirements. Other reasons can also be the organisation culture, to get a break from mundane work or the workspace, which compels one to work more. The sick and leave policies can also be a factor leading to leaveism. The in-degree here also suggests the consequences of leaveism, which are employee burn out, work life imbalance and not being able to get time for personal life. Hence, in the sociogram some people have referred leaveism as a bad thing and this gives a reason for the organisation's to address the practice and ill effects of leaveism.

Bar Graph

The theme analysis has broadly defined the themes of the responses but how many people believe a particular theme can be known by the following diagram:



The diagram shows that how many percentages of people have contributed to the responses to the themes observed in theme analysis of the responses generated from the questionnaire.

The first indicator is time; about 6.5% of respondents of open-ended questions say that time is a factor for leaveism. The reasons for it are their type of work is such that they end up practicing leaveism, their organizational culture is such that work more or the way that they spend their time on work. It means that long working hours prevail, and they also think that it affects their work-life balance and results in employee burn-out.

The leaveism in the graph indicates that around 4% of respondents

believe that leaveism is present and they are practicing it. The reasons for the same are that in modern times and it is hard to stay un-connected to the organization, their working space and culture is such that they work for un-paid working hours and due to this they are also not able to give time to their personal life and it affects their work-life balance negatively.

Work is another thing that is affected by leaveism and it is believed by 3.5% of the respondents. Work is affected negatively as one works for long hours doing the same mundane work and that too unpaid. It also results into employee burn-out and related side-effects.

There is also a part where around 1.5% have agreed that there are effects of leaveism. This means that leaveism has effects like burnout and fatigue resulting from long working hours, imbalance in work-life balance, ill effects related to working long hours, etc. These all factors affect people's personal and professional lives negatively and must be addressed for wellbeing of employees, students, self-employed and faculties and also for the organization itself.

QUANTITATIVE ANALYSIS THROUGH SPSS

The tools of SPSS that are used in this research are- Cronbach's Alpha, Kaiser-Meyer-Olkin (KMO), Bartlett's test of sphericity, One Way ANOVA and T-test, which are as follows:

Cronbach's Alpha

Validity and reliability are the major elements in the evaluation of the instruments like survey questionnaires, simulations, and skill or aptitude tests. Cronbach's alpha, developed by Lee Cronbach in 1951, measures the internal consistency of the test. The consistency helps in understanding the extent to which all the items of the test measure or determine the same construct or concept, which means that the items are inter-related (Tavakol&Dennick, 2011). It is vital to conduct this test to measure the internal consistency before the test is employed for examination or research purpose. It also ensures validity and reliability of the test. The range of Cronbach's Alpha reliability test is between 0 and 1. There is no lower limit of the coefficient but closer the coefficient to 1, greater the internal consistency of the test's items. As the score the test increases, i.e. the reliability increases, the test score attributable to the error decreases (Gliem&Gliem, 2003).

The Cronbach's Alpha obtained by running the test in SPSS is 0.762. It is closer to 1 and so there is strong internal consistency of the items (factors) of the test. It also ensures validity and reliability of the test which is an indication to carry the study forward for the research purpose. The greater value reflects that the factors of leaveism are interrelated and have consistency amongst them. It also ensures that the test score attributable to the error is less and hence the research can be taken ahead.

Kaiser-Meyer-Olkin (Kmo) and Bartlett's Test of Sphericity

Many scientists use Factor Analysis to ensure whether that variables that have been used to measure the concepts are measuring the intended content. The KMO and Bartlett's test measures the suitability of the data for the Factor Analysis. It measures the adequacy of the variables of the sample, either as separate models or for the complete model. It measures the saturation of the data and whether the number of responses received is adequate. KMO test also measures the proportion of variance among the variables under study. The KMO values between 0 and 1. There are some thumb rules for the interpretation of the test. If KMO value is between 0.8 and 1, it means that the sampling is adequate. If the KMO value is less than 0.6, it indicates that sampling or the sample size is not adequate and corrective actions are required. If the KMO value is closer to 0, it means that there is widespread correlation, which becomes a problem for the Factor Analysis (Stephanie, 2011).

The KMO and Bartlett's value obtained for the study is 0.889. This signifies that the sampling is adequate and can be used for the research purpose. It also means that the data is saturated. As the KMO value is greater than 0.8, it means that the data is suitable for Factor Analysis and so the factors that lead to leaveism can be studied further by conducting T-test or One-Way ANOVA test.

The analysis using SPSS is done by interpreting the data generated from two tests. The first is the one-way ANOVA test which does the analysis of variance and the second is the non-parametric T- test for paired samples which analyses the degree of correlation and level of significance between pairs of factors.

ANOVA

ANOVA is used to identify the statistically significant differences between the means of two or more unrelated and independent variables. ANOVA is predominantly used when there are more than three groups to be analysed as two groups can be analysed using a simple T-test (One-way ANOVA in SPSS, 2007). ANOVA will generate the sum of squares, the degrees of freedom, mean square, F statistics, and the significance level of the variance of two or more groups/variables. In simple sense, different groups are tested to know the difference between them through ANOVA. If the two means are equal, i.e., there is no variance between them, then it becomes a null hypothesis. The significant result will mean that two means are unequal and have variance, this will lead us to accept the intended hypothesis (ANOVA Test: Definition, Types, Examples, 2004).

Analysis of One-Way Analysis Of Variance (Anova)

ANOVA is used to see and determine that if there is any significant difference, statistically, between two means of independent groups i.e. groups that are unrelated.

Here, in the analysis, different factors are tested against many groups. Each factor is taken up against groups- gender, age group, occupation, doing extra work at home and working on vacation to see how these groups affect the factors and the level of significance they have on that particular factor.

Significance of gender and working on leave/vacation is more on proving dedication. The Significance Probability (Sig.) of gender is 0.529 and of working on leave/vacation is 0.551 (as per Table 1) which means that while one is on a vacation/leave, he/she has an urge to prove their dedication by working more even though they are on leave and end up practicing leaveism.

As per Table 2, significance of age group is more on peer pressure. The Significance Probability (Sig.) of age group is 0.891 which means that as and when age grows, the individual feels more peer pressure and hence work more to be apart with their peers.

Significance of gender and doing extra work at home is more on Organizational CitizenshipBehavior (OCB). The Significance Probability (Sig.) of gender is 0.972 and of doing extra work at home is 0.761 (as per Table 3) which means that the more one feels that they owe much to their organization and should pay back by working more end up working extra at home and practice leaveism.

As per Table 4, significance of doing extra work at home and working on leave/vacation is more on Low Individual Productivity. The Significance Probability (Sig.) of doing extra work at home is0.556 and that of working on leave/vacation is 0.596 which means individuals who have low productivity levels, do more work at home and also work while they are on leave/vacation and hence do more unpaid overtime.

Significance of age group, doing extra work at home and working on leave/vacation is more on work as escape. The Significance Probability (Sig.) of age group is 0.635, that of doing extra work at home is 0.808 and of working on leave/vacation is 0.685 (as per Table 5) which means that people use work as an excuse to avoid many things that they do not want to address to and while trying to do so they work more than their paid hours.

As per Table 6, significance of doing extra work at home and age group is more on employee gratification. The Significance Probability (Sig.) of doing extra work at home is 0.558 and that of age group is 0.973 which means that with age and people feel need to be or their work to be gratified and to make it happen they work more from home over paid-up working hours.

Significance of doing extra work at home is more on stress and fatigue. The Significance Probability (Sig.) of doing extra work at home is 0.819 (as per Table 7) which means that while works more from home, it eventually leads to more stress and fatigue.

As per Table 8, significance of doing extra work at home is more on impact of leaveism. The Significance Probability (Sig.) of doing extra

work at home is 0.771 means that the more people work extra from home more is the impact of leaveism on their lives.

Significance of age group and doing extra work at home is more on free time. The Significance Probability (Sig.) of age group is 0.811 and of doing extra work at home is 0.909 (as per Table 9) which means that as people get more free time, they tend to work extra to use their time. It also refers to people working from home during the COVID-19 crises which defines work from home as a new form of leaveism.

As per Table 10, significance of doing extra work at home, gender and age group is more on leaveism. The Significance Probability (Sig.) of doing extra work at home is 0.600, of gender is 0.618 and that of age group is 0.618 which means that with age and with people working more from home, it depicts that people practice leaveism.

Analysis of T-Test

T-test is a non-parametric test which is a part of inferential statistics. It is used to identify the significance of difference in the mean or the variance between two variables, independent and dependent. There are two types of T-tests, independent and paired tests. The analysis for this research includes T-test for paired samples. The degrees of freedom indicate the number of pairs. The level of significance or the alpha level that should be accepted is usually greater than or equal to 0.05. This conveys that the independent factors or variables which show the correlation are analysed and then the level of significance among them is considered for arriving at a final decision.

From Table 11 attached in the Annexure, it was identified that 9 pairs reflected positive correlation between the two variables under consideration with respect to leaveism. These include organisational citizenship behaviour, work environment, work pile up, technology, self-serving motives, fear of dismissal, organisational structure, delegation, new task or project, work as escape, insecurity, work expectation and work life balance. From all these pairs, self-serving motives reflected 0.255 degree of correlation, which is the highest among the others. While fear of dismissal and organisational structure shows 0.014 degree of correlation, which is the lowest among all. But all these 9 pairs cannot be claimed to be the main factors of leaveism till their significance is checked.So, from Table 12 attached in the Annexure, it was found that out of these 9 identified files, only 3 pairs reflected significance above the ideal alpha level. The ideal alpha level to accept the pairs is taken as 0.05. Accordingly, work pile up and technology, fear of dismissal and organisational structure, and delegation and new task/project reflected higher significance as factors affecting leaveism. Work pile up and technology showed 0.811 level of significance which is the highest among the pairs identified.

Thus, we can conclude that 6 factors namely – work pile up, technology, fear of dismissal, organisational structure, delegation and new

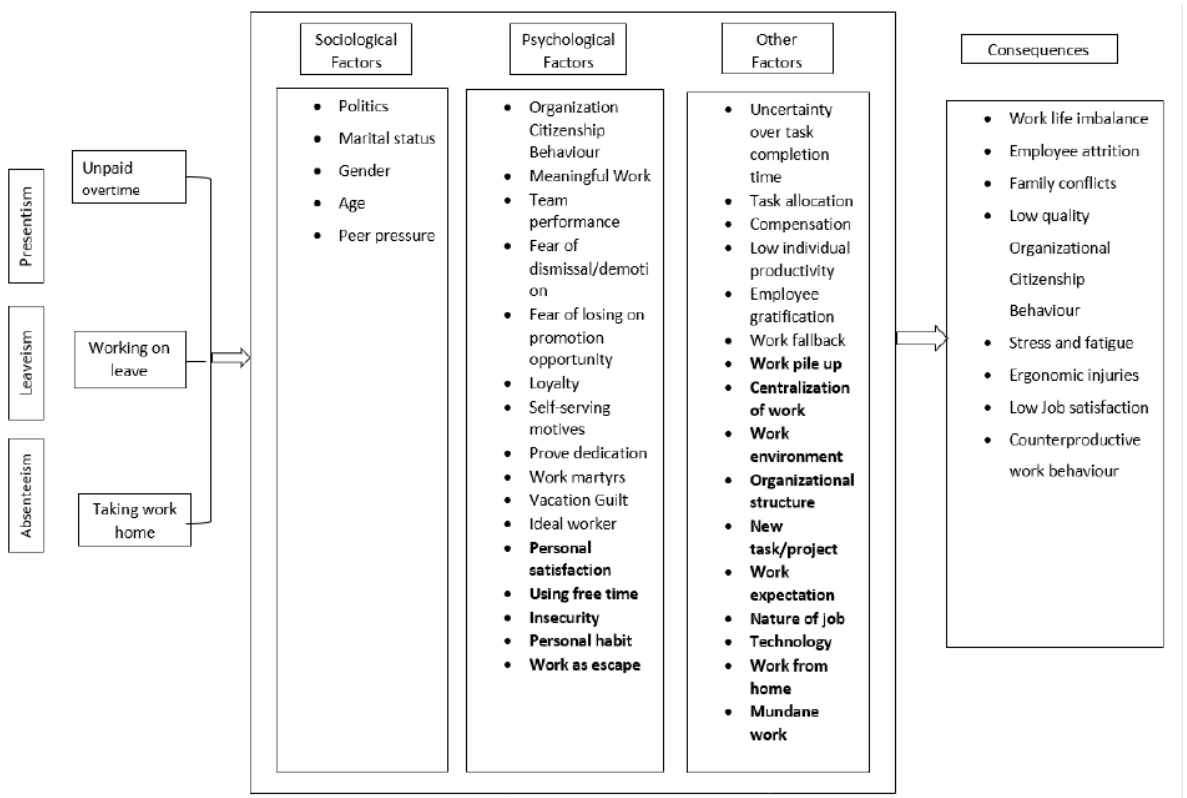
task/project not only have positive correlation among each other, but also show higher level of significance with reference to leaveism.

RESULTS

The results of the above-mentioned tests prove that leaveism exists. It also proves that peer pressure; organisational citizenship behaviour; fear of dismissal; self-serving motives; working more to prove dedication; using free time to work; low individual productivity; compensation; employee gratification; work pile up; organisational structure; allocation of new task/project; technology; mundane work are the factors leading to leaveism. Finally, it also shows that stress and fatigue are proven consequences of leaveism.

The above-obtained results from various tools helped in knowing the presence of leaveism and in identifying the factors that lead to it. Out of the 16 hypotheses that were tested, two have been accepted, which means that compensation and self-serving motives are not the key factors that lead to leaveism. This implies that all the other factors tested by the hypotheses lead to leaveism.

REVISED THEORETICAL MODEL



After analysing the responses collected and the results obtained above, some new factors are identified which are also responsible for existence of practice of leaveism. These factors mainly fall in category of psychological factors and organizational structure, which are explained in the following paragraphs.

Some people work for longer unpaid hours just because of their organizational structure i.e. they are short staffed, or nature of their job is such or people expect them to do so at work. Then there are some factors that depend on an individual i.e. he/she might have it as their personal habit and are shaped into working that way or they might start feeling insecure when they do not complete task and role that they are expected to do or they might do it only when a new task/project is handed over to them.

It was also seen that, because of COVID-19 crises, work from home is emerging as a new form of leaveism. People have started working more unpaid hours because of change in set-up, because they started enjoying it or because they do not want to waste their free time and hence utilized it.

Thus, these are the sum total of factors that are identified in our study to be the factors leading to leaveism.

CONCLUSION

An organization's success and effectiveness depend upon various factors, and employee well-being is one of the major factors that can affect the growth and development of an organization and its business. If employees' well-being is not taken care of, then it can lead to issues like stress and fatigue, low job satisfaction, employee attrition, etc., which negatively affect the organization's productivity. Thus, to identify and analyse what affects organizations effectiveness and employee well-being in a way that the factors known are not considered independently, but in totality, the study of Leaveism has been conducted. From this, it was concluded that leaveism is practiced in work context and there are various factors that lead to it. Some of the major factors are peer pressure, Organizational Citizenship behavior, fear of dismissal, working more to prove dedication, using free time, low individual productivity, employee gratification, work pile up, organizational structure, allocation of new task/project, technology and mundane work. When leaveism is practiced due to these factors its leads to consequences like work life imbalance, low quality Organizational Citizenship Behavior, family conflicts, ergonomic injuries, stress and fatigue, low job satisfaction, employee attrition and counter workplace behavior. Hence, it is important to address these issues to improve organisation's performance.

It is to be noted that if leaveism is practiced because of one's own willingness, it does not affect the organization negatively, but when it is practiced because of other factors, it becomes a problem that needs to

be addressed. Therefore, leaveism in work context should be considered and necessary managerial steps to minimize the presence of leaveismshould be taken, like restructuring the workplace environment in a way that does not pressurize employees, focusing on employee wellbeing and leave policies or framing a separate policy on leaveism itself in order to improve employee's as well as organisation's efficiency.

RECOMMENDATIONS/ MANAGERIAL IMPLICATIONS

- Various factors affecting leaveism that are identified through this research are found to negatively impact the employees, businesspeople, faculties, and students subconsciously. Organisation should make efforts to increase awareness about the term leaveism, the factors affecting it and the consequences it has so that the people at work understand the importance of changing their working methodology for their wellbeing. This shall ultimately result in improving the efficiency of the organisations.
- Many organisations have a method of working which creates an environment that pressurizes employees, students and faculties to work more (excluding the condition that they are willing to work more) which results into over-work and poor quality of work-life balance. Here it is important to restructure the way that particular organisation is working so that a conducive environment is created which shall improves the productivity of the people at work and therefore contribute to increased organisation's effectiveness.
- The practice of leaveism leads to low productivity of employees, businesspeople, faculties, and students and that affects the overall efficiency and development of an organisation. To reduce the impact of the consequences of leaveism identified through this research, organisations should step forward and redesign policies like sick leaves in such a way that it is ensured that the people at work take a definite number of minimum leaves as dictated by their national government. The organisation should not allow the employees or faculties to accumulate and encash these minimum leaves. It may further give an option of choosing to accumulate the extra leaves that it provides and encash it or use them whenever needed. This strategy can prove to be win-win for both.
- Managers should maintain healthy amount of competition throughout the organisation which is conducive for the development of the same. Excessive competition can lead to burnout and counter-productive workplace behaviour. Thus, it should be ensured that the employee gratification policies are not in direct relationship with the amount of time or the extra work put in by the worker beyond a certain limit. This will naturally result into less cut-throat competition and more fair competition that shall benefit the organisation overall.
- From our survey it was found that 70% of the organisations under consideration do have an employee well-being policy. But, because the

concept of leaveism is unknown to people and is quite subtle in nature, its consequences are not addressed in these policies. Thus, it is important to reframe certain policies to include points that limit the practice of leaveism at workplace.

- Apart from the changes in other policy, it would be more effective if a policy on leaveism is designed and implemented. This will help in holistically reducing the number of factors leading to it and decrease in the ill-effects of leaveism. This policy can include a major condition that promotes people to work in their official working hours only. This means that in the current era, when people are wasting quite a lot of time on social media and other things while working on their laptops at the workplace, it is essential to have a policy that keeps a check on the time wasted by the employees while at work so that they can be free from overwork post official working hours. This will directly result in reduction of the practice of leaveism.
- Lastly, an important way to address the problem of leaveism can be through Organisational Development (OD) intervention. As OD interventions are personalized solutions, it can precisely address the problem of leaveism at workplace, may it be at corporate houses, business houses, start-ups and educational institutions, as per the factors/situation at a particular place. This method shall help in creating a new system that can drastically reduce the practice of leaveism and hence improve organisations productivity and effectiveness.

LIMITATIONS OF THE STUDY

Following are the limitations of this study:

- Due to COVID-19 crises, simulation experiments and scenario based activities to map the behaviour of people could not be done which might have been useful to capture the honest thoughts and opinions of people, leading to new factors responsible for leaveism.
- As this cross-cultural study includes responses from over 22 countries across the globe, it was difficult to set up a common time to conduct a focused group discussion and hence a common medium to enquire people face to face could not be done especially due to COVID-19 crises.
- Due to COVID-19 crises, convenient sampling had to be followed as it was not possible to reach out to people from different sectors like finance and IT which are commonly seen as victims of leaveism.
- Online questionnaire was used to collect responses from people as an alternative method to map people's behaviour and opinions relating to leaveism. It was found that open ended questions were not answered by many respondents. Hence, the possibility to find new factors reduced.
- Reliability of the responses obtained might not be up to the mark because it is possible that the respondents might have misinterpreted

the questions or could have selected options haphazardly as the questionnaire was circulated online and it not possible to monitor the same.

- With sample size around 250 people, only a few new factors affecting leaveism could be identified. Had the sample size been larger, more factors leading to leaveism might have been known.

ANNEXURE

Table 1: One-way ANOVA for factor-Prove Dedication

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.102	1	.102	.400	.529
	Within Groups	19.885	78	.255		
	Total	19.987	79			
q56agegroup	Between Groups	.252	1	.252	1.233	.270
	Within Groups	15.948	78	.204		
	Total	16.200	79			
q1occupation	Between Groups	.000	1	.000	.	.
	Within Groups	.000	78	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	1.875	1	1.875	2.000	.161
	Within Groups	73.125	78	.937		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	.675	1	.675	.359	.551
	Within Groups	146.813	78	1.882		
	Total	147.488	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q4r2Provededication

Table 2: One-way ANOVA for factor-Peer Pressure

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.453	1	.453	1.810	.182
	Within Groups	19.534	78	.250		
	Total	19.988	79			
q56agegroup	Between Groups	.004	1	.004	.019	.891
	Within Groups	16.196	78	.208		
	Total	16.200	79			
q1occupation	Between Groups	.000	1	.000	.	.
	Within Groups	.000	78	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	2.451	1	2.451	2.635	.109
	Within Groups	72.549	78	.930		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	1.855	1	1.855	.994	.322
	Within Groups	145.632	78	1.867		
	Total	147.487	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q4r3PeerPressure

Table 3:One-way ANOVA for factor-Organizational Citizenship Behaviour

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.000	1	.000	.001	.972
	Within Groups	19.987	78	.256		
	Total	19.988	79			
q56agegroup	Between Groups	3.738	1	3.738	23.400	.000
	Within Groups	12.462	78	.160		
	Total	16.200	79			
q1occupation	Between Groups	.000	1	.000	.	.
	Within Groups	.000	78	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	.128	1	.128	.134	.716
	Within Groups	74.872	78	.960		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	.834	1	.834	.443	.507
	Within Groups	146.654	78	1.880		
	Total	147.487	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q5OrganizationalCitizenshipBehaviour

Table 4: One-way ANOVA for factor- Low Individual Productivity

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	1.154	2	.577	2.359	.101
	Within Groups	18.833	77	.245		
	Total	19.988	79			
q56agegroup	Between Groups	.377	2	.188	.917	.404
	Within Groups	15.823	77	.205		
	Total	16.200	79			
q1occupation	Between Groups	.000	2	.000	.	.
	Within Groups	.000	77	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	1.136	2	.568	.592	.556
	Within Groups	73.864	77	.959		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	1.968	2	.984	.521	.596
	Within Groups	145.520	77	1.890		
	Total	147.487	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q6Lowindividualproductivity

Table 5: One-way ANOVA for factor- Work as escape

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.419	2	.210	.825	.442
	Within Groups	19.568	77	.254		
	Total	19.987	79			
q56agegroup	Between Groups	.190	2	.095	.457	.635
	Within Groups	16.010	77	.208		
	Total	16.200	79			
q1occupation	Between Groups	.000	2	.000	.	.
	Within Groups	.000	77	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	.414	2	.207	.214	.808
	Within Groups	74.586	77	.969		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	1.442	2	.721	.380	.685
	Within Groups	146.045	77	1.897		
	Total	147.488	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q8workasescape

Table 6: One-way ANOVA for factor- Employee Gratification

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.655	2	.328	1.305	.277
	Within Groups	19.332	77	.251		
	Total	19.988	79			
q56agegroup	Between Groups	.012	2	.006	.028	.973
	Within Groups	16.188	77	.210		
	Total	16.200	79			
q1occupation	Between Groups	.000	2	.000	.	.
	Within Groups	.000	77	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	1.127	2	.564	.588	.558
	Within Groups	73.873	77	.959		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	4.657	2	2.329	1.255	.291
	Within Groups	142.830	77	1.855		
	Total	147.488	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q9employeegratification

Table 7: One-way ANOVA for factor- Stress and Fatigue

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	1.195	2	.597	2.447	.093
	Within Groups	18.793	77	.244		
	Total	19.988	79			
q56agegroup	Between Groups	.352	2	.176	.854	.430
	Within Groups	15.848	77	.206		
	Total	16.200	79			
q1occupation	Between Groups	.000	2	.000	.	.
	Within Groups	.000	77	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	.388	2	.194	.200	.819
	Within Groups	74.612	77	.969		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	1.819	2	.910	.481	.620
	Within Groups	145.668	77	1.892		
	Total	147.488	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q10stressandfatigue

Table 8: One-way ANOVA for impact of Leaveism

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.614	2	.307	1.220	.301
	Within Groups	19.374	77	.252		
	Total	19.987	79			
q56agegroup	Between Groups	.219	2	.110	.528	.592
	Within Groups	15.981	77	.208		
	Total	16.200	79			
q1occupation	Between Groups	.000	2	.000	.	.
	Within Groups	.000	77	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	.505	2	.252	.261	.771
	Within Groups	74.495	77	.967		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	6.284	2	3.142	1.713	.187
	Within Groups	141.204	77	1.834		
	Total	147.487	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q12impactofleaveism

Table 9: One-way ANOVA for factor- Free Time

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	1.188	5	.238	.935	.463
	Within Groups	18.799	74	.254		
	Total	19.987	79			
q56agegroup	Between Groups	.480	5	.096	.452	.811
	Within Groups	15.720	74	.212		
	Total	16.200	79			
q1occupation	Between Groups	.000	5	.000	.	.
	Within Groups	.000	74	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	1.507	5	.301	.304	.909
	Within Groups	73.493	74	.993		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	18.329	5	3.666	2.100	.075
	Within Groups	129.159	74	1.745		
	Total	147.488	79			

ONEWAY q55gender q56agegroup q1occupation q2r2doingextraworkathome q2r3workingonleavevacation BY q15r1freetime

Table 10: One-way ANOVA for practice of Leaveism

ONEWAY q55gender q56agegroup q1occupation

		Sum of Squares	df	Mean Square	F	Sig.
q55gender	Between Groups	.461	3	.154	.598	.618
	Within Groups	19.527	76	.257		
	Total	19.988	79			
q56agegroup	Between Groups	.296	3	.099	.472	.703
	Within Groups	15.904	76	.209		
	Total	16.200	79			
q1occupation	Between Groups	.000	3	.000	.	.
	Within Groups	.000	76	.000		
	Total	.000	79			
q2r2doingextraworkathome	Between Groups	1.812	3	.604	.627	.600
	Within Groups	73.188	76	.963		
	Total	75.000	79			
q2r3workingonleavevacation	Between Groups	5.331	3	1.777	.950	.421
	Within Groups	142.157	76	1.870		
	Total	147.488	79			

q2r2doingextraworkathome q2r3workingonleavevacation BY q13leaveism

Table 11: Paired Samples Correlations

		N	Correlation	Sig.
Pair 2	q44basisofemployement & q45r1unpaidovertime	110	.	.
Pair 3	q45r2takingworkhome & q45r3workingonleavevacation	110	-.061	.524
Pair 4	q46r1OCB & q46r2workequirement	110	.154	.107
Pair 5	q46r3workpileup & q46r4technology	110	.025	.797
Pair 6	q46r5lowproductivity & q46r6disability	110	.	.
Pair 7	q46r7selfservingmotives & q46r8selfservingmotives	110	.255	.007
Pair 8	q46r9fearofdismissal & q46r10Organizationalstructure	110	.014	.882
Pair 9	q46r11delegation & q46r12newtaskproject	110	.047	.625
Pair 10	q46r13workasescape & q46r14insecurity	110	.061	.527
Pair 11	q46r15workexpectation & q46r16worklifebalance	110	.139	.147
Pair 12	q46r17natureofjob & q46r18workload	110	-.044	.650
Pair 13	q46r19personalhabit & q46r20performancebasedtask	110	-.051	.599
Pair 14	q46r21workasescape & q46r22organizationalpolicy	110	.	.
Pair 15	q46r23workmartyr & q46r24others	110	-.030	.753
Pair 16	q47FlexiTime & q47r1FlexiTimeyes	110	-.012	.905
Pair 17	q47r2FlexiTimeno & q47r3FlexiTimeprefernottosay	110	-.056	.562

Pair 18	q49employeewellbeing & q50degreeofleaveism	110	.102	.287
Pair 19	q51sectorofemployement & q52r1freetime	110	.135	.160
Pair 20	q522r2workasescape & q52r3taskallocation	110	.	.
Pair 21	q52r4environment & q52r5personalsatisfaction	110	-.027	.781
Pair 22	q52r6notworkingduringthepandemic & q52r7otherreasons	110	-.069	.474

Table 12: Paired Samples Test

		Sig. (2-tailed)
Pair 2	q44basisofemployement - q45r1unpaidovertime	.000
Pair 3	q45r2takingworkhome - q45r3workingonleavevacation	.636
Pair 4	q46r1OCB - q46r2workequirement	.000
Pair 5	q46r3workpileup - q46r4technology	.811
Pair 6	q46r5lowproductivity - q46r6disability	.083
Pair 7	q46r7selfservingmotives - q46r8selfservingmotives	.009
Pair 8	q46r9fearofdismissal - q46r10Organizationalstructure	.242
Pair 9	q46r11delegation - q46r12newtaskproject	.341
Pair 10	q46r13workasescape - q46r14insecurity	.000
Pair 11	q46r15workexpectation - q46r16worklifebalance	.009
Pair 12	q46r17natureofjob - q46r18workload	.000
Pair 13	q46r19personalhabit - q46r20performancebasedtask	.412
Pair 14	q46r21workasescape - q46r22organizationalpolicy	.320
Pair 15	q46r23workmartyr - q46r24others	.006
Pair 16	q47FlexiTime - q47r1FlexiTimeyes	.000
Pair 17	q47r2FlexiTimeno - q47r3FlexiTimeprefernottosay	.123
Pair 18	q49employeewellbeing - q50degreeofleaveism	.000
Pair 19	q51sectorofemployement - q52r1freetime	.000
Pair 20	q522r2workasescape - q52r3taskallocation	.000
Pair 21	q52r4environment - q52r5personalsatisfaction	.028
Pair 22	q52r6notworkingduringthepandemic - q52r7otherreasons	.312

BIBLIOGRAPHY

Allen, H., Slavin, T., & Bunn, W. I. (2007). Do long work hours impact health, safety, and productivity at a heavy manufacturer. *journal of occupational & environmental medicine vol 49*, 148-171.

ANOVA Test: Definition, Types, Examples. (2004). Retrieved from Statistics How To: https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/

Aronsonn, G., Gustafsson, K., &Dallner, M. (2000). Sick but yet at work. An Empirical study of sickness presenteeism. *Journal of epidimiology and community health*, 502-509.

Bandura, A., & Locke, E. (2003). Negative self-efficacy and goal effects revisited. *Journal of applied psychology*, 87-99.

Barnett, R. C. (2004). *Work Hours as a predictor of stress outcomes.*

Retrieved from http://www.cdc.gov/niosh/topics/workschedules/abstracts/barnett.html

Bass, J. (1998). Structured Experiences, Role Plays, Case Studies, Simulations and Games at Pfeiffer & Company. *21*(2nd).

Bell, N. F., Hart, D. A., & Robert. (1999). Unpaid Work. *Economica, New Series*, 271-290.

Bolino, M., Klotz, A., Turnley, W., & Harvey, J. (2013). Exploring the dark side of organizational citizenship behavior. *Journal of Organizational Behavior*, 542- 559.

Bolino, M., Turnley, W., &Niehoff, B. (2004). The other side of the story: Reexamining Prevailing Assumptions about Organizational Citizenship Behaviour. *Human resource management review*, 229-246.

Boogaard, K. (2019, June 12). Why Do You Get Vacation Guilt.

Brief, A., &Motiwidlo, S. (1986). Prosocial Organizational Behaviours. *Academy of management Review*, 710-725.

Casner-Lotto, F. a. (2003). *Overtime and Wellbeing at home.*Institute of Workspace Studies .

Chaitanya, S., & Tripathi, N. (2001). Dimensions of Organisational Citizenship Behaviour. *Indian Journal of Industrial Relations*, pp. 217-230.

Crystal Raypole, H. (2019, April 29). How to Identify and Manage Workplace Bullying.

Ferris, G. R. (2002). Perception of organisational politics. *Theory and researchdirections*.

George, J., & Brief, A. (1992). Feeling good- Doing good: A conceptual analysis of the mood at work- organizational spontaneity relationship. *Psychological Blletin*, 310- 329.

Gliem, J. A., &Gliem, R. R. (2003). *Calculating, Interpreting, and Reporting*

Cronbach's Alpha Reliability Coefficient for Likert-Type Scales.

Golden, L. (2006). *The effects of working time on productivity and firm performance.*

Graham, J. (1991). An essay on organizational citizenship behaviours. *Employee responsibilities and rights journal*, 249-270.

Hart, D. N. (1999). *Unpaid work.*Wiley on behalf of The London School of Economics and Political Science and The Suntory and Toyota International Centres for Economics and Related Discipline.

Hayes, A. (2019, November 16). Harmonic Mean.

Health assured. (2018, August 29). The Effects Of leaveism. Heppner, R. S. (2013). *The Ideal Worker.* Springer Link

Hesketh, I., & C., C. (2014). Leaveism and public sector reform: will the practice continue?*Research gate*.

Hesketh, I., Cooper, C., & Ivy, J. (2014). Leaveism and Work–Life Integration: The Thinning Blue Line?*policing: A Journal of Policy and Practice*, 183–194.

Ian Hesketh, C. C. (2014). *Leaveism and public sector reform: Will the practice continue?*Research gate.

Johns, G. (2009). Presenteeism in the workplace: a review and research agenda.*journal of organizational behaviour*, 519-542.

Kenton, W. (2019, June 04). Absenteeism.

Koopmanschap, M., Burdorf A., J., K., M., W. J., B. W., &Severens, H. (2005). Measuring productivity changes in economic evaluation. *pharmacoeconomics*, 47-54.

Leondhart, M. (2019, April 27). *Only 28% of Americans plan to max out their vacation days this year*.

Retrieved from CNBC- Make it: https://www.cnbc.com/2019/04/26/only-28percent-of- americans-plan-to-max-out-their-vacation-days-this-year.html

Linn, V. D., Jill W., G., & Richard, M. D. (1994). Organizational Citizenship Behavior: Construct Redefinition, Measurement, and Validation. *The Academy of Management Journal, Vol. 37*, 765-802.

Lucy P., E., Wulff, P., & Sabrina. (2010). Bringing work home: implications for BLS productivity measures. *Monthly Labor Review*, 18-35.

McClintock-Comeaux, M. (2013). *Ideal Worker.* SAGE Knowledge

One-way ANOVA in SPSS. (2007). Retrieved from Laerd Statistics: https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

Palo, s., &Pati, s. (2013). The Determinants of Sickness Presenteeism. *Indian journal of Industrial relations*, 256-269.

Parli, K. (2018). Presenteeism, Its Effects and Costs: A Discussion in a Labour Law Perspective.

International Journal of Comparative Labour Law and Industrial Relations, 53-75.

Richard G., N., James, S. B., Daryl O., M., & Robert, M. (1997). An Investigation into the Antecedents of Organizational Citizenship Behaviors in a Personal Selling Context. *Journal of Marketing*, 85-98.

Robinson, S. L. (1995). Psychological contracts and OCB: The effect of unfulfilled obligations on civic virtue behaviour. *Journal of organizational behaviour*, 289-298.

Rose, A. M. (1967). Sociological factors affecting economic development.*Springerlink*, 169-183.

StatisticsSolutions. (2019). AMOS & SPS. StatisticsSolutions. (2019). Exploratory Factor Anlaysis.

Staw, B., &Boettger, R. (1990). Task Revision: A neglected form of work Performance.*Academy of management journal*, 534-559.

Tavakol, M., &Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*.

Timely Blog. (2018, July 24). The anxiety trade-off: should you work on vacation? Oslo, Norway.

Stephanie. (2011). Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy. *Statistics How To*.

Vaynshteyn, G. (2019, October 28). *'Work Guilt' Is Real and It's Keeping You From Using Your Vacation Days*. Retrieved from appartment therapy: https://www.apartmenttherapy.com/work-guilt-not-taking-vacation-36661480

Vice, Michaela Cross. (2018, December 07). Why You Have Vacation Guilt.

Yammarino, J. (n.d.). Research in multi-level issues. *Oxford, UK: Elsevier science*, 179-254. Yvette Brazier, M. (20018, February 01). What is psychology and what does it involve?

DECISION SCIENCES INSTITUTE
Why do employees withhold knowledge: individual and organizational antecedents

Tony Tung
CASE Western University
tht11@case.edu

Kalle Lyytinen
CASE Western University
kjl13@case.edu

James Gaskin
Brigham Young University
james.eric.gaskin@gmail.com

ABSTRACT

With increased specialization comes increased potential for silos of knowledge and expertise. In such an environment, knowledge sharing becomes a vital component of productive processes, while knowledge hiding may lead to more detrimental outcomes. It is, therefore, critical to better understand why, in today's prevalently sharing society, some employees intentionally or inadvertently engage in knowledge hiding. A better understanding in this regard should help prevent and mitigate the negative consequences of such behavior. We, therefore, investigate the antecedents of knowledge hiding. The study was conducted with 310 full-time job employees across many industries. Our study draws from extant theories of employee behavior, integrating both organizational identification and perceived organizational support to explain why knowledge hiding behavior arises in an organization.

KEYWORDS: knowledge sharing, knowledge hiding, knowledge management, perceived organizational support, psychological safety, knowledge-sharing climate, organizational identification

INTRODUCTION

Knowledge management is an important activity for companies in contemporary, knowledge intensive and highly competitive environments. Knowledge management concerns the acquisition, sharing, transfer, storage, extraction, re-creation, and integration of knowledge. It helps companies identify and address risks, improve efficiency and effectiveness, and pursue product and process innovation (Marin, Cordier, & Hameed, 2016). Various knowledge management activities are indispensable for a firm's competitiveness and viability. Therefore, focusing on knowledge management has become a strategic prerogative. One element of the prerogative is how to encourage the sharing of knowledge between different organizational members (Intezari, Taskin, & Pauleen, 2017; Jeung, Yoon, & Choi, 2017).

Generally, knowledge sharing refers to organizational processes by which a *knowledge provider* transfers or shares his/her knowledge (e.g., professional knowledge, ideas, insights, or implicit or explicit knowledge) to the *knowledge receiver* (Rusly, Corner, & Sun, 2012; Teng & Song, 2011; Titi Amayah, 2013). Usually, the knowledge provider "controls the recipient's ability to know and access the provided content while customizing the shared knowledge to the recipient's needs and situation" (Ho & Ganesan, 2013 p. 93). According to Beck, Pahlke, & Seebach (2014), organizations do not "own" the intellectual assets of their employees and

therefore cannot force employees to pass their knowledge on to other members of the organization (Connelly, Zweig, Webster, & Trougakos 2012). Peng (2013) pursued this idea and applied psychological ownership theory to explain that employees have a right also to take any action to protect their knowledge including hiding knowledge- i.e. not displaying what they know.

In consequence, employees will generally have a disposition to intentionally or inadvertently refrain from sharing their knowledge. Instead, they will hide what they know (called knowledge hiding), because of they feel that they have the right to hide as they own their knowledge. Per Connelly et al. (2012), we define knowledge hiding as any intentional attempt made by an individual to withhold or conceal knowledge when they receive requests from knowledge seekers. For example, a knowledge seeker who has missed an important meeting may ask a colleague who attended that meeting for information about it. The knowledge provider may intentionally withhold the information the knowledge seeker is requesting or convey false information, because they are concerned with losing their organizational value and they can sustain it by controlling the knowledge. Also, the knowledge providers are concerned that their performance will be reviewed based on the knowledge they possess providing a disincentive to share such knowledge (Peng, 2013; Peng, Wang, & Chen, 2019). Although this seems dysfunctional for the firm, the occurrence of knowledge hiding is sufficiently common to the extent that Zhao, Xia, He, and Sheard (2016) define knowledge hiding as workplace ostracism.

Despite some plainly selfish motives, knowledge hiding can also be triggered by positive intentions. For example, employees may engage in knowledge hiding behavior to protect the feelings of other groups, adhere to confidentiality obligations, or protect the interests of third parties (Connelly & Zweig, 2015). In focusing on the knowledge-hiding behavior of knowledge providers, the scope of our study is limited to the inappropriate and unethical hiding behavior of knowledge hiders rather than situations where it is driven by possible acceptable and good intentions.

Several studies (e.g., Hajro, Gibson, & Pudelko, 2017; Trialih, Wei, & Anugrah, 2017) have explored factors that promote knowledge sharing. At the same few studies have explored why employees hide knowledge (Labafi, 2017) and the factors are not likely to be symmetrical. Knowledge hiding is also common. A survey of more than 1,700 subjects in Canada showed that 76% of employees hide their knowledge from colleagues (Connelly et al., 2012; The Globe and Mail, 2006). A study recent indicate that 60% of employees have had experiences where their coworkers hide knowledge vital to their work, and found that ineffective knowledge-sharing behavior cost more than $47 million of damage annually (Jiang, 2019). Connelly, Černe, Dysvik, & Škerlavaj (2019) note that knowledge hiding also inflicts damage on an organization, such as decreasing both organizational and individual levels of performance, and it decreases individuals' knowledge-sharing intentions in the future. (Abubakar, Behravesh, Rezapouraghdam, & Yildiz, 2019; Connelly, & Zweig, 2015; Pan, Zhang, Teo, & Lim, 2018; Zhao, Xia, He, Sheard, & Wan, 2019) by creating distrust between peers (Connelly et al., 2012). Therefore, we need to identify factors that reduce dispositions to hide knowledge.

Generally, knowledge hiding damages organizations and individuals, often characterized as a lose-lose situation. Knowledge-hiding negatively impacts organizational cohesion, creativity, innovative work, performance, and trust (Connelly & Zweig, 2015; Connelly, Černe, Dysvik, & Škerlavaj, 2019). Previous research (Tsay, Lin, Yoon, & Huang, 2014) has also pointed out that knowledge withholding by peers causes negative relationships among peers resulting in ostracism (Zhao et al., 2019). Therefore, it is important to find out not only why and how individuals hide their knowledge but what are the organizational factors that affect knowledge hiding. Understanding factors that explain why peers hide the knowledge can later aid in creating successful knowledge management interventions. Our primary research question in this study is: *What organizational and individual factors affect knowledge hiding?*

By addressing this question our study begins to fill the research gap related to organizational and individual antecedents of knowledge hiding. The objective of the study is to explain knowledge hiding by formulating a more comprehensive model of why knowledge hiding occurs in organizations and how the organization might prevent it. We first leverage individual (personality and psychology) factors to explain knowledge hiding. Beyond psychological ownership theory, this study leverages the perceived organizational support theory (POS), which involves employees' perceptions and psychology concerning how they feel in their workplace. In addressing the questions we also draw upon extant literature on knowledge management to identify salient organizational factors that drive knowledge hiding. Among them we posit that sharing climate and organizational support have a significant role (Connelly et al., 2012; Singh, 2019). Among individual factors we consider selfishness, psychological safety, seniority, identification with the organization, altruism, and basic demographics.

This study uses a quantitative, survey based approach, appropriate to confirm and extend previous studies (Connelly et al., 2012; Connelly & Zweig, 2015) to test our theory. Our target population for study only requires that subjects be 18 years old and have a full-time job. We evaluate and analyze our data with the IBM SPSS (v.25) and AMOS (v.25).

The remainder of the paper is structured as follows. We next present a conceptual framework to explain factors that influence knowledge-hiding behaviors. Then, we formulate our theoretical model and hypotheses. Subsequently, we use structural equation modelling to validate the hypotheses. We finish the study by reporting the results and discuss their implications for further theory development and practice.

THEORETICAL DEVELOPMENT/MODEL

Connelly et al. (2012) were the pioneers in defining the knowledge-hiding issue and built up a multidimensional measure of this construct. Černe, Nerstad, Dysvik, and Škerlavaj (2014) emphasized that knowledge hiding hurt both knowledge providers and receivers with poor creativity. In an effort to clarify what is meant by knowledge hiding, consider the following example. John was Mary's manager before the reorganization of the company. Realizing that Mary would take over John's role before the reorganization, John renewed a big contract a few months before the reorganization. However, the contract would be expired after the effective date of the reorganization. Then, John carried this renewed contract with him to his new position without any notification to Mary. Even though John's new role and responsibilities do not require this contract, he still wants to maintain ownership of the contract. When leaders like John behave in this way, they set a negative example for employees. This knowledge hiding can involve both direct wrongdoing and a range of negative behaviors.

A few studies have examined why and when employees hide knowledge (Connelly & Zweig, 2015; Peng, 2013; Zhao, Liu. Li, and Yu, 2019). Originally, a qualitative study confirmed that individuals, to maintain their own benefits, prefer to hide their knowledge (Labafi, 2017). Geofroy and Evans (2017) pointed out that increased teamwork, trust, and organizational commitment can reduce knowledge hiding in the organization. By adopting the psychological contract theory, Pan, Zhang, Teo, and Lim (2018) investigated the perceptions personality traits create about knowledge hiding, which was one of three studies that discussed knowledge hiding from a psychological point of view. Zhao et al., (2019) applied social identity theory to build a theoretical model linking leader-member exchange (LMX) and knowledge hiding. The concept of leader-signaled knowledge hiding (LSKH) has been established from social learning theory by Offergelt, Spörrle, Moser, and Shaw (2019).

Scholars believe organizations should create internal policies to reward those knowledge providers who like to share with knowledge receivers (Abubakar et al., 2019; Lu, Singh, & Sun, 2015; Peng, Wang, & Chen, 2019). This study argues the organization could prevent

knowledge-hiding behavior if the management level can satisfy employees' spiritual and emotional dimensions. When employees feel that the organization truly cares, they not only don't hide knowledge to hurt the organization, but they give back to the organization with their enthusiasm and performance. Knowledge-hiding theory is still new, and there are few studies that mention the individual psychological level (Jiang, Hu, Wang, & Jiang, 2019; Pan et al., 2018; Peng, Wang, & Chen, 2019).

If the knowledge provider can control the distribution or diffusion of their knowledge, it is important for the knowledge provider to have the ability to negotiate and bargain with the organization, such as through promotions, positions, and rewards. There are difficulties to obtaining solid evidence to use to punish individuals who hide their knowledge. For example, knowledge hiders often provide incomplete information, engage in rationalized hiding, hide evasively, and play dumb to prevent the investigation from reaching the management level (Connelly et al., 2012; Ladan, Nordin, & Belal, 2017).

This study categorizes both POS and knowledge-sharing climate (KSC) as organizational-level variables and all other variables, like selfishness, psychological safety, employee seniority, organizational identification, and knowledge hiding, as individual-level variables. As shown in figure 1, this study defines the knowledge hiding as our dependent variable and will evaluate the theoretical model through personal traits, motivational factors, and organizational context. This study examines as how these factors interact in a specific setting, such as organizational identification, perceived organizational support, and psychological safety, which are assumed to influence knowledge hiding as our outcome variable. Figure 1 shows our conceptual model.

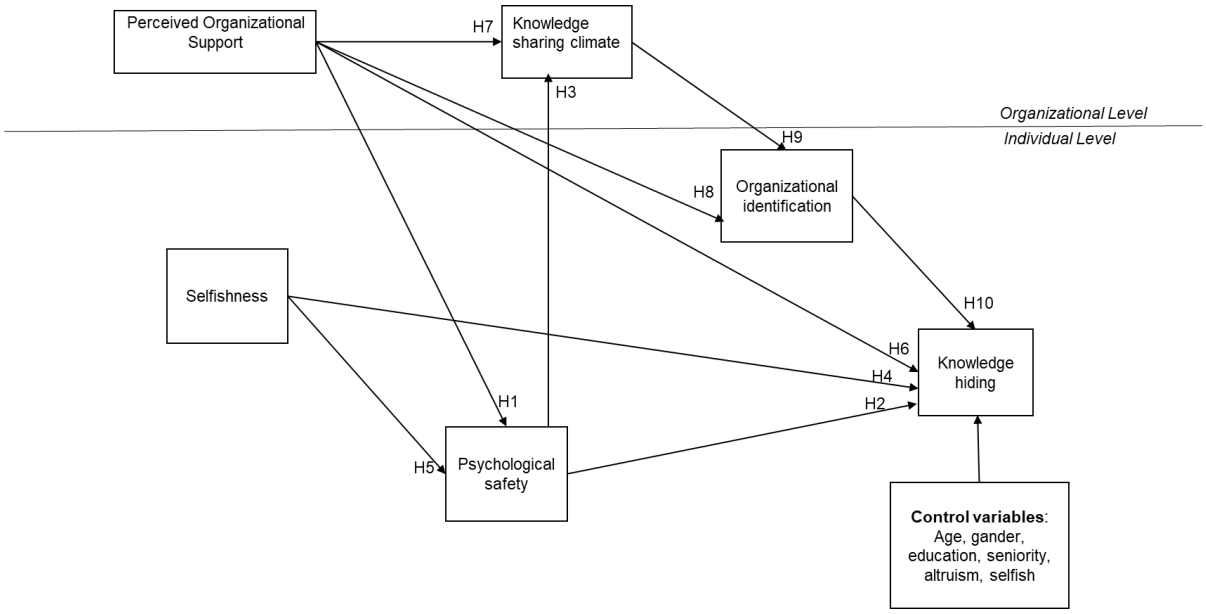


Figure 1. Conceptual Model

Psychological Safety

This study defines psychological safety (PS) as present when staff feel that they can express themselves to managers without any negative consequences, such as retaliation, that

might threaten their organizational position (Liu, Song, Li, & Liao, 2017). Jiang, Hu, Wang, and Jiang (2019) emphasized that PS implies that when staff make mistakes or provide different opinions than those of the management level, they do not feel they will be denied or turned down for help if they need it.

It is important for employees to experience confidence, safety, and freedom in their professions, especially when seeking to satisfy management's expectations. An example of a situation that would lead to a lack of psychological safety, is if management mentions some unfavorable information socially, such as a layoff or furlough that might makes individuals feel nervous, anxious, or psychologically unsafe about experiencing these negative consequences (Peng, Wang, & Chen, 2019).

Another important and significant factor is the intensification of social competition. A firm that makes increasing demands of its employees will increase competition between employees. When employees lack a sense of professional security, they might feel afraid that they don't have control over their situation anymore. Because a sense of security is a basic need of human beings (Bouzenita, & Boulanouar, 2016), when employees have a sense of security and belonging can they devote themselves wholeheartedly to achieve better performance in their work and not be distracted by feelings of insecurity.

POS indicates that employees develop general beliefs regarding the extent to which their organization values and appreciates their involvement and is concerned about their well-being (Eisenberger, Cummings, Armeli, & Lynch, 1997; Eisenberger, Huntington, Hutchison, & Sowa, 1986; Rhoades & Eisenberger, 2002). From a psychological safety point of view, POS is about feeling comfortable enough to take interpersonal risks in the organization. Scholars also have explained that POS stems from the employees' experience of organizational policies, norms, regulations, and procedures. More studies show that POS plays a significant role in meeting individuals' needs for respect and endorsement (Eisenberger et al., 1986). Individuals feel that organizations value their contributions and give more when they think they have a more secure position in the organization.

Staff members who perceive high organizational support have positive expectations about the organization's likely reaction to their efforts as well as their mistakes, and thus they have less reason to fear harmful penalties as a result of minor faults in their work roles (Edmondson, 1999). This is because employees feel that the organization appreciates their contributions and efforts, which in turn increases employees' tolerance for making mistakes. Therefore, individuals can feel that their future carrier is not threatened. This is the other aspect of PS for an individual. Thus, we hypothesize the following:

H1: Perceived organizational support has a direct positive relationship with individual psychological safety.

Knowledge Hiding

As noted, this study defines knowledge hiding as the intentional concealment of knowledge that a knowledge seeker requests. Such concealment may occur for many different reasons (prosocial motivation, instrumental motivation, laziness, etc.). Knowledge hiding, which does not necessarily stem from harmful intentions, also differs from repulsive workplace behavior, workplace infringement, and social destruction in the workplace. It does not necessarily involve intentions to cause injury. Knowledge hiding is not necessarily deceptive, although some people think that deception and knowledge hiding are similar. Deception refers to a sender passing a message intended to cause the receiver to establish a wrong belief or conclusion, which can be achieved through means such as distortion, ambiguity, and concealment, whereas hiding involves refraining from sending any message.

Staff without PS are reluctant to share knowledge, thoughts, and ideas with colleagues without regard to recourse (Peng, Wang, & Chen, 2019). However, when staff feel threatened and insecure in their organization, they feel their effort and contributions to the organization are not valued and appreciated. To get ahead of other competitors in the organization, individuals may be more prone to hide because knowledge hiding leads to the underdevelopment of other individuals; the knowledge hider believes they ultimately lose the competitive advantage if the knowledge receiver gains more knowledge continually (Labafi, 2017; Peng, 2013).

The first step they take is to protect themselves against job insecurity; therefore, hiding their knowledge will appear to be a good choice. On the other way, staff with high levels of PS may likely be more concerned with both organizational performance and benefits than with hiding knowledge for their own benefit (Miao, Eva, Newman, & Cooper, 2019). We propose the following:

Hypothesis 2: Psychological safety effectively decreases knowledge hiding.

The knowledge-sharing climate is the employees' shared perception of the state of knowledge sharing within the organization at the cultural level, which influences individual behavior, and affects the individual's sense of self (Li, Zhu, & Luo, 2010). Because individuals recognize that their future careers are less threatened, they would be more willing to help the organization build a knowledge-sharing climate. At the same time, the greater the sense of PS, the more dedicated employees are to helping the management team to create a KSC.

H3: An individual's psychological safety has a direct positive relationship with a knowledge-sharing climate.

Selfishness

At an individual level, selfishness occurs when a staff member puts his/her own considerations above the interests of other peers in the organization (Ritzenhöfer, Brosi, Spörrle, & Welpe, 2019; Takaki, Taniguchi, & Fujii, 2016). Recent studies (Ritzenhöfer et al., 2019; Mok, & De Cremer, 2018; Yam, & Reynolds, 2016) agree that selfishness is an unethical business behavior that influences organizational performance. For example, some knowledge receivers try to obtain as much knowledge as possible from knowledge providers in their interaction and pay as little as possible for that knowledge. The selfish desires of staff members often manifest as greed (Farrugia, 2002; Ives, Freeth, & Fischer, 2020). When the working environment is interdependent and requires collaboration, coworkers may dislike an individual's hiding behaviors because hiding critical knowledge on purpose is a selfish and unethical behavior (Evans, Hendron, & Oldroyd, 2014; Zhu, Chen, Wang, Jin, & Wang, 2019). (Evans, Hendron, & Oldroyd, 2014; Zhu et al., 2019). From the knowledge receivers' view, "Knowledge hiding is driven by feeling of egoism, selfishness and greed accompanied by a rational assessment of costs and benefits of the action or non-action" (Serenko & Bontis, 2016, p. 1215). Because individuals consider their own interests first, hiding their knowledge to increase their competitive advantage is perceived as a necessary behavior. Thus, we hypothesize the following:

H4: Selfishness in the workplace increases knowledge-hiding behavior.

Individuals whose behaviors considered only their interests and ignoring others' interests are exhibiting selfishness. Individuals who behave more selfishly obtain benefits and take advantage of their peers; they feel more comfortable questioning others in the organization because they believe they will have a greater competitive advantage in the organization. In the

modern organizational structure, the high-stress, high-competition environment only tends to make us more proactive in protecting our turf, being aware of our surroundings, and being vigilant and proactive in eliminating "threats" that should be seen as meaningless. Back to our previous example, John hides that he used Mary's resources to renew the contract to consolidate his new position, which was a selfish behavior. The more resources he takes from Mary, the more stable he feels in his new role.

The more benefits they get from their peers because of their selfishly, they would feel relaxed, become more open, more engaged, and more vocal, and they have less threat in their career.

Therefore, we propose the following:

H5: When an individual's level of attitude of selfishness in the workplace is high, psychological safety becomes significantly stronger.

Perceived Organizational Support (POS)

POS refers to employees' general beliefs regarding the extent to which their organization values and appreciates their involvement and cares about their well-being (Eisenbergeret al., 1997; Eisenberger et al., 1986; Eisenberger, Stinglhamber, Vandenberghe, Sucharski, & Rhoades, 2002). From a psychological safety perspective, POS is about feeling comfortable enough to take interpersonal risks in the organization. Scholars have also explained that POS stems from employees' experience of organizational policies, norms, regulations, and procedures. Studies increasingly show that POS plays a significant role in meeting employees' needs for respect and endorsement (Eisenberger et al., 1986; Shore & Shore, 1995). POS provides a foundation for reciprocally beneficial social exchange between team members and organizations (Eisenberger et al., 1986), with employees reciprocating the support received in various ways, including more exceptional job performance, loyalty, organizational citizenship behavior, and organizational commitment (Eisenberger, Fasolo, & Davis- LaMastro, 1990; Eisenberger et al., 1986; Shore & Wayne, 1993). POS provides a foundation for reciprocally beneficial social exchange between team members and organizations (Eisenberger et al., 1986), with employees reciprocating support received in various ways, including greater job performance, loyalty, organizational citizenship behavior, and organizational commitment (Eisenberger et al., 1990; Eisenberger et al., 1986; Shore & Wayne, 1993). When staff perceive kind concern from their employer, they are not motivated to engage in unethical behavior such as hiding their knowledge from peers in the organization. Therefore, this study hypothesizes the following.

H6: Perceived organizational support has a direct negative relationship with knowledge-hiding behavior.

Knowledge-Sharing Climate (KSC)

This study defines KSC as a behavior norm in which knowledge providers share their perceptions and expectations regarding knowledge sharing with knowledge receivers (Buvik, & Tvedt, 2017). It is a critical goal for the management team to build a KSC because studies indicate that a KSC positively influences job performance and creativity (Alzghoul, Elrehail, Emeagwali Okechukwu, & AlShboul Mohammad, 2018). Gupta (2008) defined a framework (Figure 2) to explain the knowledge-sharing climate, which includes three major categories: human values, knowledge management consequences, and knowledge management. The human values category, which includes integrity, innovation, creativity, and cost and time consciousness, helps management levels to generate a knowledge-based culture (Gupta,

2008). The practice of human values creates many outcomes, such as knowledge generation, knowledge sharing, cost and time consciousness, and employee/customer satisfaction (Gupta, 2008). Knowledge management as an organizational climate links human values and consequences (Gupta, 2008).

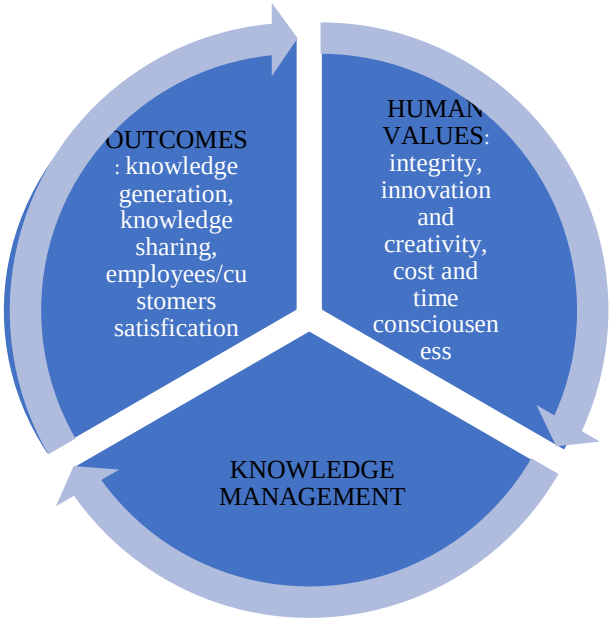


Figure 2. Framework of knowledge sharing climate

An organization's KSC influences staff members' decisions to share knowledge with coworkers. Bock, Zmud, Kim, and Lee (2005) indicated that social norms positively influence staff's knowledge-sharing intentions. As noted by Constant, Kiesler, and Sproull (1994, p. 404), "believing that information sharing is usual, correct, and socially expected workplace behavior should increase information sharing, independent of the information possessor's personal feelings about his or her coworkers." When employees have a positive perception of organizational support, they also have more positive views and beliefs about the organization itself. This positive belief makes it easier for employees to contribute to the mission of helping the organization build a KSC. Thus, we propose the following:

H7: Perceived organizational support has a direct positive relationship with a knowledge-sharing climate.

Organizational Identification (OI)

"Organizational identification (OI)…is the extent to which employees define themselves in terms of their organization and derive value from that self-definition" (Cornwell, Howard-Grenville, & Hampel, 2018). Employees' behaviors and ideas are consistent with the organizations they join in many respects, and employees feel that they have both a rational contract and a sense of responsibility in their organization, as well as irrational attribution and dependence (Hatch, & Schultz, 1997). Higher levels of OI are more important than individual self-concept but also more likely to lead to employee adherence to group norms (van Gils, Hogg, Quaquebeke, & Knippenberg, 2017). Employees' POS has a positive impact on their emotional and organizational commitment, indicating that with an increase in organizational support, employees' cohesion and organizational identity also increase accordingly (Tremblay,

Gaudet, & Vandenberghe, 2019). In addition, past research (Kurtessis, Eisenberger, Ford, Buffardi, Stewart, & Adis, 2017) confirms that the sense of organizational support has a positive impact on employee job satisfaction.

H8: Perceived organizational support has a direct positive relationship with organizational identification.

A KSC establishes an organizational culture that builds a trusting relationship between the organization and the individuals, along with an open and learning-oriented workplace on the management level (Li et al., 2010; Yoo, 2017). Individuals realize the KSC plays a significant role in knowledge management to create a fantastic working environment to which they would be proud to contribute their efforts. Once the KSC has been established, not only do individuals have more confidence to work with their peers, but the organization also will provide more understanding and tolerance if they make a mistake because the knowledge is free flowing (Li et al., 2010). This working climate makes individuals feel both a rational sense of contract and responsibility and an irrational sense of belonging and dependence on the organization.

Hypothesis 9: A knowledge-sharing climate is positively related to organizational identification.

Based on social identity theory, scholars believe that when staff have a high level of OI, turnover rates are low; by contrast, when staff have a low level of OI, turnover rates are relatively high (Conroy, Becker, & Menges, 2017; Pan, Gruber, & Binder, 2019). When staff identify with their organization more, they realize that hiding their knowledge always reduces organizational value (citation needed). Therefore, these staff not only refuse to hide and hoard their knowledge but also feel more pressure and responsibility to engage in building a KSC as an essential factor for the success of the entire organization (citation needed).

H10: An individual's organizational identification has a direct positive relationship with knowledge-hiding behavior.

METHODOLOGY

Measurement Development

This study used a quantitative survey methodology to help identify the dimensionality of the knowledge-hiding items, examine the psychometric properties of the knowledge-hiding scale, and test the discriminant validity of the knowledge-hiding construct. The questionnaire items were measured by using a Likert scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = disagree, and 5 = strongly agree). The study was conducted with adults 18 years of age or older with full-time job employment. Appendix A summarizes the construct definitions and example measures, also described below.

Psychological safety. Staff members reported their extent of PS during each interaction with management. The survey used five positively described items adapted from the scale proposed by Liang, Farh, & Farh (2012). A sample item was “In the interaction with the leader just now, I feel that expressing my true opinions is welcomed by management”.

Extrinsic Rewards. Staff members reported the extent of their extrinsic rewards during each interaction with other coworkers and management in daily operations. The survey used five positively described items adapted from the scale proposed by Malik, Butt, & Choi (2015). A sample item is “When I perform creatively, I receive financial rewards, such as incentives or bonuses”.

Knowledge hiding. Staff members reported the extent of their knowledge-hiding intention during each interaction with other coworkers in daily operations. The survey used five positively described items adapted from the scale proposed by Connelly et al. (2012). A sample item is "If I needed to work with this coworker in the future, I would likely keep my ideas to myself if this person asked for my help".

Knowledge-sharing climate. Staff members reported the extent of the KSC during each interaction with other coworkers in daily operations. The survey used five positively described items adapted from the scale proposed by Connelly, Kelloway (2003). A sample item is "People in this organization share their ideas openly".

Perceived Organizational Support. Staff members reported the extent of POS during each interaction with other coworkers and management in daily operations. The survey used five positively described items adapted from the scales proposed by Eisenberger et al. (1986) and Scott, Zagenczyk, Schippers, Purvis, & Cruz (2014). A sample item is "My organization really cares about my well-being".

Pre-testing (Q-sort and Pilot study)

We created a Q-sort (Funder, Furr, & Colvin, 2000) of the survey questions in Qualtrics with 31 items sorted into 8 different groups. The purpose of the Q-sort test was to examine our survey questions to determine which measures are appropriate for a specific group. All 11 participates had been trained and used this technique before. The placement hit ratio ranged from 60% to 90% for knowledge hiding items, 63.64% to 72.73% for psychological safety, 54.55% to 81.82% for organizational identification, 81.82% to 90.91% for KSC, and 54.55% to 72.73% for POS. All incorrectly answered questions were revised, and no questions were deleted. We modified the wording of a few questions to make them easier to understand, although all were used in previous studies.

In addition, we conducted a pilot test (Gamito, Oliveira, Rosa, Morais, Duarte, Oliveira, & Saraiva, 2010) with 51 respondents through Qualtrics. The Kaiser-Meyer-Olkin (KMO) test for measuring sampling adequacy in the responses yielded a value of .653, greater than the minimum recommended value of .50 for proceeding with this test (Kaiser, 1974). Reliability was also confirmed, as all values were above the acceptable Cronbach's alpha value of .7 (Hair, Black, Babin, & Anderson, 2010).

Control Variables

Control variables – age, gender, education, and team size – were incorporated into the model. These variables were thought to have some impact on both dependent variables. Control variables – age, gender, education, and team size – were incorporated into the model. There are many studies that have provided that individual background differences, such as age, gender, education, may impact organizational behaviors (Jiang, Hu, Wang, & Jiang, 2019; Peng, 2013). Age was coded as 1= 18-30 years old, 2= 31-40 years old, 3= 41- 50 years old, 4= 51-60 years old, and 5= 61+ years old. Gender was coded as 0= male, 1=female, and 3= prefer not to say. Education was code 1= less than high school, 2= high school, 3- college, 4= postgraduate, and 5= Ph.D. Team size was asked the number of full-time employees on the team, and it was coded 1= below 5, 2= 6-10, 3= 11-19, 4= 20-29, and 5= 30+.

Data Collection and Sample

We used Qualtrics to collect our samples through online surveys. The data were collected in November 2019 after the IRB approved our study. Table 1 describes the basic demographic profile of the 310 respondents, who were all at least 18 years of age and had a full-time job when responding to the survey.

Table1 Demographic profile of the respondents		
	Variable	%
Gender	Male	58.7
	Female	40.4
	Prefer not to say	.9
Education Background	Less than high school	1.3
	High school	28.4
	College	51.3
	Postgraduate	15.5
	Ph.D.	3.2
Age	18-30	30.0
	31-40	32.8
	41-50	21.1
	51-60	12.3
	61+	3.8
Primary industry	Education	12.3
	Information technology	12.9
	Government	6.6
	Retail	13.2
	Health care	11.7
	Manufacturing	12
	Agriculture	2.8
	Utilities	1.6
	Other	26.8

Data Analysis

Two-stage analytical procedures were used to test the measurement model and structural model (Anderson & Gerbing, 1988). The measurement model was measured by running an exploratory factor analysis (EFA) with SPSS (v. 25) with principal axis factoring extraction and Promax rotation, and the structural model was examined by using IBM AMOS (v.25).

Data Screening

The initial observations consisted of the entire dataset of 325 observations, including all 33 questions. A total of two observations were removed in the first run due to missing education, experience, and gender. Then, we removed 5 observations with missing education and 4 with missing experience. Finally, we removed data for 4 cases because the respondents were unengaged (missed attention traps). The adjusted N upon removal of 15 sample observations (1.53%) was 310 (N = 310). Regarding statistical power, significance levels, and effect size, this sample size is appropriate for conducting EFA and confirmatory factor analysis (CFA) on the collected data, as mentioned in previous research (Hair et al., 2010). Additionally, the skewness values ranged from -.744 to .559, and the kurtosis values ranged from -.923 to .236. Thus, the skewness values were lower than 3, and the kurtosis values were lower than 3 (Kline, 2015).

Exploratory Factor Analysis (EFA)

Maximum likelihood (ML) extraction with oblique ML estimation was used to examine the factor structure because this technique maximizes the differences between factors and is the approach used in AMOS, which is the measurement model analysis software used for subsequent CFA and structural modeling (Costello & Osborne, 2005). The Kaiser-Meyer-Olkin (KMO) test for measuring sampling adequacy in the responses yielded a value of .853, which

was considered superb and far greater than the minimum recommended value of .50 for proceeding with factor analysis (Kaiser, 1974). Additionally, the result of Bartlett's test of sphericity was adequate (χ^2 = 2825.558, DF = 153, p< .000), indicating sufficient intercorrelations to reject the null hypothesis and confirming that the correlation matrix was not an identity matrix (Olawale & Garwe, 2010). Moreover, the four-factor model with 18 items accounted for 67.642% of the total variance explained. There were 4% nonredundant residuals with absolute values greater than .05. Communality was reviewed and considered adequate, as the values ranged between .405 and .782 (Arnold & Feldman, 1981; Crowne & Marlowe, 1964). The eigenvalues ranged from .952 to 7.134, indicating that the 4-factor model was most likely adequate. This conclusion was supported by the data being considered suitable for an EFA. The remaining communality values confirmed that each item shared common variance with the other items. The exploratory factor analysis using Maximum Likelihood resulted pattern matrix (Appendix A).

Evidence of high reliability was also present (Appendix B), as all values were above the acceptable Cronbach's alpha value > .7 (Hair et al., 2010).

The above EFA indicated a solution with variables split into three factors (OI composed of OI3 and OI4; POS composed of POS1 and POS2; KSC composed of KSC2, KSC3, and KSC4). In summary, the validity of each indicator was verified, and the results are summarized in Appendix B.

Confirmatory Factor Analysis (CFA)

To examine the factor loadings from the EFA, this study imported the pattern matrix using the Gaskin plug-in and conducted a CFA (Gaskin, Lim, 2016) to establish the validity of the measurement model. The initial CFA measurement model resulted in an acceptable model fit.

Goodness-of-fit (GOF) measures were considered to evaluate the overall fit of the model. Overall model fit was assessed by the comparative fit index (CFI; Hair et al., 2010) and the incremental fit index (IFI; Hair et al., 2010). These fit indices were evaluated with the traditional cutoff value of .90. In addition, the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992) was used to assess a lack of model fit. The fit is considered close at RMSEA values of .06 or lower; reasonable at values between .05 and .08; mediocre at values between .08 and .10; and poor at values larger than .10 (Browne & Cudeck, 1992). The 2nd-order factor model showed a good fit in Table 2; this model was excellent. Thus, the validity of the data examined in this section was estimated to be sufficient for use in further analysis.

Table 2 Overall model fit indices of the measurement model (Hu and Bentler, 1999).

Fit indices	Recommended valu	GOF statistics
χ^2		197.716
Degrees of freedom (df)		126
Goodness-of-Fit Index (GFI)	> .95	.936
Adjusted Goodness-of-Fit Index (AGFI)	> .80	.913
SRMR	< .09	.046
Comparative Fit Index (CFI)	> .90	.974
Root Mean Square Error of Approximation (RMSEA)	< .08	.043
PClose	> .05	.851

Common Method Bias

We used the common latent factor (CLF) approach to evaluate common method bias (CMB) (Podsakoff, 2003). The χ^2 test for the zero constrained model was significant (i.e., measurable bias was detected). Therefore, a bias distribution test (of equal constraints) was performed, and a test of equal specific bias demonstrated unevenly distributed bias (Gaskin & Lim, 2017). Factor scores will be imputed to ensure CLF adjusted construct to be used in the strutural model.

We analyzed the factorial validity, convergent validity, and discriminant validity in the measurement model. Factorial validity was supported by the factor loadings, which were above .50. The loadings ranged (Appendix A) from .624 to .908 and were sufficient, as the sample size was > 300 (Hair et al., 2010). For convergent validity, the composite reliability (CR) and average variance extracted (AVE) of each construct exceeded the minimum criteria of .70 and .50, respectively (Fornell & Larker, 1981). Finally, according to Fornell & Larker (1981), the square of the correlations was compared with the AVE to examine discriminant validity. We confirmed that all of the AVE values of a construct were greater than the correlation coefficients between the constructs (Fornell & Larker, 1981). The means, standard deviations, and correlations are presented in Table 3. Thus, we concluded that the discriminant validity between all latent factors was acceptable (Hair et al., 2010).

Table 3. Correlations among the variables. N=310

No.	Characteristic	M	S.D	1	2	3	4	5	6	7	8
1	Gender	1.59	.49	-.11							
2	Education	2.90	.77	.04	-.165*						
3	Team size	3.93	1.14	.115*	-.09	.08					
4	Hiding	2.64	1.19	-.235**	-.148**	.211*	.190**				
5	Psychology safety	3.13	.93	.02	.08	.123*	0	.06			
6	Organizational identification	3.90	.93	.07	-.05	.221**	.10	.08	.424**		
7	POS	9.80	1.06	-.01	-.05	.212**	.10	.04	.483**	.679**	
8	Knowledge-sharing climate	3.82	.84	.11	-.06	.310**	.06	.07	.486**	.691**	.680**

Note. *$p < .05$; **$p < .01$; ***$p < .001$. Gender was coded as 0= female and 1= male. Education was coded as 1= Less than high school, 2= high school, 3= college, 4=postgraduate, and 5= Ph.D. Team size was coded as 1= 5 or fewer full-time employees, 2= 6-10 full-time employees, 3= 11-19 full-time employees, 4= 20-29 full-time employees, and 5= 30 or more full-time employees. POS indicates perceived organizational support.

Structural model

The fit for the proposed structural model was in an acceptable range, according to Table 4: SRMR = .056, which is below the acceptable threshold of .08 (Hu and Bentler, 1999); GFI = .966, which is above the threshold of .95 (Tabachnick and Fidell, 2007); and CFI = .950, which is above the threshold of .9 (Hu and Bentler, 1999). However, an unacceptable fit was suggested as well: RMSEA = .087, which is slightly above the threshold of .08 (Hu and Bentler, 1999), and PCLOSE = .005. We have analyzed the modification indices and removed non-significant paths to improve model fit results. Nonetheless, RMSEA remained slightly above threshold.

Table 4 Overall model fit indices of the structural model (Hu and Bentler, 1999).

Fit indices	Recommended valu	GOF statistics
X^2		69.674
Degrees of freedom (df)		21
Goodness-of-Fit Index (GFI)	> .95	.966
Adjusted Goodness-of-Fit Index (AGFI)	> .80	.874
SRMR	< .09	.056
Comparative Fit Index (CFI)	> .90	.950
Root Mean Square Error of Approximation (RMSEA)	< .08	.087
PClose	> .05	.005

RESULTS

The path analyses from IBM AMOS confirmed most of the proposed hypotheses. Figure 3 presents the results of the structural model. Except for altruism, all of the control variables impacted knowledge hiding significantly.

H1 was supported: POS was positively related to PS (*b*= .269, *p* < .001). However, the results did not support H2, as PS was not associated with knowledge hiding (*b*= .047, n.s). H3 was supported, as an individual's PS had a direct positive relationship with a KSC (*b*= .185, *p* < .01). The results supported H4, showing that POS was significantly and negatively related to knowledge hiding (*b*= -.195, *p* < .05). Moreover, H5 was supported as an individual's level of acceptance of selfishness was positively related to PS (*b*= .293, *p* < .001). Furthermore, H6 was supported, as an individual's level of acceptance of selfishness was positively associated with knowledge hiding (*b*= .152, *p* < .05). H7 was also supported: POS was positively associated with a KSC (*b*= .460, *p* < .001). Additionally, H8 was supported, as POS was positively related to OI (*b*= .331, *p* < .001). H9 was supported, as a KSC was positively related to OI (*b*= .456, *p* < .001). Finally, H10 was unsupported: an individual's OI did not have a direct positive relationship with knowledge hiding behavior (*b*= .069, n.s).

About the control variables, age and gender were negatively associated with knowledge hiding; both education and team size were positively associated with knowledge hiding.

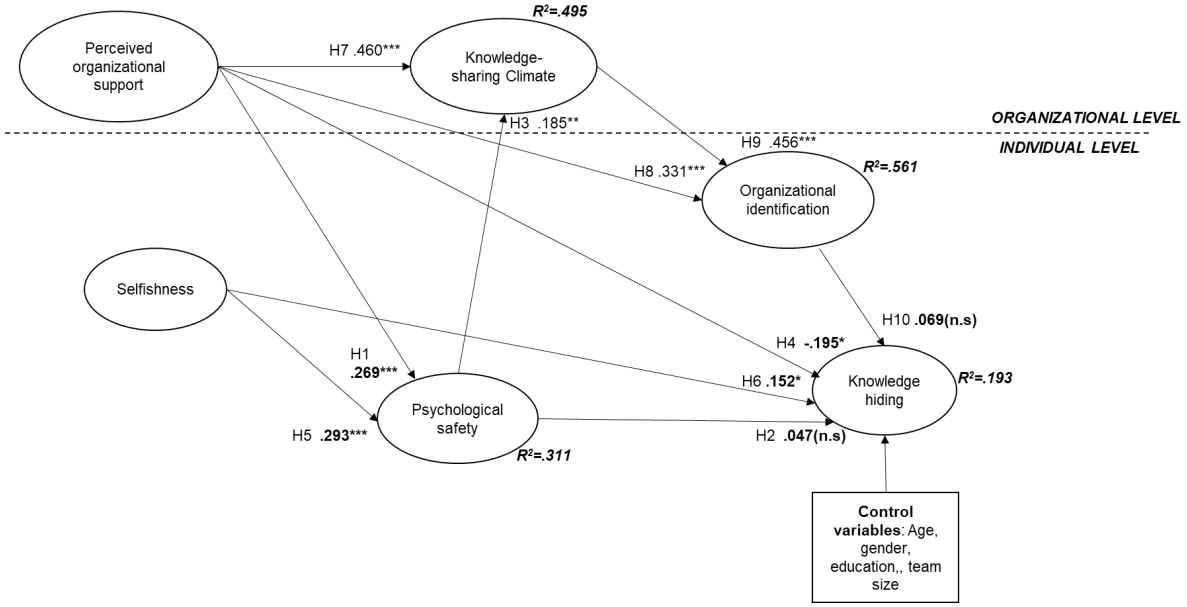


Figure 3. Summary of the Results

DISCUSSION

Knowledge hiding is an exciting and interesting topic in knowledge management research, although researchers had seldom studied this subject deeply until the study by Connelly et al. (2012). Following recent studies (Abubakar et al., 2019; Singh, 2019; Gagné, Tian, Soo, Zhang, Ho, & Hosszu, 2019), the present study established a theoretical model to explain how and why knowledge-hiding behavior arises in an organization when the staff perceive kindness and appreciation from their organization. Our primary research question in this study is: What organizational and individual factors affect knowledge hiding? The results answer our research question well in that POS (as an organizational factor) and selfishness (as an individual factor) affected knowledge-hiding behavior.

According to our findings, most selfish employees seem to avoid sharing knowledge; this avoidance is inappropriate and unethical organizational behavior. However, selfishness is a human characteristic that we believe is widespread and normal in organizations when people feel uncomfortable and nervous in their careers.

PS was not related to hiding behavior in our study. Although PS was not related to hiding behavior in our study, our results presented a few entertaining areas to discuss. Many individuals say they are insecure and anxious about their jobs under high-pressure working conditions. In this organizational climate, where competition is cruel and full of opportunities, fierce competition in the workplace causes disorganization, individual fear of unemployment, lack of belonging, a threat to dignity, and the uncertainty of possible failure (Kelly, & Noonan, 2008; Krasnyanskaya, & Tylets, 2018). One of our results indicated that POS plays a role position in helping individuals feel the organization appreciates their diligent effort and contributions. Therefore, individuals should not be afraid anymore about their psychological concerns; they can provide kindness to the organization reciprocally. Another significant finding was after individuals feel they are safe in an organization, they want to help the organization to build up a knowledge-sharing climate as an interchange.

In terms of external factors, the most direct factor triggering PS is the internal management and climate of the organization, including talent strategy, salary, competition mechanism, work content, management mechanism, working environment, politics, and other aspects (Krasnyanskaya, & Tylets, 2018). For example, some companies will set up a backup plan, also known as a staff replenishment plan, for core and critical positions in case an incumbent suddenly becomes unavailable. If individuals leave their position or do not perform well, the organization can always use a backup candidate to replace them. Many companies have established internal competitions that keep the performance of employees efficient. Still, the employees feel (a) the pressure of competition and (b) the resulting lack of job security. Under this kind of stressful working climate, a selfish personality thrives, according to our results.

An organization's concern for the welfare of its employees and the recognition and importance it attaches to their contributions will lead to a certain level of POS. Consequently, individuals provide a few responses. First, individuals have positive work emotions and reduced work stress. Our outcomes presented that individuals feel more psychological safety in the organization. Second, individuals produce results-reward expectations, and they, in turn, care about the organization and its goals, demonstrated, for instance, by helping the management team build up the knowledge-sharing climate to make another contribution to the organization. Finally, per our results, individuals will strengthen their commitments to the organization and improve their performance, reducing the turnover rate, which is the desired outcome of the organization. Also, this behavior identifies the entire organization.

Theoretical implications

Our study is the first to integrate both OI theory and POS theory into a knowledge-hiding model. Kleinert, Johnson, Webel, & Fitzpatrick (2020) and Shanock, Eisenberger, Heggestad, Malone, Clark, Dunn, & Woznyj (2019) emphasized that POS theory has concentrated on how firms use human resource benefits such as rewards, promotion and other methods to motivate staff to perform better. Our findings emphasize staff's emotional experience of receiving support from their organization and their resulting actions to protect the organization. Employees with such a positive perception believe in helping the firm to build a KSC and identify with their organization.

PS is an emotion that does not need to be expressed loudly and refers to one's deeper experience of interaction between an individual and a leader. According to Gupta (2008), there many factors that generate a knowledge-sharing climate; our findings indicated there are other ways to consider how human emotions positively influence the knowledge-sharing climate to enhance both PS and POS.

A reciprocal relationship (Moeller & de Vreese, 2019; Teasdale, 1983) emerges as a result of this kindness, representing a trade-off for employees. Social exchange theory (Roch et al., 2019; Tsai & Kang, 2019) is based on trust and mutual benefit, and the relationship of mutual rights and obligations developed between the employer and employees can even surpass what is specified in an actual job contract. This study applies POS to social exchange theory to explain the phenomenon of how individuals respond to the kindness or carelessness from an organization. Our results strengthen social exchange theory, POS theory, and organizational identification theory; that is, the results provide a few different perceptions when we examine knowledge hiding theory.

Practical implications

This research has important practical implications for both managers and employees. This study provides different insights to discuss knowledge hiding at organizational and individual levels.

POS has a strong relationship with all of the variables from our study, such as KSC, PS, OI, and KI. These results explain that the more individuals expect the results of their work to bring the rewards they desire, the higher the motivation or incentive to pursue those results. Individuals will view their relationship with the organization as an interpersonal relationship between themselves and a more powerful individual. In this relationship, individuals will see the organization's policies, practices, and norms and the words and actions of the supervisors' as the organization's values toward its employees. All these will bring a certain understanding and feeling about the organization to the staff. For managers, on the contrary, the consideration, motivation, and situation of the organization in providing such support may be more important. If the support is obtained by the employees through striving, the employees will not feel that it is because of the goodwill and care of the organization.

Also, PS has a strong relationship with KSC. Based on the results of this study, we make the following recommendations for employees. When there is a lack of PS, there is often fear and resistance to work. Because a sense of security is a basic need of people, only people who have a sense of security and a sense of belonging will wholeheartedly put effort into their work. The sense of occupational security of employees is not only conducive to individual employees; for the enterprise, it is also very important. If the organization cannot give employees enough sense of job security, the turnover rate of employees will be particularly high; thus, the workplace will be relatively tense. It is the manager's responsibility to prevent a situation where employees' work becomes inefficient and performance suffers. Hence, it is a necessary and a win-win task for companies to take on the mission of improving employee job security.

For employees, insecurity is a temporary state, not a permanent one (Weber & Petriglieri, 2018). We recommend employees know their strengths in terms of abilities and qualities and believe that these strengths will enable them to cope with a professional crisis. The best strategy to expand one's competitive advantage from peers is to improve one's capability and to help the management team to build up the knowledge-sharing climate.
If individuals do their jobs well and gain recognition from others, individuals will naturally feel more and more secure in their careers. On the other hand, regardless of one's position, one needs to demonstrate a humble and friendly attitude and maintain harmonious interpersonal relationships. Earn the respect of leaders and colleagues and gain solid interpersonal support. Sharing your knowledge with peers is the best way to gain respect.

Limitations and future research directions

Although many studies have discussed how and why knowledge hiding occurs in an organization, only a few studies (Connelly, & Zweig, 2015; Peng et al., 2019) have evaluated the *consequences* of staff hiding their knowledge in the organization. Future research may consider the effects of knowledge hiding, such as how to recover from knowledge hiding and how to prevent this unethical behavior.

Future research should consider whether knowledge hiding has positive benefits for organizations. A few studies have examined the ethical issues associated with knowledge hiding in businesses to develop a theoretical model of why and how staff follow ethical leadership in hiding their knowledge. For example, a list of names of laid-off employees is an example where knowledge hiding can be necessary to protect a firm. Future research should investigate this idea to compare it with the negative motivation for knowledge hiding.
Based on our findings, we suggest that CMB may, in part, potentially influence the relationship in our causal model. However, after we imputed the data, we found that CMB was not a significant concern in our structural model.

Conclusion

Knowledge asymmetry is the most crucial advantage in competition, with effects at both the organizational and individual levels. Individuals may seek to increase the level of knowledge asymmetry to guard what they perceive as their territory and seek competitive advantage in the organization because KS behavior involves expending resources such as time; therefore, hiding knowledge becomes a choice. This study emphasizes that feelings such as nervousness and anxiety and selfish behaviors originating from deep within an individual's mind can be relevant to the knowledge asymmetry occurring in an organization.

References

Abubakar, A. M., Behravesh, E., Rezapouraghdam, H., & Yildiz, S. B. (2019). Applying artificial intelligence technique to predict knowledge hiding behavior. International *Journal of Information Management*, 49, 45-57.

Alzghoul, A., Elrehail, H., Emeagwali Okechukwu, L., & AlShboul Mohammad, K. (2018). Knowledge management, workplace climate, creativity and performance: The role of authentic leadership. *Journal of Workplace Learning*, 30(8), 592-612. doi:10.1108/JWL-12-2017-0111

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3), 411.

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal*, 24(2), 377-385.

Beck, R., Pahlke, I., & Seebach, C. (2014). Knowledge exchange and symbolic action in social media-enabled electronic networks of practice: A multilevel perspective on knowledge seekers and contributors. *Management Information Systems Quarterly*, 38, 1245-1270.

Bock, G.-W., Zmud, R. W., Kim, Y.-G., & Lee, J.-N. (2005). Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate. *MIS Quarterly*, 29(1), 87-111. doi:10.2307/25148669

Bouzenita, A. I., & Boulanouar, A. W. (2016). Maslow's hierarchy of needs: An Islamic critique. *Intellectual Discourse*, 24(1).

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230-258.

Buvik, M. P., & Tvedt, S. D. (2017). The influence of project commitment and team commitment on the relationship between trust and knowledge sharing in project teams. *Project Management Journal*, 48(2), 5-21.

Černe, M., Nerstad, C. G., Dysvik, A., & Škerlavaj, M. (2014). What goes around comes around: Knowledge hiding, perceived motivational climate, and creativity. *Academy of Management Journal*, 57(1), 172-192.

Connelly, C. E., & Kevin Kelloway, E. (2003). Predictors of employees' perceptions of knowledge sharing cultures. *Leadership & Organization Development Journal*, 24(5), 294-301.

Connelly, C. E., & Zweig, D. (2015). How perpetrators and targets construe knowledge hiding in organizations. *European Journal of Work and Organizational Psychology*, 24(3), 479-489.

Connelly, C. E., Černe, M., Dysvik, A., & Škerlavaj, M. (2019). Understanding knowledge hiding in organizations. *Journal of Organizational Behavior*, 40(7), 779-782.

Connelly, C. E., Zweig, D., Webster, J., & Trougakos, J. P. (2012). Knowledge hiding in organizations. *Journal of Organizational Behavior*, 33(1), 64-88. doi:10.1002/job.737

Connelly, E. C., Zweig, D., (2015) How perpetrators and targets construe knowledge hiding in organizations. *European Journal of Work and Organizational Psychology*, 24(3), 479-489.

Conroy, S. A., Becker, W. J., & Menges, J. I. (2017). The meaning of my feelings depends on who I am: Work-related identifications shape emotion effects in organizations. *Academy of Management Journal*, 60(3), 1071-1093.

Constant, D., Kiesler, S., & Sproull, L. (1994). What's mine is ours, or is it? A study of attitudes about information sharing. *Information Systems Research*, 5(4), 400-421.

Cornwell, T. B., Howard-Grenville, J., & Hampel, C. E. (2018). The company you keep: How an organization's horizontal partnerships affect employee organizational identification. *Academy of Management Review*, 43(4), 772-791.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1), 7.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*: New York: Wiley.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, 44(2), 350-383. doi:10.2307/2666999

Eisenberger, R., Cummings, J., Armeli, S., & Lynch, P. (1997). Perceived organizational support, discretionary treatment, and job satisfaction. *Journal of Applied Psychology*, 82(5), 812-820. doi:10.1037/0021-9010.82.5.812

Eisenberger, R., Fasolo, P., & Davis-LaMastro, V. (1990). Perceived organizational support and employee diligence, commitment, and innovation. *Journal of Applied Psychology*, 75(1), 51-59. doi:10.1037/0021-9010.75.1.51

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, 71(3), 500-507. doi:10.1037/0021-9010.71.3.500

Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I. L., & Rhoades, L. (2002). Perceived supervisor support: contributions to perceived organizational support and employee retention. *Journal of Applied Psychology*, 87(3), 565.

Evans, J. M., Hendron, M. G., & Oldroyd, J. B. (2014). Withholding the ace: The individual-and unit-level performance effects of self-reported and perceived knowledge hoarding. *Organization Science*, 26(2), 494-510.

Faraj, S., & Sproull, L. (2000). Coordinating expertise in software development teams. *Management Science*, 46(12), 1554-1568.

Farrugia, D. (2002). Selfishness, greed, and counseling. Counseling and Values, 46(2), 118-126.

Ferrari, I. (2019). The effectiveness of incentives to postpone retirement: evidence from Italy. *Journal of Pension Economics & Finance*, 18(2), 220-246.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* (JMR), 18(1), 39-50.

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-sort: A Tool for the Description of Social Behavior. Journal of Personality, 68(3), 451-489. doi:10.1111/1467-6494.00103

Gagné, M., Tian, A. W., Soo, C., Zhang, B., Ho, K. S. B., & Hosszu, K. (2019). Different motivations for knowledge sharing and hiding: The role of motivating work design. *Journal of Organizational Behavior*, 40(7), 783-799.

Gamito, P., Oliveira, J., Rosa, P., Morais, D., Duarte, N., Oliveira, S., & Saraiva, T. (2010). PTSD Elderly War Veterans: A Clinical Controlled Pilot Study. Cyberpsychology, Behavior, and Social Networking, 13(1), 43-48. doi:10.1089/cyber.2009.0237

Gaskin, J., & Lim, J. (2016). Model fit measures. *Gaskination's StatWiki*, 1-55.

Gaskin, J., & Lim, J. (2017). CFA tool. *AMOS Plugin*.

Geofroy, Z., & Evans, M. M. (2017). Are Emotionally Intelligent Employees Less Likely to Hide Their Knowledge? *Knowledge and Process Management*, 24(2), 81-95.

Gupta, K. S. (2008). A comparative analysis of knowledge sharing climate. Knowledge and Process Management, 15(3), 186-195.

Hair, J., Black, B., Babin, B., & Anderson, R. (2010). Multivariate Data Analysis 7th Pearson Prentice Hall. *Upper Saddle River, NJ.*

Hajro, A., Gibson, C. B., & Pudelko, M. (2017). Knowledge exchange processes in multicultural teams: Linking organizational diversity climates to teams' effectiveness. *Academy of Management Journal*, 60(1), 345-372.

Hatch, M. J., & Schultz, M. (1997). Relations between organizational culture, identity and image. *European Journal of marketing*, 31(5-6), 356-365.

Ho, H., & Ganesan, S. (2013). Does knowledge base compatibility help or hurt knowledge sharing between suppliers in coopetition? The role of customer participation. *Journal of Marketing*, 77, 91-107. doi:10.1509/jm.11.0570

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Intezari, A., Taskin, N., & Pauleen, D. J. (2017). Looking beyond knowledge sharing: an integrative approach to knowledge management culture. *Journal of Knowledge Management*, 21(2), 492-515. doi:10.1108/JKM-06-2016-0216

Ives, C. D., Freeth, R., & Fischer, J. (2020). Inside-out sustainability: The neglect of inner worlds. Ambio, 49(1), 208-217.

Jeung, C.-W., Yoon, H. J., & Choi, M. (2017). Exploring the affective mechanism linking perceived organizational support and knowledge sharing intention: a moderated mediation model. *Journal of Knowledge Management*, 21(4), 946-960. doi:10.1108/JKM-12-2016-0530

Jiang, Z. (2019, November 26). Why withholding information at work won't give you an advantage. Harvard Business Review. Retrieved from https://hbr.org/2019/11/why-withholding-information-at-work-wont-give-you-an-advantage

Jiang, Z., Hu, X., Wang, Z., & Jiang, X. (2019). Knowledge hiding as a barrier to thriving: The mediating role of psychological safety and moderating role of organizational cynicism. *Journal of Organizational Behavior*, 40(7), 800-818.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.

Kelly, S., & Noonan, C. (2008). Anxiety and psychological security in offshoring relationships: the role and development of trust as emotional commitment. *Journal of Information Technology*, 23(4), 232-248.

Kleinert, D., Johnson, J., Webel, A. R., & Fitzpatrick, J. (2020). Pediatric nurse practitioner's perceived organizational support in academic and nonacademic institutions. *Journal of the American Association of Nurse Practitioners*, 32(1), 45-51.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*: Guilford publications.

Krasnyanskaya, T., & Tylets, V. (2018). Resources of individual psychological security with respect to the employment status of a pensioner. *Advances in Gerontology*, 8(3), 256-261.

Kurtessis, J. N., Eisenberger, R., Ford, M. T., Buffardi, L. C., Stewart, K. A., & Adis, C. S. (2017). Perceived organizational support: A meta-analytic evaluation of organizational support theory. *Journal of Management*, 43(6), 1854-1884.

Labafi, S. (2017). Knowledge hiding as an obstacle of innovation in organizations a qualitative study of software industry. *AD-minister* (30), 131-148.

Ladan, S., Nordin, N. B., & Belal, H. M. (2017). Influence of transformational leadership on knowledge hiding: mediating role of organizational psychological ownership: A proposed framework. International Journal of Business & Management Science, 7(2).

Li, Z., Zhu, T., & Luo, F. (2010). A Study on the Influence of Organizational Climate on Knowledge-Sharing Behavior in IT Enterprises. *JCP*, 5(4), 508-515.

Liang, J., C. Farh, C. I., & Farh, J.-L. (2012). Psychological antecedents of promotive and prohibitive voice: a two-wave examination. *Academy of Management Journal*, 55(1), 71-73. doi:10.5465/amj.2010.0176

Liu, W., Song, Z., Li, X., & Liao, Z. (2017). Why and when leaders' affective states influence employee upward voice. *Academy of Management Journal*, 60(1), 238-263.

Liu, W., Tangirala, S., Lam, W., Chen, Z., Jia, R. T., & Huang, X. (2015). How and when peers' positive mood influences employees' voice. *Journal of Applied Psychology*, 100(3), 976.

Lu, Y., Singh, P. V., & Sun, B. (2015). Is core-periphery network good for knowledge sharing? A structural model of endogenous network formation on a crowdsourced customer support forum. *In A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum* (August 3, 2015).

Malik, M. A. R., Butt, A. N., & Choi, J. N. (2015). Rewards and employee creative performance: Moderating effects of creative self-efficacy, reward importance, and locus of control. *Journal of Organizational Behavior*, 36(1), 59-74.

Marin, A., Cordier, J., & Hameed, T. (2016). Reconciling ambiguity with interaction: implementing formal knowledge strategies in a knowledge-intensive organization. *Journal of Knowledge Management*, 20(5), 959-979. doi:10.1108/JKM-11-2015-0438

Miao, Q., Eva, N., Newman, A., & Cooper, B. (2019). Ceo entrepreneurial leadership and performance outcomes of top management teams in entrepreneurial ventures: The mediating effects of psychological safety. *Journal of small business management*, 57(3), 1119-1135.

Moeller, J., & de Vreese, C. (2019). Spiral of political learning: The reciprocal relationship of news media use and political knowledge among adolescents. *Communication Research*, 46(8), 1078-1094.

Mok, A., & De Cremer, D. (2018). Too tired to focus on others? Reminders of money promote considerate responses in the face of depletion. *Journal of Business and Psychology*, 33(3), 405-421.

Offergelt, F., Spörrle, M., Moser, K., & Shaw, J. D. (2019). Leader-signaled knowledge hiding: Effects on employees' job attitudes and empowerment*. Journal of Organizational Behavior*, 40(7), 819-833

Olawale, F., & Garwe, D. (2010). Obstacles to the growth of new SMEs in South Africa: A principal component analysis approach. *African journal of Business management*, 4(5), 729.

Pan, N. D., Gruber, M., & Binder, J. (2019). Painting with all the colors: The value of social identity theory for understanding social entrepreneurship. *Academy of Management Review*, 44(1), 213-215.

Pan, W., Zhang, Q., Teo, T. S., & Lim, V. K. (2018). The dark triad and knowledge hiding. *International Journal of Information Management*, 42, 36-48.

Pan, W., Zhang, Q., Teo, T. S., & Lim, V. K. (2018). The dark triad and knowledge hiding. International Journal of Information Management, 42, 36-48.

Peng, H. (2013). Why and when do people hide knowledge? *Journal of Knowledge Management*, 17(3), 398-415.

Peng, J., Wang, Z., & Chen, X. (2019). Does self-serving leadership hinder team creativity? A moderated dual-path model. *Journal of Business Ethics*, 159(2), 419-433.

Podsakoff, N. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 885(879), 10.1037.

Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. Journal of Applied Psychology, 87(4), 698-714. doi:10.1037/0021-9010.87.4.698

Ritzenhöfer, L., Brosi, P., Spörrle, M., & Welpe, I. M. (2019). Satisfied with the job, but not with the boss: Leaders' expressions of gratitude and pride differentially signal leader selfishness, resulting in differing levels of followers' satisfaction. *Journal of Business Ethics*, 158(4), 1185-1202.

Roch, S. G., Shannon, C. E., Martin, J. J., Swiderski, D., Agosta, J. P., & Shanock, L. R. (2019). Role of employee felt obligation and endorsement of the just world hypothesis: A social exchange theory investigation in an organizational justice context. *Journal of Applied Social Psychology*, 49(4), 213-225.

Rusly, F. H., Corner, J. L., & Sun, P. (2012). Positioning change readiness in knowledge management research. *Journal of Knowledge Management*, 16(2), 329-355.

Scott, K. L., Zagenczyk, T. J., Schippers, M., Purvis, R. L., & Cruz, K. S. (2014). Co-worker exclusion and employee outcomes: An investigation of the moderating roles of perceived organizational and social support. *Journal of Management Studies*, 51(8), 1235-1256.

Serenko, A., & Bontis, N. (2016). Understanding counterproductive knowledge behavior: antecedents and consequences of intra-organizational knowledge hiding. *Journal of Knowledge Management*, 20(6), 1199-1224. doi:10.1108/JKM-05-2016-0203

Shanock, L. R., Eisenberger, R., Heggestad, E. D., Malone, G., Clark, L., Dunn, A. M., . . . Woznyj, H. (2019). Treating employees well: The value of organizational support theory in human resource management. *The Psychologist-Manager Journal*, 22(3-4), 168-191. doi:10.1037/mgr0000088

Shore, L. M., & Wayne, S. J. (1993). Commitment and employee behavior - comparison of affective commitment and continuance commitment with perceived organizational support. *Journal of Applied Psychology*, 78(5), 774-780. doi:10.1037/0021-9010.78.5.774

Singh, S. K. (2019). Territoriality, task performance, and workplace deviance: Empirical evidence on role of knowledge hiding. *Journal of Business Research*, 97, 10-19.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5): Pearson Boston, MA.

Takaki, J., Taniguchi, T., & Fujii, Y. (2016). Confirmation of Maslow's Hypothesis of Synergy: Developing an Acceptance of Selfishness at the Workplace Scale. *International journal of environmental research and public health*, 13(5), 462.

Teasdale, J. D. (1983). Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behaviour Research and Therapy*, 5(1), 3-25.

Teng, J. T. C., & Song, S. (2011). An exploratory examination of knowledge- sharing behaviors: solicited and voluntary. *Journal of Knowledge Management*, 15(1), 104-117. doi:doi:10.1108/13673271111108729

The Globe and Mail. (2006). The weekly web poll. Friday May 12, 2006, C1.

Titi Amayah, A. (2013). Determinants of knowledge sharing in a public sector organization. *Journal of Knowledge Management*, 17(3), 454-471. doi:doi:10.1108/JKM-11-2012-0369

Tremblay, M., Gaudet, M.-C., & Vandenberghe, C. (2019). The role of group-level perceived organizational support and collective affective commitment in the relationship between leaders' directive and supportive behaviors and group-level helping behaviors. *Personnel Review*, 48(2), 417-437. doi:10.1108/PR-06-2017-0172

Trialih, R., Wei, H.-L., & Anugrah, W. (2017). Knowledge sharing behavior and quality among workers of academic institutions in indonesia. *International Journal of Business and Society*, 18(S2), 353-368.

Tsai, J. C.-A., & Kang, T.-C. (2019). Reciprocal intention in knowledge seeking: Examining social exchange theory in an online professional community. *International Journal of Information Management*, 48, 161-174.

Tsay, C. H.-H., Lin, T.-C., Yoon, J., & Huang, C.-C. (2014). Knowledge withholding intentions in teams: The roles of normative conformity, affective bonding, rational choice and social cognition. Decision Support Systems, 67, 53-65.

van Gils, S., Hogg, M. A., Van Quaquebeke, N., & van Knippenberg, D. (2017). When organizational identification elicits moral decision-making: A matter of the right climate. *Journal of Business Ethics*, 142(1), 155-168.

Weber, S., & Petriglieri, G. (2018, June 27). To Overcome Your Insecurity, Recognize Where It Really Comes From. *Harvard Business Review*. Retrieved from https://hbr.org/2018/06/to-overcome-your-insecurity-recognize-where-it-really-comes-from

Yam, K. C., & Reynolds, S. J. (2016). The effects of victim anonymity on unethical behavior. *Journal of Business Ethics*, 136(1), 13-22.

Yoo, D. K. (2017). Impacts of a Knowledge Sharing Climate and Interdisciplinary Knowledge Integration on Innovation. *Journal of Information & Knowledge Management*, 16(02), 1750010. doi:10.1142/s0219649217500101

Zhao, H., Liu, W., Li, J., & Yu, X. (2019). Leader–member exchange, organizational identification, and knowledge hiding: The moderating role of relative leader–member exchange. *Journal of Organizational Behavior*, 40(7), 834-848.

Zhao, H., Xia, Q., He, P., Sheard, G., & Wan, P. (2016). Workplace ostracism and knowledge hiding in service organizations. *International Journal of Hospitality Management*, 59, 84-94.

Zhao, H., Xia, Q., He, P., Sheard, G., & Wan, P. (2019). “Workplace ostracism and knowledge hiding in service organizations”: Corrigendum. *International Journal of Hospitality Management*, 77, 290-291. doi:10.1016/j.ijhm.2018.07.008

Zhu, Y., Chen, T., Wang, M., Jin, Y., & Wang, Y. (2019). Rivals or allies: How performance-prove goal orientation influences knowledge hiding. *Journal of Organizational Behavior*, 40(7), 849-868. doi:10.1002/job.2372

APPENDIX

Appendix A Pattern matrix resulted in exploratory analysis			
	1	2	3
HI3	.878		
HI2	.863		
HI5	.853		
HI4	.841		
HI1	.740		
POS2		.766	
OI3		.740	
POS1		.694	
OI4		.681	
KSC4		.666	
KSC3		.634	
KSC2		.624	
PS2			.908
PS1			.740
PS3			.688

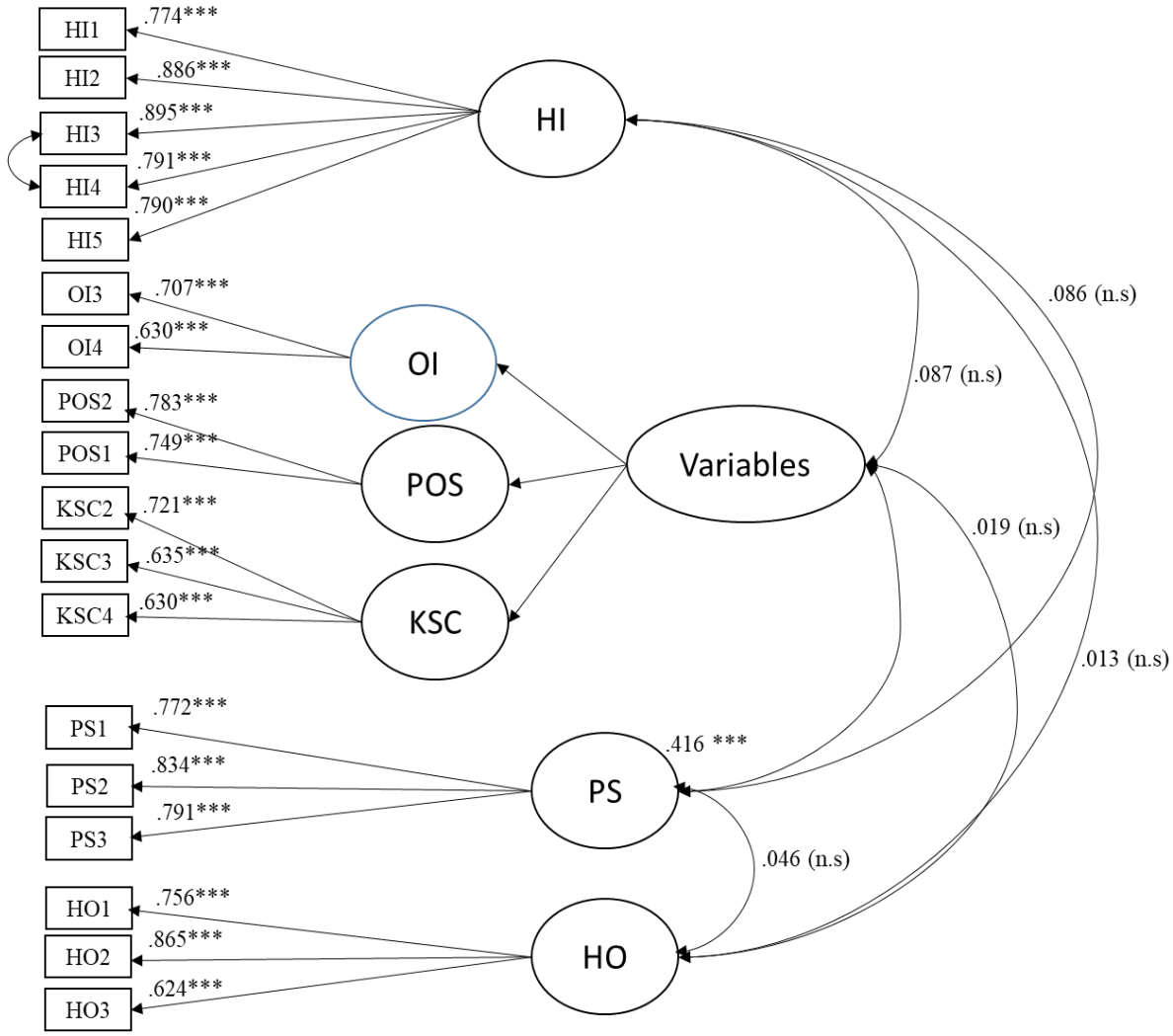
Appendix B Construct table			
Construct/Dimensions	Definition	Items	References
Knowledge hiding (Cronbach aplhpa=.919)	Intentional attempts made by individuals to withhold or conceal knowledge when they receive requests from knowledge seekers (Connelly et al., 2012).	1. If you needed to work with this coworker in the future, I would likely keep my ideas to myself if this person asked for my help.	Connelly et al. (2012)
		2. If you needed to work with this coworker in the future, I would agree to help him/her but never really intend to.	
		3. If you needed to work with this coworker in the future, I would agree to help him/her but instead give him/her information different from what s/he wanted.	
		4. If you needed to work with this coworker in the future, I would pretend that I did not have the information.	
		5. If you needed to work with this coworker in the future, I would tell him/her that my boss would not let	

		anyone share this knowledge.	
Psychological safety(Cronbach aplhpa=.836)	Staff members' feeling that they can express themselves to managers without any negative consequences, such as retaliation, that might threaten their position in their organization (Liu, Song, Li, & Liao, 2017).	1. In my work unit, I can express my true feelings regarding my job.	Liang, Farh, & Farh (2012)
		2. In my work unit, I can freely express my thoughts.	
		3. In my work unit, expressing your true feelings is welcomed.	
		4. Nobody in my unit will pick on me even if I have different opinions.	
		5. I'm worried that expressing true thoughts in my workplace would do harm to myself (reverse-coded).	
Knowledge-sharing climate (Cronbach aplhpa=.745)	"Believing that information sharing is usual, correct, and socially expected workplace behavior should increase information sharing, independent of the information possessor's personal feelings about his or her coworkers"	1. My manager would like me to share more information with other coworkers in the organization.	Connelly, & Kelloway (2003); Connelly et al., (2012); Faraj, & Sproull (2000)

	(Constant, Kiesler, and Sproull, 1994, p. 404).		
		2. My manager expects staff with expert knowledge to be willing to help other coworkers in this organization.	
		3. People in this organization share their ideas openly.	
		4. My manager expects staff in our team to share their special knowledge and expertise with one another.	
		5. More knowledgeable team members freely provide other members with hard-to-find knowledge or specialized skills.	
Perceived Organizational Support (Cronbach aplhpa=.739)	Employees develop general beliefs regarding the extent to which their organization values and appreciates their involvement and cares about their well-being (Eisenberger, Cummings, Armeli, & Lynch, 1997; Eisenberger, Huntington, Hutchison, & Sowa, 1986; Rhoades & Eisenberger, 2002).	1. My organization really cares about my well-being.	Eisenberger et al., (1986); Scott et al., (2014)
		2. My organization takes pride in my	

		accomplishments at work.	
		3. My organization would forgive an honest mistake on my part.	
		4. The organization strongly considers my goals and values.	
		5. The organization cares about my opinions.	
Organizational Identification (Cronbach aplhpa=.614)	Organizational identification is a ‘‘perception of oneness with or belongingness to an organization, where the individual defines him or herself in terms of the organization(s) in which he or she is a member’’ (Mael and Ashforth 1992, p. 104).	1. When someone criticizes my company, it feels like a personal insult.	Conroy, Becker, & Menges (2017); Mael, & Ashforth (1992)
		2. I am very interested in what others think about my company.	
		3. This company’s successes are my successes.	
		4. When someone praises this company, it feels like a personal compliment.	
		5. If a story in the media criticized the school, I would feel embarrassed.	

Appendix C
Measurement model results for 2nd order CFA



DECISION SCIENCES INSTITUTE
The Role of Digitization in Knowledge Creation: A Case of Hospice Industry

Mehrnaz Khalaj Hedayati
Georgia College & State University
Email: mehrnaz.khalajhedayati@gcsu.edu

Carla Curado
ISEG - Lisbon School of Economics & Management
Universidade de Lisboa
Email: ccurado@iseg.ulisboa.pt

Dara Schniederjans
The University of Rhode Island
Email: schniederjans@uri.edu

ABSTRACT

The dynamics of organizational knowledge creation and its drivers represent a major research topic in business and management studies. This study uses Nonaka knowledge creation theoretical framework coupled with qualitative data from semi-structured interviews of hospice practitioners in order to examine the role of digital technologies in knowledge creation in the context of hospice industry. Drivers and inhibitors are then identified. The results confirmed that it is only through the integration of the three pillars of knowledge management 'people', 'process', and 'technology' driven by 'awareness', 'motivation', and 'capability' that organizations can enhance the value to the patient.

KEYWORDS: Knowledge Creation, Digital Technology, Service Industry, Organizational Knowledge

INTRODUCTION

In 2017, hospice providers admitted 1.3 million unduplicated Medicare hospice patients, indicating 13.1% increase since 2014 (NHPCO, 2018). There are challenges for hospice to meet the demands for care knowledge while facing new care methods, guidelines, and communications to provide service-quality for the patient and their families (Abidi et al., 2005). Knowledge management (*KM*) which involves the process of creating, sharing, using, and management of knowledge and information in an organization can be considered as a potential solution to the current operational and technical hurdles for the management and dissemination of hospice care knowledge (Jackson, 2000; O'Leary, 1998; Abidi, 2001; Abidi et al., 2005).

There are various valuable theoretical frameworks underlying the dynamics of organizational knowledge creation (Nonaka, 1994; Nonaka & Takeuchi, 1995; Nonaka & Konno, 1998; Nonaka et al., 2000). Nonaka (1994) explained a dynamic theory of knowledge creation including socialization, externalization, combination, and internalization (*SECI*), which provides a foundational concept for further theoretical elaboration. According to Curado & Bontis (2011), *SECI* is underdeveloped and more empirical analysis is necessary to enhance the theoretical contribution. Their recommendations include operationalizing the processes of this model into semi-structured interview questions for qualitative research. An

integrated perspective can provide more conceptual clarity. Thus, this study collects qualitative data in order to integrate existing *SECI* theory developments to enhance and extend through digital technologies in the context of hospice industry. The aim is to add to theoretical knowledge by the exploration of the relationships of digital technologies on *SECI* framework in the context of hospice industry.

Using software applications and information technology has revolutionized the way of doing processes and collaborations in the organizations (Laurenza et al., 2018). Although the previous research examines the role of information technology and knowledge management in various industries (Obeidat, 2019; Bongku & Kurniawan, 2019; Ali et al., 2014), little research has been done on the role of digitization through new digital technologies in the service industry and particularly hospice care. Hospice providers and/or home health providers do not yet have the capability to understand how digital technologies will aid in the current processes. As such, it is now more important than ever to examine the knowledge management of these organizations and distinguish what drives this process so that technology can be better utilized. More specifically, this research aims to explore the role of digital technologies to support socialization and engagement for the purposes of sharing tacit knowledge, thus, the following research questions are explored:

- How do organizations conduct socialization, externalization, combination, and internalization?
- How does digitization induce and/or maintain movement between the four modes in the *SECI* model?

Interpretation of the results for each of the four modes of *SECI* framework are presented as well as the interpretation for movements from one mode to another such as socialization to externalization. Findings are categorized based on the three pillars of knowledge management that is 'people', 'process', and 'technology'. After doing the analyses, the research model is revised.

This paper follows a qualitative interpretive approach, with three main sections. First, the manuscript puts forward our theoretical framework following the literature review on digital technologies and *SECI* in service industries and introducing home health hospice. The next section explains our methodology and the data collection procedure. Then, the results are presented clarifying the relationship between digital technologies and knowledge creation modes investigating drivers and/or inhibitors of the movements in the four modes of *SECI* framework. Finally, the qualitative data is discussed with specific recommendation to the weaknesses in the conclusions.

LITERATURE REVIEW

According to Kayikci (2018), digitization converts the physical or analog signal to a digital form in order for electronic accessibility, storage, and processing capability. "Digitization makes information and communication available anywhere, anytime, within any context, and for any user using any device and type of access" (Kayikci, 2018, p.2), thus, a higher level of digitization means more and better access to information and business processes. This study defines digitization as the use of digital technologies for the purpose of working toward stakeholders' value. Digital technologies include anything from internet of things (*IOT*), advanced robotics, artificial intelligence, machine learning, drones, augmented reality, blockchain, cloud computing, sensors, etc. The purpose is to provide a well-connected network with the ability to collect, share, and use information to make better decisions and be more productive through decentralizing technology processes.

Although digital technologies provide various advantages and opportunities for storing, processing, and real-time sharing information, the dissemination and managing knowledge

is significantly needed and necessary in order to create value. Based on existing literature, limited human knowledge exists for how to implement and utilize a large amount of data in order to achieve strategic value through shared knowledge in an organizational network (Feng & Shanthikumar, 2018). Xue (2014) highlights the significance of knowledge management in digitization by proposing that knowledge creation in IT departments, with more empowerment in decision making, provides firms enhanced benefits in operational and strategic management. Research in supply chain indicated that there are four key enablers for supply chain digitization: technology, process, organization, and knowledge. The integration of technology and applications with effective knowledge management across firms are crucial factors to the success of supply chain digitization (Stuermer et al., 2017).

The new computing era requires persistent managing and sharing of data. Unfortunately, there lacks transfer of knowledge and experience to these new systems and environments (Peckham, 1999). Individuals and individual groups in organizations are needed to extract knowledge from the data to optimize the use of digital technologies for the purpose of enhancing performance (Schniederjans et al., 2019). Knowledge management which is a process of creating, managing, transferring, sharing, and increasing knowledge, involves technology, processes, and people (Carrion et al., 2004). Applying digital technologies aids routines that support knowledge management practices within industrial applications (Wilkesmann & Wilkesmann, 2018). In order to overcome managerial challenges of digitization such as challenges in planning and implementation, strategy and analysis, cooperation and networks, business models, human resources, change, and leadership firms must use a combination of the three pillars of knowledge management that is technology, processes, and people.

There are a number of studies investigated the role of *IT* in knowledge creation based on *SECI* framework. Obeidat (2019) identifies the extent of *IT* in knowledge conversion process through *SECI* framework developing questionnaire. He developed a model to examine the information technology adaptation with knowledge conversion process. Kovacs & Szabo (2019) investigated knowledge creation patterns in the supply chain companies. They explored information and communication technology (*ICT*) solutions in knowledge creation and collaboration with suppliers and customers. Ali et al. (2014) studied some concepts of knowledge management in higher education institutions and examined the roles of information technologies for knowledge management in higher education institutions. They identified information technologies that can contribute to knowledge creation in higher education using *SECI* framework.

While the role of *IT* in knowledge management is established in the current literature, the role of knowledge management and its relation to digitization has just started to emerge in academic literature.

Literature Review of Digitization Service Industry

In service firms, the organizational assets reside in the staff experience and knowledge, rather than in factory and equipment (Tsui et al., 2009). Offered services often vary in nature in order to address various customers' needs and demands (Tsui et al., 2009). Nowadays, firms manage and deal with information rather than raw materials to provide customized services (Agrifoglio et al., 2017). The collaboration and information sharing between various actors in a firm network provides customer value and improves processes. Using emergent technical capabilities can provide higher levels of customization to consumers (Agrifoglio et al., 2017).

Emerging digital technologies are profoundly changing the way in which the service industry

and in particular healthcare and hospice care processes are managed. It provides value by promoting cooperation of several healthcare players or actors. Healthcare industry services and in particular hospice care and processes strongly rely on the information and knowledge as well (Lenz et al., 2012; Laurenza et al., 2018). Therefore, knowledge and information management as well as management and support of digital technologies play essential roles in the healthcare industry (Laurenza et al., 2018).
Furthermore, healthcare and caregiver organizations, more than others, are being challenged by resource restrictions and growing regulatory frameworks and procedures. Caregiver companies such as hospice are trying to provide and increase service quality and simultaneously reduce costs with the purpose of maximizing value. Hospice care services include treatment of medical conditions that requires multiple interdisciplinary expertise and several interventions(Porter et al., 2010). Value for the patient is created by combined efforts of an interdisciplinary team over the full cycle of care (Laurenza et al., 2018). Process management, simplifying services and efficient processes, delivering high quality and reducing response times are critical factors in the hospice industry. Therefore, the enormous changes in providing high quality services in hospice industry requires making a new business model to create value in order to get better service performance (Laurenza et al., 2018).

Literature Review of the Roles of IT and Digitization in Hospice Care

Hospice not only provides comprehensive comfort for the patient, but also supports for the family as well. Hospice is provided for a person with a terminal illness whose doctor believes they have 6 months or less to live if the illness runs its natural course (Vitas, 2020). When a cure for the patient is no longer an option, the hospice care team is involved to make the patient's remaining life as comfortable as possible by addressing distress and physical symptoms as well as emotional and spiritual issues. Hospice care involves stopping any efforts to cure a person's illness and refocusing efforts to help terminally ill patients physically, spiritually, and mentally at the end of life. In palliative care the same philosophy of full-patient care is delivered to non-terminal patients, helping them live longer, happier lives (VITAS, 2020). Hospice is an approach to care, and it is not tied to a specific place. Thus, it can be offered at home or in a facility such as a nursing home, hospital, or even in a separate hospice center. With a focus on interdisciplinary care, hospice brings a team of people with special skills and expertise together. All members of this professional team provide the best in necessary medical treatment through a philosophy of values (i.e. caring, competent, reliable, honest and ethical) (NHPCO, 2018).
Due to multidisciplinary nature of hospice care as a service industry, considering human-factor aspects of the digital systems is very important in this particular service industry (Abernethy et al., 2011). As the main goal of hospice care is to increase satisfaction and enhance quality of life for the patients, health information technology (HIT) such as electronic medical records (EMR) have been increasingly used to improve the quality of hospice care services by increasing accuracy, reducing time of travel and care documentation (Resnick & Alwan, 2010). The common thought in hospice care as a service industry is to put more importance on the face-to-face interactions and less on technology approaches with patients (Parker Oliver & Demiris, 2004). Although nurses and staff in hospice care have an overall positive attitude toward adopting information technologies, they are concerned that technology may be used as replacement for face-to-face communications instead of using as a tool to improve the processes and quality of services (Courtney et al., 2005). Essentially individuals as teams, information sharing and knowledge processes, and technology as a tool can aid in the creation of this value when applied

appropriately. All members in hospice care ought to be committed to working collaboratively in order to assure patients and families have the best and most comfortable possible experience in a most challenging time. As such knowledge management becomes vital in the creation of value for the hospice care industry.

THEORETICAL FRAMEWORK

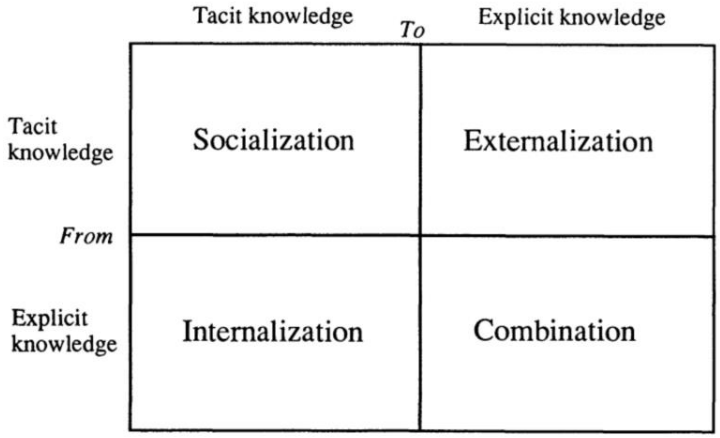
Knowledge has been identified as an important resource for the organizations to sustain competitive advantages (Easa, 2012). Value is created for firms when knowledge is gathered and utilized for their strategic purposes (Curado & Bontis, 2011). The value created from knowledge assets is generally maintained and often increased when it is applied, while other conventional assets typically depreciate when they are applied and utilized (Bontis & Serenko, 2007). Therefore, creating and managing knowledge is crucial for the strategic value (Curado & Bontis, 2011).

Nonaka's dynamic theory of organizational knowledge creation proposes that organizational knowledge is created and reinforced by the individual knowledge in an organization through converting tacit to explicit knowledge and moving knowledge from the individual to the group, organizational, and inter-organizational levels (Nonaka, 1994). This process occurs through a four-category model, referred to as *SECI*. This framework is based on the process of continuous interactions between explicit and tacit knowledge to create knowledge.

This study uses the definition of knowledge from the organizational viewpoint with the purpose of increasing individual knowledge as part of the knowledge network of the organization (Easa, 2012; Nonaka & Takeuchi, 1995). In this view, organizational knowledge is embedded in individuals and the repositories such as documents and databases as well as in organizational processes and practices (Wallace, 2007; Easa, 2012).

There is dynamic and continuous interaction between tacit and explicit knowledge that moves across individuals, groups, organizations, and back to individuals again (Dierkes et al., 2003). The combination of these two types of knowledge was utilized by Nonaka (1994) to conceptualize the creation of knowledge. Figure 1 shows the interaction of tacit and explicit knowledge.

Figure 1. Modes of knowledge Creation and Spiral Evolution of Knowledge Conversion (Nonaka, 1994)



The theory of knowledge creation proposes that knowledge is created and increases over an endless and continual cycle in which knowledge is shaped by a series of shifts/movements between different modes of knowledge conversion.

Socialization is defined as face-to-face communications among organization members to exchange tacit knowledge. The socialization process motivates organizations to provide enough space and time for interaction and communication among individuals (Curado & Bontis, 2011). They stated that to motivate socialization and transformation to happen, organizations might locate the existing comfortable rooms and social spaces.

Externalization requires an individual's mental models to be articulated. For example, writing down thoughts on paper or typing out content within a document to transform tacit into explicit knowledge thereby allowing it to be captured and accessible by others in an organization (Nonaka, 1994). When tacit knowledge is articulated and converted into explicit knowledge, the codification process happens. This conversion might be through using words, images, concept definitions, metaphors, figures, and analogies (Curado & Bontis, 2011).

Combination consists of transforming explicit knowledge into more complex forms like integrating business plans by teamwork in a meeting from the shared documents between them. Some examples of operationalized tasks and process of this mode include collaboration via virtual meetings (Curado & Bontis, 2011; McKnight & Bontis, 2002).

Internalization converts explicit knowledge back into tacit form allowing individuals to include it into their personal routines. For example, when an organization has integrated and developed its strategic plan, a manager must then internalize what that plan means for his/her own team and/or themselves.

METHODOLOGY

Theoretical Elaboration on Nonaka's Knowledge Creation Framework

Theory elaboration is used when there are preexisting conceptual ideas or a preliminary framework as the starting point, but there is some controversy, inadequacy, ambiguity, or uncertainty, in the existing theoretical explanation
(Fisher & Aguinis, 2017). In theory elaboration, the researcher focuses on the contextualized logic of a general theory to elaborate on the theory itself. Theory elaboration is typically utilized when a researcher may be able to apply an existing general theory, but the context is not known well enough to obtain sufficiently detailed premises to be used with the general theory to deduce testable hypotheses. Thus, empirical data are used not only to test a theory but also to challenge it. Furthermore, in theory-elaborating case research, the researcher identifies a general theory that can be used to approach the empirical context (Ketokivi & Choi, 2014).

There are various ways in which theories can be elaborated including introducing new concepts, conducting an in-depth investigation of the relationships among concepts, or examine boundary conditions (Whetten, 1989). In theory elaboration, the researcher investigates the general theory and the context simultaneously, in a balanced manner. Thus, the purpose of theoretical elaboration is described as reconciliation (synthesis) of the general with the particular. In theory elaboration, a general theoretical logic is applied without anticipating the empirical findings by a priori formulation of propositions (Ketokivi & Choi, 2014). Merton (1957) proposes that in theory elaboration, case research remains open to unanticipated findings considering the possibility that the general theory requires

considerable reformulation (Alvesson & K¨arreman, 2007). Theory elaboration can be described as the iteration between general theory and the empirical data (Merton, 1957). "Ultimately, theory elaboration lies in between theory testing and theory building. It seeks situational groundedness using similar logic as grounded theory, with the exception that it engages in more theoretical abstraction. In establishing a sense of generality, theory elaboration relies on general theory as well (Ketokivi & Choi, 2014, P. 5)."
The *SECI* framework provides a foundational concept for further theoretical elaboration. According to Curado & Bontis (2011), *SECI* is underdeveloped and more empirical and qualitative analysis is needed to enhance the theoretical contribution. Thus, this study collects qualitative data in order to integrate existing *SECI* theory through the integration of the 'people', 'process', 'technology' as overlapping components of knowledge management and digitization in the context of hospice (Fisher & Aguinis, 2017). Through theoretical elaboration, this study seeks to address the following research questions of 'how do organizations conduct socialization, externalization, combination, and internalization?' and 'how does digitization induce and/or maintain movement between the four modes in the *SECI* model?'.

Qualitative Analyses

Qualitative research approach makes contributions to the literature by adding to the theoretical knowledge such as using a theory-generating approach (Johnson, 2015; Curado et al., 2019). Qualitative study relies on meaning-based analysis of data rather than numerical analysis. It emphasizes on measuring things by words rather than measuring by numbers (Elliott & Timulak, 2005). "Qualitative study emphasizes on understanding phenomena in their own right (rather than from some outside perspective); open, exploratory research questions (vs. closed-ended hypotheses); unlimited, emergent description options (vs. predetermined choices or rating scales); use of a special strategies for enhancing the credibility of design and analyses; and definition of success conditions in terms of discovering something new (vs. confirming what was hypothesized)." (Elliott & Timulak, 2005, P. 147).
When research questions of a study seek to explain 'how' a certain phenomenon occurs, the qualitative analyses, sometimes done through case studies, become more relevant (Yin, 2009). In addition, as the purpose of this study is to discover the perspectives, experiences, beliefs, and individuals' motivations on specific important issues e.g. in our study we are discovering the role of digitization in knowledge creation and in a specific industry, the qualitative study is conducted. Qualitative study with unstructured interview provides a 'deeper' understanding of the phenomena rather than purely quantitative methods, such as questionnaires. This study examines an industry where interdisciplinary communication is of high importance in the creation of value. Specifically, this study targets the hospice industry.

Method of Sampling and Control Variables

The National Hospice Locator (*NHL*) was used as a starting point in the initial collection of potential interviewees. The *NHL* provides information about every known hospice in the United States. Information is gathered from medicare data files; numerous hospice, state, and national public websites, and from surveying hospice directly (NHL, 2020). Detailed information such as address, email, size of the hospice, accreditations, can be found in this website. Since this is a qualitative study, no certain sample size is determined and the interviews continue until theoretical saturation occurs, (i.e. when marginal improvement and

learning become trivial and phenomena become repetitive (Lewis, 1998; Eisenhardt, 1989). In selecting the sample, accreditation (*CHAP*), type of hospice program (Home Health Based), and medicare certified hospital, home health, skilled nursing providers (Home Health) are controlled. "*CHAP* is accreditation to evaluate processes to assess the quality of health care providers. A *CHAP*-accredited organization has demonstrated that it has met regulatory requirements and nationally recognized standards (CHAP, 2020a). The rigorous evaluation by *CHAP* requires unscheduled in-person visits to hospice locations to evaluate the program's clinical care quality, patient and family experiences, and processes for ensuring long-term viability (Compassus, 2020). Accreditation leads to better quality care and betterbusiness performance (CHAP, 2020b).

We control this factor since it affects the communication process based on the defined policies and procedures. Here are some examples of these policies; It evaluates communication between the attending physician and other members of the interdisciplinary team is ongoing and documented in the hospice clinical record. It checks whether the hospice facilitates communication about changes in the patient's status between interdisciplinary team members and the patient'sattending physician. It also ensures establishing and maintaining effective channels of communication, including integrations of technology. *CHAP* accreditation also includes policies for the responsibility for coordination of care and fostering communication between the patient and the interdisciplinary team. It also assures communication between the hospice and other health care providers involved in the patient's care. It identifies family dynamic, communication patterns and assures special needs related to cultural diversity including communication, space, the role of family members, and special tradition. Furthermore, it emphasizes on the communication of information to appropriate interdisciplinary team members and demonstrating knowledge in communications (Group, 2003).

We also controlled for type of hospice program by selecting 'Home Health Based' hospice. Home health based hospice provide services to patients anywhere they call home, including assisted living communities or nursing homes (Vitas, 2020). In home hospice, nurse communications are considered as triggers or prompts that are operationalized as a "distinct change of subject, direction, or opening to a discussion" (Ellington et al., 2012). Hospice care at home requires many responsibilities, from medical care, communication, and taking care of patient self-care and the household. Furthermore, considering regulations for this type of hospice, doing a survey is less restrictive than the other type of hospice. With the control variable inputs, the initial sample yielded 92 hospice service providers on the *NHL* website.

Data Collection Procedure

The primary source of data collection for this research are semi-structured interviews. Protocol questions have been developed through pre-existing literature i.e. (Nonaka, 1994; Nonaka et al., 2003; K'o et al., 2019; Fernando et al., 2019; Bongku & Kurniawan, 2019), then altered and changed from comments provided by two experts in the supply chain digitization and knowledge management fields. We obtained institutional review board (*IRB*) procedures for the further development and approval of the protocol.

Interviews were conducted via phone calls. All members of the sample were contacted by telephone to have a convenient, fast, and secure participation. After contacting and explaining the purpose of the interview followed by *IRB*, they were asked to introduce to anyone in their interdisciplinary team who is available and willing to interview. The interviews were guided by the developed protocol including a brief purpose of the study, vocabulary (terminologies used in interview questions), and a range of questions. Four

questions addressing each of the four modes i.e. socialization, externalization, combination, and internalization and four questions addressing the role of digitization to induce/maintain the movement from socialization to externalization (*S-E*), movement from externalization to combination (*E-C*), movement from combination to internalization (*C-I*), and movement from internalization to socialization (*I-S*).

Pretesting reduces measurement error and determines whether interviewees are understanding and interpreting questions correctly, and to make sure that the order of questions does not influence the way of answering by a interviewee (Bolton, 1993). The interview protocol was first pre-tested by twenty professionals. The pre-test included individuals from Non-profit organizations, manufacturing, consumer goods retail, healthcare and security industries. Numerous industries were included to provide a broad range of how questions may be interpreted, given the interdisciplinary nature of the hospice industry. However, special attention was paid to the healthcare and service industry interviewees, to ensure that the questions being asked accurately to reflect the information we intend and that the interviewee can and will understand and answer the questions. Based on the pre-test results the interview protocol was adjusted accordingly.

The interviews were conducted during Spring 2020 and each interview lasted between 20 and 30 minutes. Follow up questions were asked considering the semi-structured nature of the interview and the flow of conversation, Interviews were conducted with 6 interviewees in 5 hospice care providers. All interviewees were hospice managers and one intake nurse was also interviewed from Hospice D. To comply with the anonymity per *IRB* approval, fictitious names were assigned (e.g., A, B) to the participating hospice care providers. A detailed profile of each hospice is presented in Table 1. The unit of analysis is company.

Table 1: Profiles of the interviewed care providers

Hospice Name	Hospice Ownership	Number Served Daily
A	Nonprofit, Gov, or Other	Medium (50-124 pts/day)
B	Nonprofit, Gov, or Other	Medium (50-124 pts/day)
C	For-profit	Extra Large (200+ pts/day)
D	Nonprofit, Gov, or Other	Large (125-199 pts/day)
E	Nonprofit, Gov, or Other	Extra Large (200+ pts/day)

All interviews were recorded and transcribed and coded. Transcripts were coded by the author and also by one of the supply chain colleagues. Following to coding process, in order to assess the inter-rater reliability, Cohen's kappa coefficient (k) which is a statistic that is used to measure inter-rater reliability was calculated. Cohen's kappa measures the agreement between two raters who each classify the quotes. We got 86% for inter-rater reliability which shows a strong agreement (McHugh, 2012).

In addition, several alternative data sources from the January 2019 to April 2020 were collected for the purposes of triangulation (Yin, 2009). These materials include interviewed hospice websites, their news, and publicly available material on hospice industries. These data sources allowed us to validate what is said during interviews as well as provided more insight into the subject during interviews.

Validity and Reliability

In order to evaluate the quality of the study, construct, internal, and external validity, and reliability were tested (Tate & Ellram, 2012; Levinthal & Rerup, 2006; Yin, 2003). Also, these

sources recommend that construct, internal, and external validity and reliability should be evaluated throughout the qualitative data. Table 2 addresses the qualitative study tactics and a brief explanation of their implementation of validity and reliability issues.

Table 2: Overview of reliability and validity in qualitative study (Reprinted from "Service supply management structure in offshore outsourcing" by Tate & Ellram, 2012)

TEST	DEFINITION	TACTIC	IMPLEMENTATION
Construct validity	Tests whether the research measures what it is supposed to measure. Also, establishes the correct operational measures for the constructs being studied. Focuses on the extent that conclusions can be drawn for casual effects and establishes a causal relationship.	Use multiple sources of evidence, Establish a chain of evidence, Key informants review draft of qualitative study report.	Gathered several documents including companies' websites, news, mission, and approach documents, the document of CH AP hospice operational policies, and procedures.
Internal validity	Focuses on the extent that conclusions can be drawn for casual effects and establishes a causal relationship.	Pattern matching. Explanation building, Rival explanations. Logic models.	Investigated patterns such as Approach, mission and values, core team, and hospice services and communication processes, Examined relationships and communication process and involvement of interdisciplinary team members.
External validity	Looks at whether the research results can be applied to the populations and the settings of interest. Then, establishes a domain in which the studies' findings can be generalized."	Use replication logic in multiple cases.	Conducted qualitative study with five home health-based companies.
Reliability	Demonstrates repeatability.	Use interview protocol, Develop interviewees' database"	Implemented qualitative study protocol with all companies, Created a qualitative study database.

RESULTS

The conducted interviews are used to give interpretation and meaning to the qualitative data. Content analysis is a suitable technique to analyze the interview responses (Hsieh and Shannon, 2005). This study utilizing content analysis identifies certain keywords, concepts, themes, or sentences based on the coded transcribed interviews. This interpretive research method helps to assess the contents of recorded interviews. Also,

"content analysis can be used to analyze newspapers, advertisements, recordings of the interviews, and the like" (Sekaran & Bougie, 2010, P. 386).
"Interpretive analysis of qualitative data provides a detailed description of events, situations, and interaction between people and things that are related tothe phenomenon under investigation" (Bryman and Bell, 2011, p. 27). Content analysis allows researchers to deeply investigate the attitudes to provide an understanding of drivers, and values of a particular context in a given framework through detailed descriptions (Easa, 2012; Denzin et al., 2005).
Content analysis of transcribed interviews was done manually. All quotes in the interviews were coded first. For example, in the coding process, we utilized *S*, *E*, *C*, *I* as representative of socialization, externalization, combination, internalization, and *S-E, E-C, C-I, I-S* as representatives of movements from each mode to the other mode (e.g. *S-E* : socialization-to-externalization) for each quote in each question. For example, for question 1 (how your organization conducts socialization) the first quote addressing socialization was coded as *S* (A).1 in which *S* comes from Socialization, 'A' means company A, and 1 means the quote comes from question 1. We coded all interviews in this way, then all codes are put into our research model. After coding, the research model was mapped based on the coded transcripts. This way, we can see where we lie in terms of cross-case and within-case analysis (Eisenhardt, 1989).
At the next step based on the created map from the codes, keyword tags were assigned to each mode and for each movement. For example, Company A: socialization quote 1: 'We mostly conduct socialization by talking with one another through meetings.', two keyword tags can be extracted as 'talking' and 'meeting'. In order to maintain validity and reliability of the research as mentioned in Table 1, synonyms for each of these keywords were found to be searched in interviewed organizations' websites and news (Tate & Ellram, 2012; Ellram, 1996; Yin, 2003).

***SECI* in the Home Health Hospice**

To add to theoretical knowledge, the following qualitative data which presents the descriptive interpretation of the transcribed interviews, provide an overview for the exploration of the relationships of digital technologies on *SECI* framework.

Socialization

Socialization is defined as face-to-face communication and discussion of shared experience among organization members to exchange tacit knowledge. The socialization process motivates organizations to provide enough space and time for interaction and communication among individuals (Curado & Bontis, 2011). Interviewees acknowledged that socialization could be achieved through informal talks, formal meetings, and training programs. All interviewees stated that interdisciplinary team meetings and daily face-to-face talks play an important role in the communication process and getting information, sharing experiences, and updating every member of work. They acknowledged that the majority of the socialization process occurs via face-to-face communication. Narratives can come through face-to-face communications. Thus, they are providing opportunities for their interdisciplinary team to frequently communicate with each other this way. One such example is described by Company E and Company B's managers are quoted as below:

> "Most of the value-added conversation comes from face-to-face talking. I mean we use email and computers, but that's mostly transactional. The really deep information you get from face-to-face communication, not through computers."

> "We hold team meetings bi-weekly with the entire team for event coordinator volunteer, chaplains, social workers, nurses, the manager, myself, the physician, etc. We all go around and we speak on each patient, identify any means, any concerns, troubleshoot any issues we're having together as a team."

All of the interviewees indicated since hospice is a service industry, having faceto-face communication along with body language and sharing personal thoughts is fundamental. Having team meetings help to identify potential problems and successful interventions designed to provide a comfortable environment for the patient and their families as well. From a digital perspective, typically legacy technology (i.e. phone meetings, paging systems, email) are utilized in these processes. One of the interviewees referred to the newsletters as a tool for making daily talks and conversations. All interviewees stated that they use electronic medical records as their method of providing important information regarding patient updates. One such example is a quote from Company A's manager:

> "We utilize some of that over the phone. We use Epic as our electronic medical records. We send messages through Epics to one another. Our social workers and chaplains may join to visit the patient. We also have a paging system, so, if the nurses or social workers need to get the information to another discipline right away, make either page or call. Our nurse managers and our director are also central to making sure that information gets out on the patient system since the patient calls in."

Externalization

In terms of externalization, the majority of interviewees utilize particular mechanisms and tools including electronic medical records, interfaces, Microsoft, patient portal, and internal documentation system, to record and keep records of information. Both technology and handwritten communication were mentioned. Specifically, from the technology standpoint, staff utilized electronic dictation devices for ease of recording information. However, handwritten reports and minutes were also still utilized.

During the interviews and the secondary data collection it was made evident that the mission and interdisciplinary team approach was highly integrated into the corporate strategy for value creation. Specifically, goals were centralized on utilizing the interdisciplinary team approach to provide value defined by patient quality of life in end of life care. Provided this strategy, the willingness of transferring tacit into explicit knowledge is high and sometimes obligatory since the entire team seemed to be centralized on one strategic priority as defined by quality of life for the patient. These processes included providing continuous updates from each interdisciplinary member's visits to all members via electronic medical records keeping. This was described by Company A, Company B, and Company C's managers as below:

"For all of our committee meetings, we have a secretary who documents the meetings. We have electronic medical records in order for patient information to be utilized depends on how and what type the person is."

> "The way we typically work is that when we sit down as a team, the care plans, the medications are all listed on there, the entire team writes a note specific to their discipline, everybody writes notes there. We typically utilize computers or handwritten documents. We have to write up a report for each patient's visit etc."

> "We do primarily electronic medical records. We also have access to dictation

devices, so, we can talk to text (Speech to text programs such programs take sound recording of a person talking and dictate the sound recording based on the frequencies it picks up from the sound)."

Combination

The combination process transforms explicit knowledge into more complex forms integrating the thoughts of a team in a meeting from the shared documents between them, updating, reviews, categorization, and databases. The majority of the interviewees stated that their internal information, meetings' minutes, and reports are all recorded via a centralized database through electronic medical records. All records are properly updated and reviewed by the interdisciplinary team members. The interviewees confirmed that they daily use the updated reports and patients' visits, and then circulate them to the interdisciplinary team members using an internal documentation system, email word of mouth, and/or messaging system. These distribution systems are then used during urgent situations to provide ongoing information to team members. The use of this technology to aid in specific processes is outlined below by Company B and Company A's Managers:

"On a daily basis our entire team reports of every patient they saw using HIPPA compliant internal voice messaging system as well as a direct call for any urgent need, for example when one of our nurses or social workers saw a patient today, they leave a report about their visit today, any issues, anything new, any changes, etc. If there was an urgent need, for example, a nurse goes out and the patient is now transitioning, she not only leaves on the voice message, but she also contacts the social worker, the chaplain directly, physician, and home health, and everybody in the team is aware of transitioning status. Most of our disciplines can get out in a timely manner. We really strive to make sure that, especially in difficult times that our entire team and family have been out to see the patient and make sure that we addressing any support needs and also management needs."

"We also talked about in our DIT meetings (also known as DIT meetings, stands for Daily Interdisciplinary Team). It was distributed by word of mouth and then how the plan was going to work along with the message through our internal documentation or our messaging system, we email word of mouth and meeting."

Combination process is also done through coordinating the care among the different disciplines via gathering subcommittee together to work together by planning and distributing this information throughout the interdisciplinary team. Theinterview subjects described the combination process and specifically mentioned coordinating volunteers as central to the process driven by the interdisciplinary committees and subcommittees. As evidence, Company A, manager and Company D's nurse mentioned as below:

"When there are visits by the staff in the last seven days of life, our mission was to start coordinating the care amongst the different disciplines and getting those visits done. We got a subcommittee together, we worked through, we came up with the plan, and then our plan was distributed each discipline. We met as a whole staff and talked about it."

"We have a great group of volunteers that help us with daily tasks and give us information that the patient is comfortable speaking to them. Our line of work

coordinates through interdisciplinary committees."

All interviewees emphasized the importance of coordinating the interdisciplinary team and all involved stakeholders in order to develop strategies for supporting a patient's therapy in the combination process. Growth is dependent on this coordination process. This was made evident by Company D's nurse and Company E's manager as follows:

> "We build on the interdisciplinary team. We have therapy assistants who need to work directly with nurses. This is done to develop strategies for supporting a patient's therapy. So for example, if the nurse is in the home and the therapist works with that individual on improving upper body strength, how does the nurse incorporate that when she is doing her skills and asking the patient to do certain things as she is providing her skills? How do you keep that whole patient in mind, and use the whole team?"
> "In hospice, it's all based on a team approach, within our place we have our providers or the people who provide the service, but it's not just about that. We also have to coordinate with the patient, the patient's family or caregivers, their doctor, we have to coordinate with the care center, hospital discharge planner. In order to grow, you have to know how to coordinate."

Integrating involved people's thoughts via coordination among members is also important in the combination phase according to the interviewees. However, there is not a lot of virtual back and forth discussion as they mentioned. One such example was described by Company C's manager and Company D's nurse as below:

> "When we have a meeting, there is a discussion between team members, and we try to integrate whatever people are thinking and we write down that based on the coordination of the team members. The key is coordinating and not only that but finding the right mechanism for coordination."

> "A lot of it in our line of work is through direct communication. I mean there's not a lot of back and forth, virtually. We have a great group of volunteers that help us with daily tasks and give us information that the patient is comfortable speaking to them. Our line of work coordinates through interdisciplinary committees."

Based on the presented qualitative data, there are many advantages from the combination mechanisms, e.g. coordination among team members, continuous updating of databases/reports based on their daily visits and activities and distributing these updates to the interdisciplinary team and involved people. Thesefindings underline the importance of the combination process in home health hospice.

Internalization

According to the interviewees, the internalization process in home health hospice supported by applying new knowledge via their disciplines. When different disciplines come together and later internalize the knowledge this enhances: (1) feedback from the patient and their families, (2) individualized approaches to different patients and (3) a radical different in the quality of care for each patient. Previous research supports this finding by suggesting performance implications via learning by doing, performing, reflecting and simulation (Chatti et al., 2007). Furthermore, action is deeply rooted in internalization (Nonaka,

1994). As an example, Manager from Company A mentioned:

> "We give an example of when we were successful in those visits and it showed how different nurses or different staff came up with their way of doing and how they are going to talk to the families and get the other members in and we showed there was a success and how the families responded positively to all of our visits and we showed when there were not those visits and where we could have it improved. For more clarification: We use an example of a patient of who got multiple visits on the end of life and how the family responded positively to it, they were so grateful, they turn out they really did need the social worker, they didn't realize that and we showed how doing this actually make the big difference whereas just telling it's going to make the big difference that doesn't mean it's really will, like nobody knows for sure if it can make a big difference or not. So, we showed how it did make a big difference. And we talked about the patients who did not get those end of life visits and whether we could have improved upon."

Most interviewees stated that they support internalization by providing updated and integrated reports including new ideas, meetings' minutes and agendas prior to each meeting to the interdisciplinary team. This allowed them time to reflect and bring specific questions and ideas to the next meeting. As an example, Manager from Company B mentioned:

> "For example, when we have a staff meeting, an agenda is printed out beforehand with points of exactly what we're going to cover and exactly what the new expectations are as of a new process that is being put in place. Everybody could get a copy of that communication before the meeting usually 24 hours before the meeting so they can come with questions and then go around the room and troubleshoot any issues that may come up, we then go ahead and I take any note and I resend it out with all of the any type of additional ideas people have or specific answers to questions they will have."

The interviewees also highlighted that their organization supports internalization mechanisms by allowing every discipline access to information and outcomes via electronic medical records and interfaces. Observational differences were also mentioned as supportive of the internalization process. One such example was described by Company D's manager as below:

> "My immediate thoughts are on the observational differences. Patient functional impairments are assessed in different ways by different individuals. Therapists assess by observation not by interviews with patients. This is very different than medical doctors."

Interviewees also mentioned that they motivate individuals to make the best decision for themselves by enhancing their own receptivity to complex and differing ideas. As an example, Manager from Company E mentioned:

> "Individuals have to make the best decisions for themselves. I mean what I think is good for the patient may not be what the family, friend, primary caregiver think is appropriate. What I tell people here, is that my door is always open. If there is any questioning of the quality practices of this facility, or if I'm not doing something right, my team or the patient team is always welcome to

> stop by. The leadership team needs to hear you out and either convince or change."

Socialization – Externalization (SE), (Tacit to Explicit)

According to interviewees, there are several drivers to support the movement from socialization to externalization (i.e. *S-E* process). The drivers are not the individual people, processes or technology but rather the integration of the three facets that allow for the *S-E* process to occur. The people category refers to the interdisciplinary team which involves managers, staff, volunteers, social workers, nurses, chaplain, etc. who communicate together to share their knowledge and new ideas to provide the best possible experience for the patient and their loved ones. As an example, Manager from Company A mentioned: "We do this (*S-E*) in our staff meeting and in our individual and interdisciplinary meetings. We talk about what works out for us individually and then we talk about how we can all utilize what works out for all of us and that's written down in the minutes and the plan is put forward to utilize the new idea."
Specific processes in this category include sending messages, face-to-face communication, informal talks, interdisciplinary meeting, staff meeting, patient visits, discussing ideas for improvements via the idea ballot box, and quality assurance with shared information. This was made evident by Company D's nurse and manager as follows:

> "One of the things I can think that this is done is through an idea ballot box. Sometimes we get together to discuss ideas for improvement. And we put those ideas in a ballot box and share them at the end of the month."

> "We use the cloud for most applications. Specifically sharing information among patient committees, what happens more than you think, is that information gets jumbled up, so a ton of checks has to be done."

Specific technology mentioned were electronic medical records, dictation devices, messaging systems, internal voicemail systems, internal documentation systems, computer programs (*Microsoft*), patients portal, and cloud computing technology. As an example, Manager from Company A stated:

> "We use our electronic medical records to coordinate communication and visits. We all have laptops and we are taking every patients' home. And that's how we send messages to one other and how we see each other's charting. That's how we can get the messages that the patient is dying."

Our qualitative analyses revealed not only insights into the integration of these three pillars as driving the *S-E* process but also insights into potential inhibitors. Specifically, according to interviewees, *S-E* drivers involved people on the interdisciplinary team working jointly to provide required documents for consumer assessment of healthcare providers and systems, accounting, and auditing. What was necessary included foundations for efficient recheck of data inputs, and more aligned record keeping in order to promote adequate transfer to externalization. Interviewees also provided insight into the role of interpersonal trust as well as technology-based trust. Concerns about privacy of data were also mentioned as possible inhibitors. This was made evident by Company E's manager as follows:

> "There's an increased focus on documentation in the hospice industry since

> regulatory tightening. Auditors look for the documentation of whether or not a patient is right for hospice. Oftentimes documentation will be incomplete forcing a recheck of the data collected. Digitization might help through better record keeping. I think when you use the software, it helps calm the nerves in case of an audit."
>
> "When you digitize face-to-face communications, it becomes easier to share and more likely the data will get into the wrong hands. You have to be really careful here with privacy."

Externalization-Combination (*E-C*), (Explicit to Explicit)

The interviewees confirmed that the three pillars integrate to create a driving force inducing movement from externalization to combination (i.e. *E-C* process). As previously mentioned, the people category remained the same. Specifically, those individuals comprised of the interdisciplinary team were discussed in the people category.
According to the interviewees, *E-C* processes include distributing information (e.g. meetings' minutes), verbal and face-to-face communication, discussing ideas, reviewing reports and documents, collaboration among involved people, coordinating communications and meetings, coordinating visits, virtual visits and meetings, and integrating existing knowledge with the past experience and external knowledge. There are also many verbal communications between team members. One such example was described by Company D's manager and nurse as well as Company E's manager as below:

> "One example I can think of is through nursing visits. Nurses can conduct virtual visits to reinforce improving the patients' ability to perform Activities of daily living (ADLs). The therapist can also be linked in. I mean we really struggle to collaborate well. Maybe because we are not all in the same facility, we are not all in the same place with the same patient, we are all in our cars going places to get there. The best practice here is that when we do our assessment, we make sure we are collaborating with our whole team to try to make the plan together, and we are all working toward our patient goal."
> "Nurses naturally discuss ideas with one another, if accepted this will then be discussed upwards. But not usually."
>
> "We can review everything to make sure. We also have a lot of verbal communications between team members, me, and the physician."

In the technology category, all of the interviewees referred to some devices and tools such as electronic medical records and interfaces to support reviewing everything. As an example, Manager from Company C described:

> "We utilize electronic medical records and we all have access to patient information and internally, we also use interfaces."

From the process perspective, one of the interviewees using a coordination example of a therapist-patient relationship suggested *E-C* is the biggest hurdle for them as the therapist is expected to combine external knowledge and experience to provide the patient energy conservation techniques. Evidently, a potential driver, appears to be stakeholders' motivation to move from simply externalization knowledge to combination with other

stakeholders on the team. In the following example, motivation seems to be lacking from both stakeholders thereby inhibiting the *E-C* process. One such example was stated by Company D's manager as follows:

> "I think this is our biggest hurdle. For example, in therapist-patient relationships, when we assess physical activity. We would hope that the therapist is taking into account 'what has been familiar activities?' and 'what have been things they have done, in the past, things to maintain a certain level of activity?' And also, what can we do to provide some energy conservation techniques? vs. throwing them a home exercise programs, then let's be honest, they're not doing those."

One interesting finding was interviewee suggesting digital technologies may actually impede the *E-C* process by standardizing electronic documents without the allowance of personalized, qualitative and/or visual evidence (i.e. body language). Many of our interviewees confirmed the *E-C* process occurs typically through face-to-face communications. For example, Managers from Company A and D, and a nurse from Company D stated:

> "If you go from individualization to externalization, that means everybody has got the same form to fill out because we took your individual thought and we globalize it, right? But then once it globalized and we all are all filling out this form, it can become routine and just like fill in the blank kind of a form instead of making it personalized, because the form is on the computer and it's just you fill in the blanks. I think it (digital technologies) takes it away, I think you go from being able to take your own notes and your own thoughts, to just writing the thoughts that want to be known on the form."

> "Coordination in our operations really happens face-to-face. I don't really see any connection here to the digital world."

> "This is mostly just communication. I mean face-to-face communication."

Combination-Internalization (*C-I*), (Explicit to Tacit)

The integration of the three pillars to create driving force inducing the movement from combination to internalization was confirmed by the interviewees. The 'people' category like previous movements comprises of those people who are involved in the process of flowing and creating information and knowledge including interdisciplinary team, managers, staff, and the patient. The interviewees alluded to a few inhibitors of the *C-I* process. Specifically, motivation of the people category was mentioned. A number of interviewees commented about the capability of stakeholders as well as the motivation to absorb collective knowledge for the movement to internalization. Both adequate education and training seemed to be a common theme mentioned to help alleviate the inhibitor of capability. This allows for effective absorption of collective knowledge individually requiring time to process collective knowledge apart from the group, but also the motivation to do so. This was made evident by managers of Company A, B, and D as follows:

> "I think it's really helpful, because it gets the information out to people faster and they can read up on the information that we give them on their own and think about on their own and they are not stuck in a big group meeting trying to

> figure this out to, so, it gives some time to review the information individually."
>
> "It's hard to say whether or not each team member absorbs collective knowledge. But I would say that the best way we monitor that is through making sure that we do chart reviews that the education and training they provide are regarding. Making sure that they are communicating, medicating, treating, observing, or based on the training that we've provided as a company we can see when we do our chart reviews."
>
> "Again, there is no internalization. It's very important to follow procedures."

Motivation can be enhanced by face-to-face communication particularly in the hospice industry where the urgency and importance is relayed by the relationship you develop with your colleagues and patients. Interviewees described how digital initiatives can impede. As an example, Manager from Company E explained:

> "Communication, I mean face-to-face communication can be so much faster in my opinion. Often with digital initiatives, the meaning gets lost in the digital world. You lose the ability to read body language. And in the service industry that's often more important than digital information which is more transactional in nature. I think digitization can impeded individuals absorbing important pieces of information."

Internalization- Socialization (*I-S*), (Tacit to Tacit)

The interviewees in the process of movement from internalization to socialization, confirmed that the integration of the three pillars promotes drivers in the *C-I* process.
As mentioned in the previous movements, the people category remained the same and it is comprised of the interdisciplinary team, manager, and staff who are involved in the process of creating and flowing information and knowledge.
According to the interviewees, *I-S* processes includes sharing experiences, interdisciplinary team meetings, daily team meetings, face-to-face communication, informal talks, virtual communication, virtual meetings, and virtual sharing experiences. *I-S* provides learning by sharing internalized knowledge and experience through face-to-face communications, informal talks, and meetings, which leads to value for the patient. As an example one interviewee describes the value of the *I-S* process as follows:

> "So, when team members talk face-to-face there is a lot of interface about shared collective knowledge. For example, if we are sitting around the table and talking about a patient who is maybe very agitated, other nursing professional and physician weigh in on possible medication changes that can be made spiritual care and social work weigh in on maybe non-medication modality talk about a little more about the family and religious belief, maybe creating some of those anxiety non-pharmaceutical approaches to that."

Similar to previous processes, the interviewees described the limiting nature of the 'technology' aspect in the *I-S* process. Specifically describing digitization as a mechanism for a type of limited communication that is more transactional. Hospice is in the category of the service industry that requires a personal touch and face-to-face communication. Digital initiatives may impede the process since the meaning gets lost in the digital world

due to a loss of an ability to read body language. The capability of the technology was referenced here as a limiting factor. Another interesting insight was in regard to the people category and specifically personality. Motivation and capability of the individual to share information is essential in the *I-S* process. As an example, Manager from Company E stated:

> "This seems to be more of a personality trait in individuals. Some people are more willing to share than others. I mean you see this in patients, some just want to be left alone, while others will talk your ear off."

DISCUSSION AND CONCLUSION

Based on the qualitative analyses, we derive the following research model in depicting a visual representation of the research question of 'how does digitization induce and/or maintain movement between the four modes in the *SECI* model'. Figure 2. shows the research model based on *AMC*.

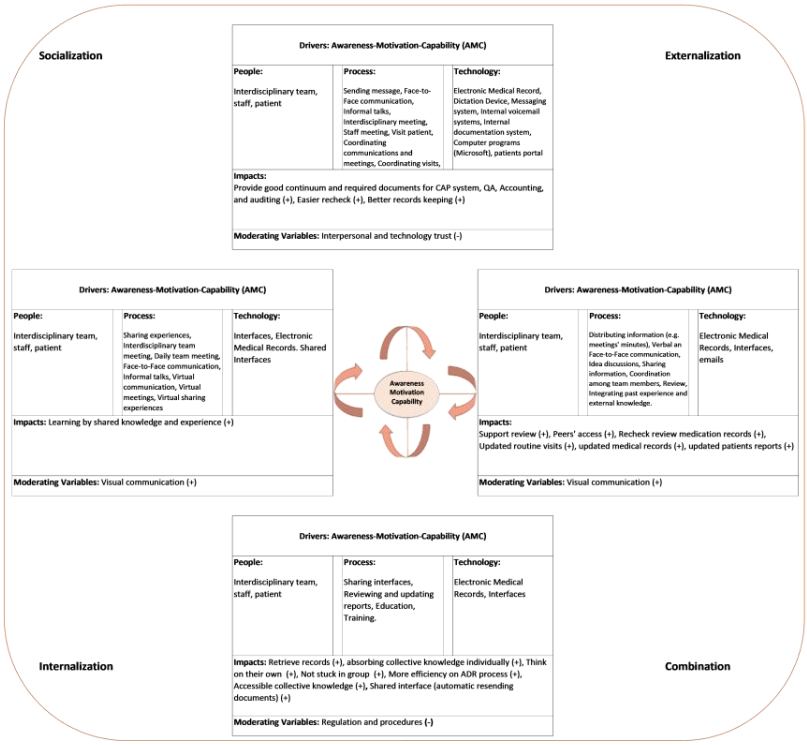
Awareness, motivation, and capability are all foundations of success in organizational action, when one is missing it is likely others will be hindered. The awareness, motivation and, capability model (*AMC* model) is a compilation of decades worth of competitive dynamics research. Simply stated, competitiveness takes place when a firm is aware of an action, motivated to act, and has the capability to respond (Udenio et al., 2018; Mitsuhashi & Alcantara, 2019; Chen & Miller, 2012). Typically resulting in competition, competitiveness is defined as the degree to which a firm can create a defendable position over its rivals (Porter et al., 1985; McGinnis & Vallopra, 1999). The operationalization of the three elements presented in the *AMC* model has differed considerably. Awareness refers to organizational alertness of the market as well as the signals it receives (Levinthal and Rerup, 2006; Lamberg et al., 2009; Udenio et al., 2018). Motivation is defined by the incentives which drive a firm to act (Smith et al., 2001; Udenio et al., 2018). Finally, capability refers to the cognitive and resource-based factors which influence a firm to act (Smith et al., 2001; Udenio et al., 2018)).

Our qualitative analyses provide us insight into the people, processes and technology that promote knowledge creation through the *SECI* model driven by *AMC*.

As already noted, it is not specific people, processes or technologies that drive the knowledge creation process but rather the integration of these three pillars. In other words, based on knowledge organization perspective, people exchange knowledge utilizing technology and through an established process (Chugh et al., 2013). In further analyzing our qualitative data, what also became apparent were specific drivers and inhibitors of these three pillars. These are also outlined in awareness, motivation, capability (*AMC*) framework literature.

The interviewees alluded to a few inhibitors of the *C-I* process. Specifically, motivation of the 'people' category was mentioned as a driver for movement in the *SECI* framework. One specific example in this category is the motivation of the patient and the therapist to

Figure 2. Derived research model



absorb collective knowledge. In fact, this was related to the capability dimension through the mentioning of adequate education and training of the interdisciplinary team in order to ensure the conversion knowledge. Both adequate education and training seemed to be a common theme mentioned to help alleviate the inhibitor of capability.

Capability of the 'people' category and 'technology' category were deemed to be drivers/inhibitors of the movement in the *SECI* model. Specific examples of this involved interdisciplinary team collaboration and sharing knowledge in the movement from socialization to externalization (*S-E*). One privilege in home health hospice is that they benefit from any type of communications including face-to-face communications, informal talks, interdisciplinary and staff meetings, and patient visits. This is strategically designed to align with the mission of the home health hospice.

Capability of the 'process' category to drive the movement in the *SECI* is also evident. Important processes in knowledge creation and conversion is involved in communications and meetings, reviewing reports and documents, distributing information, collaboration and coordination among the involved people, integration of existing knowledge with the past experience and external knowledge. Particular importance is on the nurse-therapist communications for the purpose of combining their existing knowledge and integrating with external knowledge to provide the best service for the patient. However, the capability of the technology can often impede this process through the standardization of electronic documents removing the ability to read non-verbal and/or digital language. The value derived from patients is often dependent on the capability to derive information from non-verbal cues.

Another aspect, aside from the *AMC* model was interpersonal trust as well as the trust with the technology being important drivers in the movement of *SECI*. Trust in technology may

refer to the trust of the technology to keep patient information safe and secure. In addition, it may refer to the trust in the technology's capability to provide value. Trust in the individual referred to both privacy and capability.

In summary, our findings indicate that people, processes and technology separately can only provide minimal value, whereas the integration of these processes driven by motivation and capability in people, processes and technology, can provide value in the complex interdisciplinary environment definitive of the hospice industry. In addressing more specifically how digital technology may be able to be leveraged for enhanced value, the following recommendations is provided based on specific weaknesses identified.

Weakness 1: Lack of privacy, data security, and trust

Record keeping and continuous updating reports and documents in the hospice sector plays an important role as it facilitates easy access to patient's information, saves time by reducing duplications, and also supports managing the patients in an efficient manner (Senanayake et al., 2017; Azogu et al., 2019). A good record keeping system is also important as it ensures information adequacy for the quality of life for the patients and their families. Record keeping is not only important in providing update by the interdisciplinary team for ensuring continuity of care, but also in providing the required documents for some processes such as insurance, auditing quality assurance, and other services. Results confirmed that most of the record keeping is done through electronic medical records, which help in coordinating interdisciplinary team activities and all involved people in the care system.

Although record keeping and updating through electronic medical records has provided many advantages for hospice such as convenient and continuous documentation of care records, reducing redundant services, enhanced decision making and coordination, the system still needs improvement which arises challenges for hospice such as privacy issues. Privacy issues have ethical and legal consequences. Privacy concerns in hospice may result in reducing patient's trust regarding properly protecting their health and care records and keeping confidentiality. This may also further results in reducing public trust in healthcare delivery system (Azogu et al., 2019). The need for trust is becoming more vital in an increasingly digitized environment considering the increased need to store private data in hospice.

Another issue is the security of the care information. Results shows that caregivers use their laptops or smartphones in which they carry patients' record on it to send and receive personal emails and messages in a proper time to the other members of the team. This might expose the patient's information on the unauthorized hands or to be hacked by the people to access the care records of the patients (Rezaeibagha, 2013). Consequently, it can be inferred that information insecurity in hospice service delivery reduces the trust of individuals in the electronic medical
records systems.

The challenges listed above covers the most important issues in the hospice industry. Fortunately, digitization by providing digital technologies as a potential solution can enhance information management in the hospice sector which is elaborated hereunder.

Recommendations: As mentioned in the result section, reports in hospice are currently recorded via a centralized database through electronic medical records. All records are properly updated and reviewed by the interdisciplinary team members. These records are then used during urgent situations to provide ongoing information to team members. Based on the literature on the digitization eco-system, what distinguishes the new era of digitization is the trend toward the decentralization of information sharing and automation.

Blockchain technology is a transparent decentralized distributed ledger for recording transactional data. As it cannot be changed and for any change or correcting an error, a new block should be added to the network and the history of changes are preserved in the system, thus, it is known as an incorruptible digital ledgers technology for a secure and transparent form of sharing transactional data (Tapscott & Tapscott, 2016; Underwood, 2016).

The decentralized nature of the blockchain technology ensures that interdisciplinary team and all involved people in the care network are directly connected to each other. Thus, the information stored in the blockchain is not controlled by any central authority and therefore, the information can be reviewed by the network participants who have access to the system and as a result to ensure proper redundancy (Azogu et al., 2019; Atzori, 2015).

Specific example of blockchain application in hospice industry could be the capability of "giving patients the ability to have agency over and knowledge of who can access their healthcare data (Angraal et al., 2017, P. 2). Blockchain also benefits for decision making in businesses value creation through trust. Blockchain reduces the need for third-party intermediaries by creating transparency between parties in an ecosystem. If a business requires an immutable system, then a blockchain may be an appropriate solution for achieving these purposes (Hughes et al., 2019). Utilizing blockchain technology provides a mechanism to ensure that each block is automatically authenticated and can be easily verified. A specific example in hospice care is the process of sending messages. Hospice care often uses messaging systems to inform the related members for providing a specific care.

A blockchain system, considering its unique characteristics, could provide the opportunity of having a secure and trustable system where activities and updates can be easily accessed by the participants. It could be as potential driver that moves the knowledge from socialization to externalization, then to combination, and finally to internalization.

Weakness 2: Lack of visual communications (body language)

One of the interesting findings of this study was that they suggest digitization may impede the processes of movement in the modes of knowledge creation by standardizing electronic documents without the allowance of personalized, qualitative and/or visual evidence (i.e. body language). Describing narratives can come through face-to-face communications. They believed for hospice as a service industry, face-to-face communications are very important. Specifically, the relationship between team members and patients in urgent and important situations, is verycrucial and they described how digital initiatives may impede this relationship without having visual communications.

Recommendations: Digitization has the potential to provide visual collab-oration by integrated services such as Cisco WebEx telepresence services, video robots, telehealth, etc. Based on the definition of digitization in this study, emergent *IOT* technologies are utilized as enablers to provide a well-connected network and to enhance the digital communications. *IOT* refers to a society where people, devices, and systems interact in daily life via internet connectivity and integrated communication. Based on the definition of *IOT*, 'things' includes people, devices, computers, and digitized systems bridging the digital virtual, and analog physical world (Bolton et al., 2016). People are essential part of the eco system of 'things' within the connected network through *IOT*. The integration of human into the digital *IOT* eco-system, combining artificial intelligence (*AI*) capabilities has the potential to leverage decision making by computerized decision maker on behalf of a human player (Bolton et al., 2016). *IOT* provides an integrated well-connected system that facilitates connectivity of 'any-one' to 'any-thing' in 'any-place' at 'any-time' (Wang et al., 2015). The concept of *IOT* can be extended to the exchange of ideas and information from a service giver to a service

provider within the digital environment. People are essential component of communications systems who drives the dissemination of knowledge (Bolton et al., 2016). In the hospice service, *IOT* connectivity can provide many advantages such as patient's health monitoring devices as well as family and patient access to emergency alarm devices. A plethora of research has been done on the emergence of *IOT* and *AI* applications in many industries since Shah & Chircu, (2018) conducted their systematic literature review. Their study investigates the impact of *IOT* and *AI* in healthcare industry by exploring a wide variety of publications.

Wearable devices through *IOT* connectivity are one of the other important key technological changes in healthcare and patient care. Several capabilities for emergent technologies in patient care including tracking and monitoring of patients through wearable devices, and remote diagnosis are important in providing emergency detection, information management related to therapy, and medical advice (Vongsingthong & Smanchat, 2014).

A systematic literature review in the Shah & Chircu (2018) study demonstrates high emphasis on the use of wearable sensors in the elderly care. They are easily available, easy to use, and relatively low-cost controllers for care of elderly interventions, such as *Wii* remote controller and pressure sensors as Konstantinidis et al. (2015) provided in their study's overview. These devices by integrating smart *TV* s, can provide monitoring of patients and real time alerts to the interdisciplinary team. They can also, aid by promoting exercise through games, thus enabling daily life management and independent living. Furthermore, these devices can provide continuous monitoring and remote diagnosis and data sharing through mobile health devices and applications which is critical for the hospice industries (Sun et al., 2016; Shah & Chircu, 2018; Konstantinidis et al., 2015).

While emergent digital devices such as blockchain, *AI*, and *IoT* wearables can provide many benefits, technology related risks and threats still remains and needs to be more investigated. Although applying technologies such as blockchain, *IoT*, and AI can improve healthcare in various ways, further research is needed investigating the effect of applying these technologies on ethical issues, data collection, legislations, privacy and security, their reliability and accuracy. Furthermore, there might be a need for further customization of these technologies for the purpose of providing more convenient services to patients. These issues need to be addressed in order to successfully. promote value added change in the hospice industry.

In conclusion, while digital technology can provide valuable tools to practitioners in this important and vital industry, technology itself can be a limiting factor impeding value if not integrated properly. It is only through the integration of the people, processes, and technology (driven by awareness, motivation and capability) that hospice care can enhance the value to the patient in the end of theirlife.

Limitations and future work

Considering the reduced sample size and given the exploratory and qualitative nature of the study this lacks generalization power. We suggest future research to develop a fuzzy-set qualitative comparative analysis (fsQCA) (Ragin, 2009). This technique is best-suited for working with small samples (Rihoux & Ragin, 2008) and it has been recently used in KM studies on non-profitable organizations (Li 2018; Wang & Kwek, 2018) and specifically in the healthcare industry (Gonçalves, Curado & Balle, 2019). We recommend the use of this technique to understand the functioning of a concrete and specific complex reality (Ragin, 2009; Fiss, 2011), either in the healthcare industry or other.

List of References

Abernethy, A. P., Wheeler, J. L., & Bull, J. (2011). Development of a health information technology–based data system in community-based hospice and palliative care. American journal of preventive medicine, 40(5), S217-S224.

Abidi, S. S. R. (2001). Knowledge management in healthcare: towards 'knowledgedriven'decision-support services. *International journal of medical informatics*, 63(1-2):5–18.

Abidi, S. S. R., Cheah, Y.-N., and Curran, J. (2005). A knowledge creation infostructure to acquire and crystallize the tacit knowledge of health-care experts. *IEEE Transactions on information technology in biomedicine*, 9(2):193–204.

Agrifoglio, R., Cannavale, C., Laurenza, E., and Metallo, C. (2017). How emergingdigital technologies affect operations management through co-creation. empirical evidence from the maritime industry. *Production Planning & Control*, 28(16):1298–1306.

Ali, N., Sulaiman, H., and Cob, Z. C. (2014). The role of information technology forknowledge management paradigm in higher education. *Journal of Information Systems Research and Innovation*, pages 59–67.

Alvesson, M. and K¨arreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of management review*, 32(4):1265–1281.

Angraal, S., Krumholz, H. M., and Schulz, W. L. (2017). Blockchain technology: applications in health care. *Circulation: Cardiovascular Quality and Outcomes*, 10(9):e003800.

Atzori, M. (2015). Blockchain technology and decentralized governance: Is the state still necessary? *Available at SSRN 2709713*.

Azogu, I., Norta, A., Papper, I., Longo, J., and Draheim, D. (2019). A framework for the adoption of blockchain technology in healthcare information management systems: A case study of nigeria. In *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance*, pages 310–316.

Bolton, A., Goosen, L., and Kritzinger, E. (2016). Enterprise digitization enablement through unified communication & collaboration. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, pages 1–10.

Bolton, R. N. (1993). Pretesting questionnaires: content analyses of respondents' concurrent verbal protocols. *Marketing science*, 12(3):280–303.

Bongku, J. J. A. and Kurniawan, Y. (2019). Designing the knowledge managementsystem: A case study approach in it consultant company. In *Proceedings of the 2019 3rd International Conference on Graphics and Signal Processing*, pages 48–52.

Brundrett, M. and Lungka, P. (2019). The development of teachers' knowledgeand behaviour in promoting self-discipline: a study of early years teachers in thailand. *Education 3-13*, 47(4):462–474.

Bryman, A. and Bell, E. (2011). Ethics in business research. *Business Research Methods*, 7(5):23–56.

Carrion, G. C., Gonzalez, J. L. G., and Leal, A. (2004). Identifying key knowledgearea in the professional services industry: a case study. *Journal of Knowledge Management*.

CHAP (2020a). Chap community health accreditation partner. https://chapinc. org/#1520607340631-6318b110-2f8c. (Accessed on05/07/2020).

CHAP (2020b). Why use chap for your heathcare accreditation. https://chapinc. org/accreditation-process/. (Accessed on 05/07/2020).

Chatti, M. A., Klamma, R., Jarke, M., and Naeve, A. (2007). The web 2.0 drivenseci model based learning process. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, pages 780–782. IEEE.

CHC (2020). Interdisciplinary model, center for hospice care southeast connecticut. https://www.hospicesect.org/medical-professional/interdisciplinary-model. (Accessed on 05/09/2020).

Chen, M.-J. and Miller, D. (2012). Competitive dynamics: Themes, trends,and a prospective research platform. *The Academy of Management Annals*, 6(1):135–210.

Chugh, M., Chugh, N., Punia, A., and Agarwal, D. (2013). The role of informationtechnology in knowledge management. In *Proceedings of the conference on advances in communication and control systems-2013*. Atlantis Press.

Compassus (2020). What is hospice chap accreditation? https://www.compassus. com/about-us/what-is-chap-accreditation. (Accessed on 05/07/2020).
Courtney, K. L., Demiris, G., & Alexander, G. L. (2005). Information technology: changing nursing processes at the point-of-care. Nursing Administration Quarterly, 29(4), 315-322.

Curado, C. and Bontis, N. (2011). Parallels in knowledge cycles. *Computers in Human Behavior*, 27(4):1438–1444.

Curado, C., Oliveira, M., and Antunes, M. (2019). Organizational ambidexterityand customer relationship management: A cycle of virtue. *Knowledge and Process Management*, 26(3):229–243.Denzin, N. K., Lincoln, Y. S., et al. (2005). 2005. *Handbook of qualitative research*, 3.

Dierkes, M., Antal, A. B., Child, J., and Nonaka, I. (2003). *Handbook of organizational learning and knowledge*. Oxford University Press, USA.

Easa, N. F. (2012). Knowledge management and the seci model: a study of innovation in the egyptian banking sector.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy ofmanagement review*, 14(4):532–550.

Ellington, L., Reblin, M., Clayton, M. F., Berry, P., and Mooney, K. (2012). Hospice nurse communication with patients with cancer and their family caregivers. *Journal of palliative medicine*, 15(3):262–268.

Elliott, R. and Timulak, L. (2005). Descriptive and interpretive approaches to qualitative research. *A handbook of research methods for clinical and health psychology*, 1(7):147–159.

Ellram, L. M. (1996). The use of the case study method in logistics research.

Feng, Q. and Shanthikumar, J. G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, 27(9):1670–1684.

Fernando, J. T., Popadiuk, S., and Santos, N. M. B. F. (2019). Knowledge management at unilever south america–enabling conditions and their interaction with the seci model. *Revista de Neg´ocios*, 24(2):7–26.

Fisher, G. and Aguinis, H. (2017). Using theory elaboration to make theoretical advancements. *Organizational Research Methods*, 20(3):438–464.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal, 54(2), 393-420.

Gonçalves, T., Curado, C. and Balle, A. (2019). Psychosocial conditions for knowledge sharing in healthcare research centers: a qualitative comparative analysis, in Tomé, Cesário and Soares (Editors), Proceedings of the 20th European Conference on Knowledge Management, ISBN 978-1-912764-32-7, 415-424, ECKM, Lisbon, Portugal.

Group, C. (2003). Section one. https://www.vnhcsb.org/media/Hospice/ 20updated/202012-03-01.pdf. (Accessed on 05/07/2020).

Hsieh, H.-F. and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288.

Hughes, A., Park, A., Kietzmann, J., and Archer-Brown, C. (2019). Beyond bitcoin: What blockchain and distributed ledger technologies mean for firms. *Business Horizons*, 62(3):273–281.

Jackson, J. (2000). The urgent call for knowledge management in medicine. *Physician executive*, 26(1):28.

Johnson, J. S. (2015). Qualitative sales research: An exposition of grounded theory.
Journal of business logistics, 17(2):93.
Journal of Personal Selling & Sales Management, 35(3):262–273.

K˝o, A., Vas, R., Kova´cs, T., and Szabo´, I. (2019). Knowledge creation fromthe perspective of the supply chain. the role of ict. *Society and Economy*, 41(3):311–329.

Ketokivi, M. and Choi, T. (2014). Renaissance of case research as a scientificmethod. *Journal of Operations Management*, 32(5):232–240.

Konstantinidis, E. I., Antoniou, P. E., Bamparopoulos, G., and Bamidis, P. D. (2015). A lightweight framework for transparent cross platform communication of controller data in ambient assisted living environments. *Information Sciences*, 300:124–139.

Lamberg, J.-A., Tikkanen, H., Nokelainen, T., and Suur-Inkeroinen, H. (2009). Competitive dynamics, strategic consistency, and organizational survival. *Strategic*

Management Journal, 30(1):45–60.

Laudon, K. C. and Laudon, J. P. (2004). Managing the digital firm.

Laurenza, E., Quintano, M., Schiavone, F., and Vrontis, D. (2018). The effect ofdigital technologies adoption in healthcare industry: a case based analysis. *Business process management journal*.

Lenz, R., Peleg, M., and Reichert, M. (2012). Healthcare process support: achievements, challenges, current research. *International Journal of Knowledge-Based Organizations (IJKBO)*, 2(4).

Levinthal, D. and Rerup, C. (2006). Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organization science*, 17(4):502–513.

Lewis, M. W. (1998). Iterative triangulation: a theory development process usingexisting case studies. *Journal of Operations Management*, 16(4):455–469.

Li, H. (2018). Leadership succession and the performance of nonprofit organizations: A fuzzy-set qualitative comparative analysis. Nonprofit Management and Leadership, 29(3), 1-21.

Mårtensson, M. (2000). A critical review of knowledge management as a management tool. *Journal of knowledge management*.

McGinnis, M. A. and Vallopra, R. M. (1999). Purchasing and supplier involvement in process improvement: a source of competitive advantage. *Journal of Supply Chain Management*, 35(3):42–50.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemiamedica: Biochemia medica*, 22(3):276–282.

McKnight, B. and Bontis, N. (2002). E-improvisation: collaborative groupware technology expands the reach and effectiveness of organizational improvisation. *Knowledge and process management*, 9(4):219–227.

Merton, R. K. (1957). Social theory and social structure, rev.

Mitsuhashi, H. and Alcantara, L. L. (2019). Off the rivals' radar in emerging market segments: Non-mutual rival recognition between new firms and incumbents. *Long Range Planning*, page 101888.

NHL (2020). National hospice locator — hospice analytics. http://www. nationalhospicelocator.com/. (Accessed on 05/09/2020).

NHPCO (2018). Hospice facts & figures nhpco. https://www.nhpco.org/ research/. (Accessed on 05/08/2020).

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, 5(1):14–37.

Nonaka, I. and Konno, N. (1998). The concept of "ba": Building a foundation forknowledge creation. *California management review*, 40(3):40–54.

Nonaka, I. and Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford university press.

Nonaka, I., Toyama, R., and Byosi`ere, P. (2003). Handbook of organizational learning and knowledge, cap´ıtulo a theory of organizational knowledge creation: Understanding the dynamic process of creating knowledge. *Dierkes et al*, pages 491–517.

Nonaka, I., Toyama, R., and Konno, N. (2000). Seci, ba and leadership: a unified model of dynamic knowledge creation. *Long range planning*, 33(1):5–34.

O'Leary, D. E. (1998). Knowledge-management systems: Converting and connecting. *IEEE Intelligent Systems and Their Applications*, 13(3):30–33.

Obeidat, A. (2019). It adaption with knowledge conversion process (seci)? *Management Science Letters*, 9(13):2241–2252.

Oliveira, M., Ma¸cada, A. C., and Curado, C. (2014). Knowledge managementmechanisms within the seci model. In *Atas da Confer^encia da Associa¸c˜ao Portuguesa de Sistemas de Informac¸˜ao*, volume 11.

ontis, N. and Serenko, A. (2007). The moderating role of human capital management practices on employee capabilities. *Journal of knowledge management*.

Parker Oliver, D. R., & Demiris, G. (2004). An assessment of the readiness of hospice organizations to accept technological innovation. Journal of Telemedicine and Telecare, 10(3), 170-174.

Peckham, J. (1999). Industry and practice: Data for the masses. *Journal of Database Management (JDM)*, 10(2):35–37.

Porter, M. E. et al. (2010). What is value in health care. *N Engl J Med*,363(26):2477–2481.

Porter, M. E., Millar, V. E., et al. (1985). How information gives you competitive advantage.Rezaeibagha, F. (2013). Privacy and data security of electronic patient records (epr) sharing: Case studies: Iran and sweden.

Ragin, C. C. (2009). Qualitative Comparative Analysis using Fuzzy Sets (fsQCA). Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques, 51, 87–121.

Resnick, H. E., & Alwan, M. (2010). Use of health information technology in home health and hospice agencies: United States, 2007. Journal of the American Medical Informatics Association, 17(4), 389-395.

Rihoux, B., and Ragin, C. C. (2008). Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques (Vol. 51). Sage Publications.

Schniederjans, D. G., Curado, C., and Khalajhedayati, M. (2019). Supply chain digitisation trends: An integration of knowledge management. *International Journal of Production Economics*, page 107439.

Sekaran, U. and Bougie, R. (2010). Research for business–a skill building approach.

Senanayake, S., Senanayake, B., Ranasinghe, T., and Hewageegana, N. S. (2017). How

to strengthen primary health care services in sri lanka to meet the future challenges. *Journal of the College of Community Physicians of Sri Lanka*, 23(1).

Shah, R. and Chircu, A. (2018). Iot and ai in healthcare: A systematic literaturereview. *Issues in Information Systems*, 19(3).

Smith, K. G., Ferrier, W. J., and Ndofor, H. (2001). Competitive dynamics research: Critique and future directions. *Handbook of strategic management*, 315:361.

Stuermer, M., Abu-Tayeh, G., and Myrach, T. (2017). Digital sustainability: basic conditions for sustainable digital artifacts and their ecosystems. *Sustainability science*, 12(2):247–262.

Sun, J., Guo, Y., Wang, X., and Zeng, Q. (2016). mhealth for aging china: opportunities and challenges. *Aging and disease*, 7(1):53.

Tapscott, D. and Tapscott, A. (2016). *Blockchain revolution: how the technology behind bitcoin is changing money, business, and the world*. Penguin.

Tate, W. L. and Ellram, L. M. (2012). Service supply management structure in offshore outsourcing. *Journal of Supply Chain Management*, 48(4):8–29.

Tsui, E., Fong, P. S., and Choi, S. K. (2009). The processes of knowledge management in professional services firms in the construction industry: a critical assessment of both theory and practice. *Journal of Knowledge management*.

Udenio, M., Hoberg, K., and Fransoo, J. C. (2018). Inventory agility upon demand shocks: Empirical evidence from the financial crisis. *Journal of Operations Management*, 62:16 – 43.

Underwood, S. (2016). Blockchain beyond bitcoin.

VITAS (2020). Hospice vs. palliative care: What's the difference? https://www.vitas.com/hospice-and-palliative-care-basics/about-palliative-care/ hospice-vs-palliative-care-whats-the-difference/. (Accessed on 05/08/2020).

Vitas (2020). What is the difference between home health & hospice care?https://www.vitas.com/hospice-and-palliative-care-basics/ about-hospice-care/home-healthcare-or-hospice-care/. (Accessed on 05/08/2020).

Vongsingthong, S. and Smanchat, S. (2014). Internet of things: A review ofapplications & technologies. *Suranaree Journal of Science and Technology*, 21(4):359–374.

Voss, C., Tsikriktsis, N., and Frohlich, M. (2002). Case research in operations management. *International journal of operations & production management*.

Wallace, D. P. (2007). Knowledge management-historical and cross-disciplinary themes.

Wang, D., Lo, D., Bhimani, J., and Sugiura, K. (2015). Anycontrol–iot based homeappliances monitoring and controlling. In *2015 IEEE 39th Annual Computer Software and Applications Conference*, volume 3, pages 487–492. IEEE.

Wang, Z. and Kwek, C. L. (2018). The Mediation Role of Knowledge Sharing Between Organizational Learning and Technological Innovation Practice. International Journal of Knowledge Management, 14(3), 21.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy ofmanagement review*, 14(4):490–495.

Wilkesmann, M. and Wilkesmann, U. (2018). Industry 4.0–organizing routinesor innovations? *VINE Journal of Information and Knowledge Management Systems*.

Yin, R. (2003). Case study research: Design and methods 3 rd edition sage.*Thousand Oaks*.

Yin, R. K. (2009). Case study research: Design and methods fourth edition. *Los Angeles and London: SAGE*.

DECISION SCIENCES INSTITUTE

The Correlation Between State Health Department's Tweets and Public Sentiment Toward COVID-19

Zhigang Li
Kennesaw State University
Email: zli8@kennesaw.edu

Xin Tian
Kennesaw State University
Email: xtian2@kennesaw.edu

ABSTRACT

In light of the emergence and spread of the COVID-19 across the world, keeping the public informed with current and accurate information is essential. Given the ubiquitous presence and extensive usage of social media, many government agencies have been using social media platforms to communicate with the public and dispute the misinformation regarding the pandemic. This study aims to examine the relationship between the U.S. state health organization's public Twitter account and the public sentiment toward the COVID-19 pandemic. Tweets from the official accounts of 6 U.S. state health departments and the public responses were collected and analyzed.

KEYWORDS: COVID-19, Twitter, Social Media, Public Sentiment, Data Mining

INTRODUCTION

Since the emergence of the coronavirus disease (COVID-19) in December 2019, it has widely spread around the globe and has become one of the most challenging global crises (Sesagiri Raamkumar et al., 2020). Despite being listed as a Public Health Emergency of International Concern by the World Health Organization (WHO) on January 30, 2020 (World Health Organization, 2020b), the total number of confirmed cases of COVID-19 continues to grow. It is currently approaching six and a half million globally as of June 5, 2020 (World Health Organization, 2020a).

Given the rising popularity of social media platforms such as Twitter and Facebook over the past decade, researchers worldwide have taken the opportunity to study the various aspects of the spread of epidemic diseases through social media. Before the appearance of COVID-19, researchers have conducted various types of studies using social media on infectious diseases such as Ebola (Fung et al., 2016; Hossain et al., 2016; Odlum & Yoon, 2015; Roy et al., 2020), H1N1 (swine flu) (Biswas, 2013; Chew & Eysenbach, 2010; Ding & Zhang, 2010; B. F. Liu & Kim, 2011), Zika (Laurent-Simpson & Lo, 2019; Seltzer et al., 2017; Wirz et al., 2018), H7N9 (avian flu) and the Middle East Respiratory Syndrome (MERS) (Fung et al., 2013).

As one of the leading social media platforms, Twitter has over 60 million users in the U.S. as of April 2020 (Kemp, 2020). Most, if not all, of the U.S. federal and state government agencies, have public-facing Twitter accounts for community outreach and engagement. According to the Digital 2020 April Global Statshot Report (Kemp, 2020), the usage of social media has increased 36% in the U.S. during the current COVID-19 pandemic, likely due to the shelf-in-place orders and business shutdowns placed by state and local governments. This highlights

the importance of social media as a communication platform and provides public health organizations with an excellent opportunity to communicate with the public on how to respond to the epidemic appropriately and effectively. One of the most difficult challenges during this global crisis is the disinformation, misinformation, and fake news in social media. As a result, relevant government agencies and organizations need to stay in close contact with the public so that citizens can receive accurate information and proper guidance in the fight against the coronavirus disease.

This study aims to investigate and explore the relationship between the state public health department's Twitter feed and the public sentiment toward COVID-19 cross the states of New York, California, Texas, Florida, Ohio, and Kansas. Tweets published by both the state health departments' twitter accounts and public responses from February 1, 2020 – May 31, 2020, are collected. The sentiment scores from the tweets, along with their correlation with critical factors, are analyzed and discussed.

LITERATURE REVIEW

Prior Research on Social Media and Infectious Diseases

Much work has been done in the past ten years to study social media and the outbreaks of infectious diseases. Among them, a team of researchers (Tang et al., 2018) conducted a systematic literature review to identify the research approaches of published research articles on infectious diseases and social media. They collected 30 relevant research papers ranging from January 1, 2010, to March 1, 2016, and identified three significant types of research (Tang et al., 2018):

1. Assessment of public interests and response
2. Organizations' use of social media and public response
3. Accuracy of media information

Similar to the work done by Tang and colleagues (2018), another research team (Wang et al., 2019) conducted a literature review focusing only on the spread of health-related misinformation on social media. In their study, the research team collected 57 related articles and analyzed their research topics, theoretical frameworks, and research design. Among those articles, over 50% of them are studies on infectious diseases and social media (Wang et al., 2019). In addition, the types of research spread across a broad spectrum, including areas such as epidemiological research, public health informatics, health policies, communications, vaccines, and cyberpsychology (Wang et al., 2019).

Recent Research on Social Media and COVID-19

Businesses use data analysis and big data mining to extract insightful information from the collected massive data (He et al., 2015, 2016; Tian et al., 2016). Although there have only been just a few months since the beginning of the COVID-19 global pandemic, a steady stream of research papers has already been published on the topic of social media and COVID-19, particularly in the area of medical science.

A team of researchers led by Li (C. Li et al., 2020) was among one of the earliest to publish a research article on the analysis and prediction of COVID-19 outbreak through Internet searches and social media data within China. Their results show a strong correlation between the daily

confirmed case number and Internet search data such as Google Trends, Baidu Index (a Chinese search engine), and Sina Weibo (a leading Chinese social media platform) Index.

A team of healthcare researchers lead by Li (J. Li et al., 2020) conducted a content analysis on one of the Chinese social media platform, Sina Weibo, using data gathered from December 23, 2019, to January 30, 2020. They have found a positive correlation between the daily reported cases and the number of Weibo posts. Their content analysis revealed varying levels of individual uncertainty and a shift of public perception about the origins and transmission patterns throughout the study. Similarly, Han, Wang, Zhang, and Wang (Han et al., 2020) conducted a study to analyze public opinion toward COVID-19 in China using data gathered from Sina Weibo from January 9, 2020, to February 10, 2020. Their results show that the public shared their opinions and emotions during the early stage of COVID-19, and the timely release of authoritative information by the government is effective in notifying the public and popularizing the methods of prevention and treatment of the disease.

In addition to the research studies coming out of China in the early stages of COVID-19, Abd-Alrazaq and colleagues (Abd-Alrazaq et al., 2020) conducted a study analyzing the top concerns of Tweeters during the COVID-19 pandemic. Park, Park, and Chong (2020) studied the public concerns and opinions of the Korea people regarding COVID-19 using data gathered from Twitter.

The Gap in Existing Research

Although there have already been some research studies published regarding COVID-19 and the use of social media as mentioned in the previous section, few have been done in the United States. This can be largely attributed to the time difference in the spread of COVID-19 compared to Asian and European countries. Also, the political structure in the United States allows each state to act more independently in their response to COVID-19, which presents us with a good opportunity to look at the differences among different states in their effort in the prevention and control of the disease. Thus, our study focuses on six states, New York, California, Texas, Florida, Ohio, and Kansas, chosen primarily based on their geographical location among other factors to look at how their public health departments use Twitter to communicate with the public during COVID-19.

MODEL and HYPOTHESES

Social Media Post Life Cycle

In this study, we proposed a social media post life cycle by the government agencies in Figure 1. Government agencies such as the Public Health Department post information on their social media platform and wait for a few days to interact with the public. The agencies can then extract social media data and then analyze the social media data with the official confirmed cases with infectious diseases. To make the result easy to read, the agencies can use data visualization to find the pattern and peak from the data. The decision-maker can then utilize the visualized results and make a new strategic plan and decisions if necessary. Based on another round of risk evaluation and management, the agencies can post another message on the social medial platform. Hence, we have the following hypothesis regarding the proposed model and data analysis.

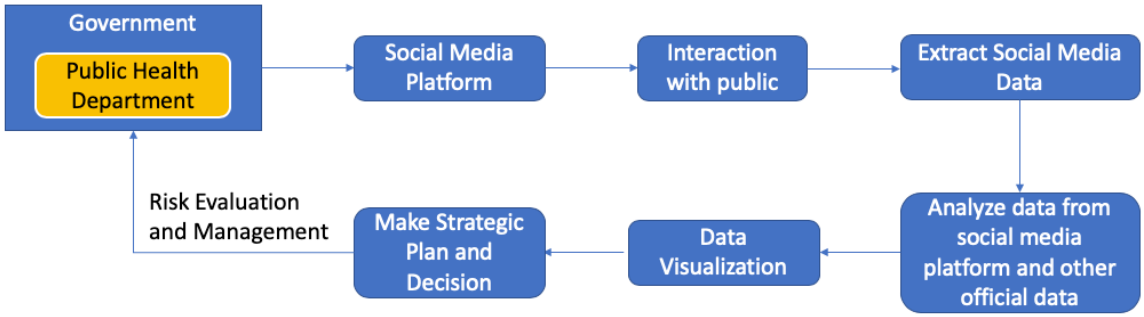


Figure 1: Social Media Posts Life Cycle by the Government Agencies

Hypotheses

Our proposed hypotheses for this study are:

Hypothesis 1: The negative sentiment of social media data aligns with the confirmed COVID-19 cases.

Hypothesis 2: The more social media public comments, the more negative of sentiment.

METHODOLOGY

Data Collection

In order to explore the relationship between social media data and the COVID-19 confirmed cases, we collected tweets from 6 U.S. state health departments' (the exact name could vary depending on the state) official twitter accounts along with the public's responses to those tweets from February 1 - May 31, 2020. Due to its large number of cases and geographical location on the east coast, we have chosen the state of New York. California is among one of the earliest states with confirmed cases and is located on the west coast. Texas is the first state to ease COVID-19 restrictions and reopen its economy. Florida was chosen both for its geographical location and the large population of elderly residents. Ohio was chosen due to its early response to COVID-19 when the state still had a relatively low number of cases. Lastly, Kansas was chosen for its geographical location in the middle of the continental U.S. The descriptive data in Table 1 shows the official Twitter account, number of tweets from the official account, the volume of tweets, number of unique users, and number of followers. The date ranges from February 1 to May 31, 2020. Florida Department of Health has the most followers and the most volume of tweets during this period, followed by Ohio, California, Texas, New York, and Kansas.

In addition to the data collected from Twitter, we have also acquired the case data for COVID-19 from the Humanitarian Data Exchange (HDX). HDX is an open-source data platform sponsored by a non-profit organization. We collected the COVID-19 cases data for each of the states mentioned above as of June 6, 2020 (HDX, 2020).

Table 1: The Descriptive Twitter Data from
the State Health Department from Feb.1 to May 31, 2020

State	Twitter Account	Num. of Tweets	Volume of Tweets	Num. of Unique Users	Num. of Followers
California	@CAPublicHealth	560	20,876	9,258	55,000
Florida	@HealthyFla	588	40,597	15,312	84,000
Kansas	@KDHE	470	5,261	1,845	14,400
New York	@HealthNYGov	230	13,025	6,935	18,100
Ohio	@OHdeptofhealth	960	35,077	14,538	71,100
Texas	@TexasDSHS	260	14,837	6,384	25,600

Table 2 shows the COVID-19 cases rate per 100,000 population in the six states in the U.S. New York state only had a total number of 13,025 tweets, and the followers are on the very lower end despite the highest confirmed cases and death. According to the U.S. Centers for Disease Control and Prevention (CDC) website, Texas, Ohio and Kansas are considered COVID-19 community transmission in defined areas, the other three states are considered widespread community transmission (CDC, 2020).

Table 2: COVID-19 Cases Per 100,000 Population as of June 6, 2020

State	COVID-19 cases rate per 100,000 population
California	310.7
Florida	281.7
Kansas	357
New York	1528
Ohio	323
Texas	249.5

Sentiment Analysis

Opinion mining and sentiment analysis contribute to the development of opinion-oriented information systems for service or products providing companies (Pang & Lee, 2008). If service providing companies want to get feedback or opinions from the public, the traditional way is for them to conduct surveys and to distribute the surveys to the public who have been impacted or can be potentially impacted. However, a survey is not enough to get the customers' feedback in this digital world because a survey needs time to be collected and analyzed. The issues of service cannot be addressed by the government agencies quickly. Fortunately, people like to post their opinions and experiences on social media platforms rather than completing a survey or complaining to the government agencies directly. Twitter and Facebook are two of the most popular social media platforms.

Twitter Sentiment Analysis is the process of determining the emotional tone behind a series of words. It is used to understand the attitudes, opinions, and emotions expressed in an online mention. Having a solid understanding of current public sentiment can be an excellent tool for any government agency. There are many ways to do sentiment analysis, and most of them follow the same general idea. Here are the three steps to do the sentiment analysis:

1. Create or find a list of words associated with strongly positive or negative sentiment.
2. Count the number of positive and negative words in the text.
3. Analyze the mix of positive to negative words. The use of many positive words and few negative words indicates a positive sentiment, whereas the use of many negative words and few positive words indicates negative sentiment.

We use positive and negative words to perform sentiment analysis and evaluate the tweets with those positive and negative lexicons. The list of positive words contains 2,003 words, and the list of negative words contains 4,782 words (Hu & Liu, 2004; B. Liu et al., 2005). We also added some misspelled words that make frequent appearances in social media content to the lists.

After having been run through the sentiment words lists, the tweets should be preprocessed before the sentiment scores can be calculated. First, the data must be cleaned. This step involves the removal of stop words, all of the numbers, and all of the white spaces, and the conversion of all of the words or letters to the lower case. And then, creating or finding a word list (also called a Lexicon). For example, social media popular acronyms or slangs are not included in the existing Lexicons. “omg”, “lol”, “thx”, and “wtf” are ubiquitous acronyms on social media platforms. We included these types of words in the Lexicons for this study. Researchers can edit the Lexicons when they study specific topics. However, some words have a double emotional meaning. For instance, “sick" is an example of a word that can have positive or negative sentiment depending on what it's referred to. The sentiment is probably negative if people are discussing a pet store that sells many sick animals. On the other hand, if people are talking about a skateboarding instructor who taught you how to do a lot of sick flips, it is probably very positive. Sentiment analysis uses machine learning algorithms. In this research, R Studio was used to load the sentiment analysis packages and to analyze the processed tweets.

Data Analysis

Our analysis for this observational study includes time series analysis, logistic regression analysis, and sentiment analysis using natural language processing to examine public knowledge and attitudes. We have collected two data sets to find the relationship between social media data and the confirmed COVID-19 cases and death. The first data set is social media data and its corresponding sentiment score for each state from February 1 to May 31, 2020. Moreover, and the second data set is the daily confirmed cases and death from open-source data platform HDX. However, due to each state's different situations, not all states reported the cases from February 1, 2020. This depends on when the first case was reported in each state. Florida reported its first case on March 1, 2020, while Ohio reported the first three cases on March 9, 2020. When we analyzed the data, we considered the date before the first reported date in each state had 0 cases on both confirmed cases and death. The logistic regression model will use the public sentiment as the dependent variable. The number of tweets, rewet, favorite, daily new COVID-19 cases, and daily new death will be the independent variables. We will test the following two models:

Model 1:

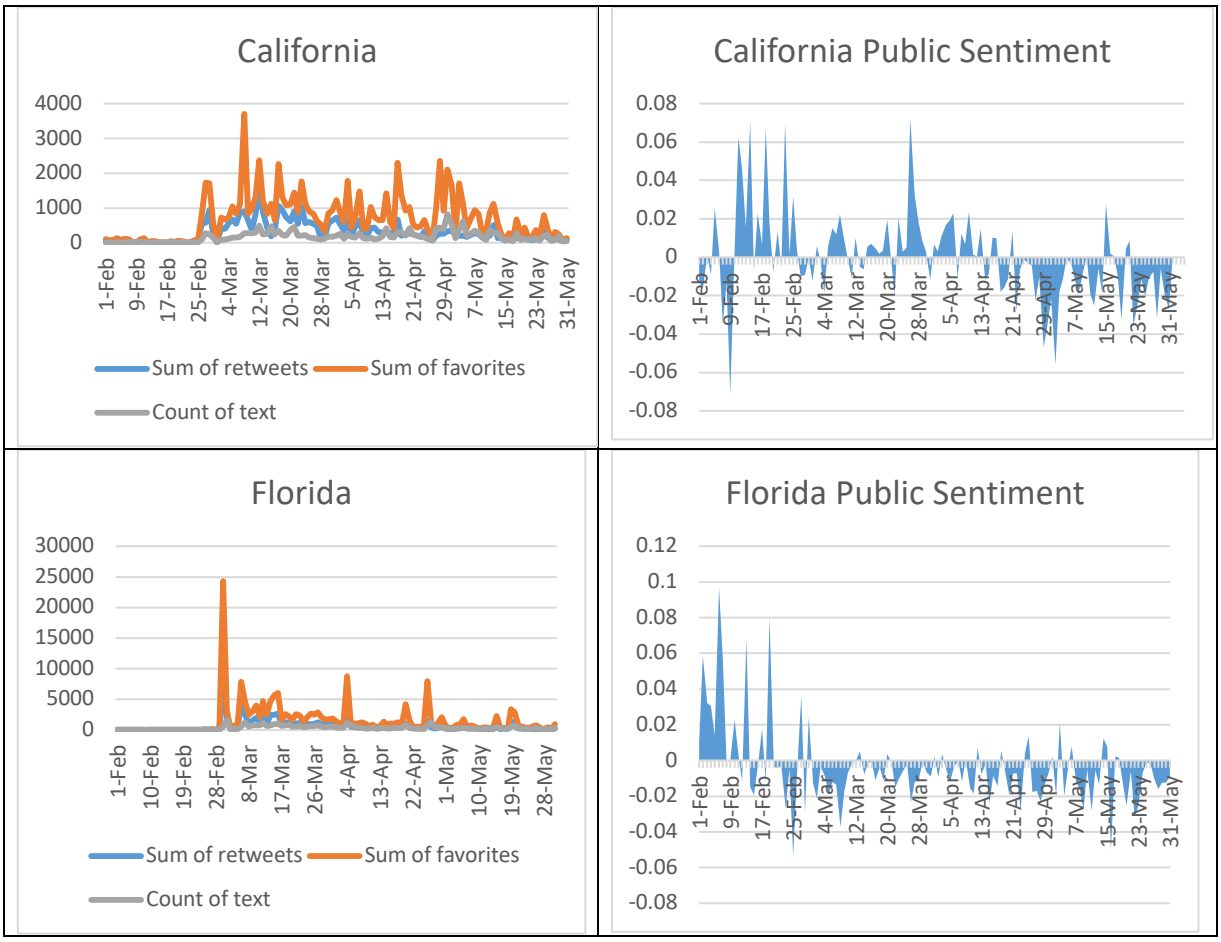
$$\boldsymbol{Public\ Sentiment} = \boldsymbol{\alpha} + \boldsymbol{\beta_1 tweets} + \boldsymbol{\beta_2 reweets} + \boldsymbol{\beta_3 favorite} + \boldsymbol{\beta_4\ NewCases} + \boldsymbol{\beta_5 NewDeath} + \boldsymbol{\varepsilon}$$

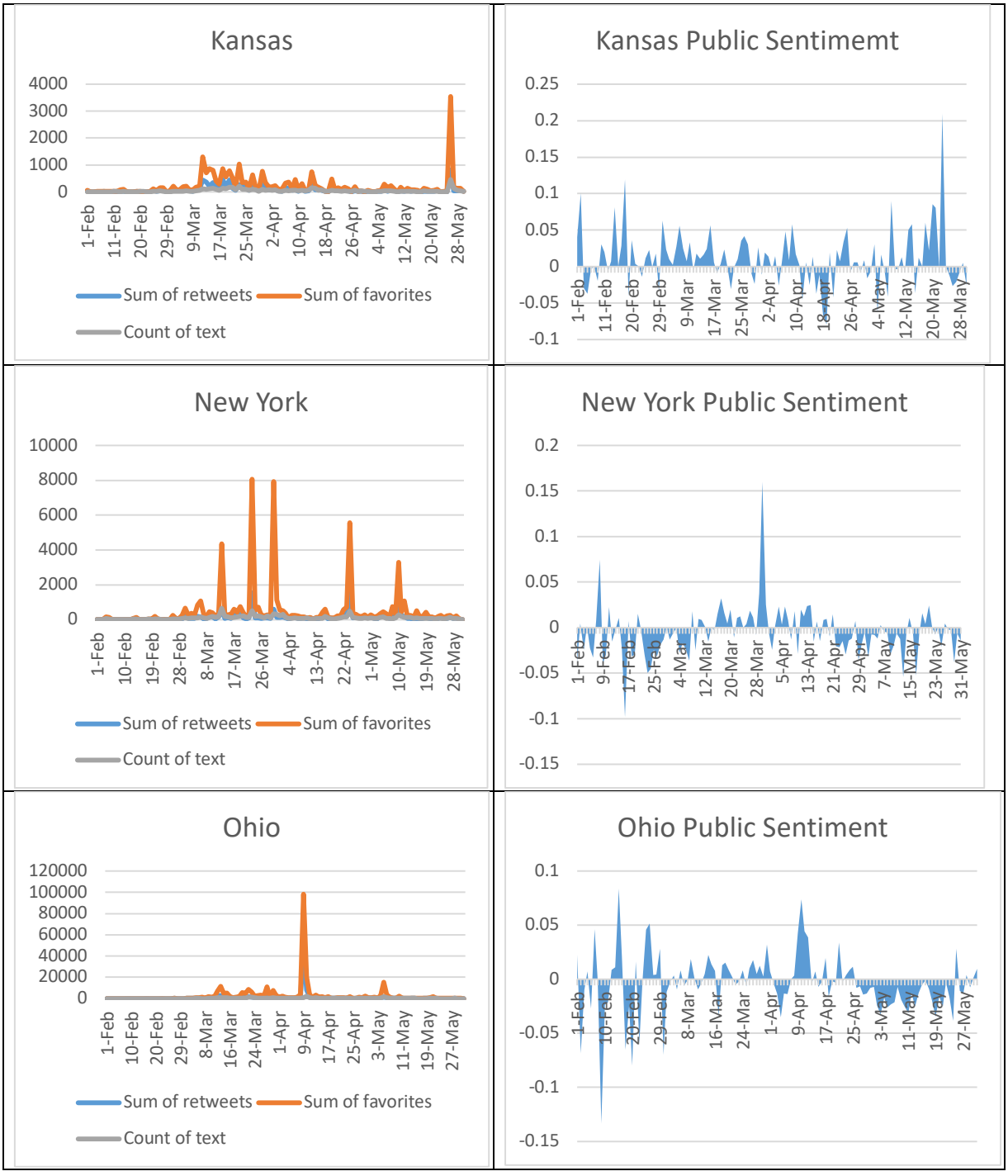
Model 2:

$$\boldsymbol{Public\ Sentiment} = \boldsymbol{\alpha} + \boldsymbol{\beta_1 tweets} + \boldsymbol{\beta_2 reweets} + \boldsymbol{\beta_3 favorite} + \boldsymbol{\beta_4\ NewCases} + \boldsymbol{\varepsilon}$$

RESULTS

We first analyzed the social media data and the daily public sentiment score for the selected six states in the U.S. The left column of Figure 2 shows the trend of the sum of retweets, the sum of favorites, and the count of tweets daily from February 1 to May 31, 2020. The right column shows the daily public sentiment for each state during this period.




Kansas
4000
3000
2000
1000
0
1-Feb
11-Feb
20-Feb
29-Feb
9-Mar
17-Mar
25-Mar
2-Apr
10-Apr
18-Apr
26-Apr
4-May
12-May
20-May
28-May
Sum of retweets
Sum of favorites
Count of text
Kansas Public Sentimemt
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
1-Feb
11-Feb
20-Feb
29-Feb
9-Mar
17-Mar
25-Mar
2-Apr
10-Apr
18-Apr
26-Apr
4-May
12-May
20-May
28-May
New York
10000
8000
6000
4000
2000
0
1-Feb
10-Feb
19-Feb
28-Feb
8-Mar
17-Mar
26-Mar
4-Apr
13-Apr
22-Apr
1-May
10-May
19-May
28-May
Sum of retweets
Sum of favorites
Count of text
New York Public Sentiment
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
1-Feb
9-Feb
17-Feb
25-Feb
4-Mar
12-Mar
20-Mar
28-Mar
5-Apr
13-Apr
21-Apr
29-Apr
7-May
15-May
23-May
31-May
Ohio
120000
100000
80000
60000
40000
20000
0
1-Feb
10-Feb
20-Feb
29-Feb
8-Mar
16-Mar
24-Mar
1-Apr
9-Apr
17-Apr
25-Apr
3-May
11-May
19-May
27-May
Sum of retweets
Sum of favorites
Count of text
Ohio Public Sentiment
0.1
0.05
0
-0.05
-0.1
-0.15
1-Feb
10-Feb
20-Feb
29-Feb
8-Mar
16-Mar
24-Mar
1-Apr
9-Apr
17-Apr
25-Apr
3-May
11-May
19-May
27-May

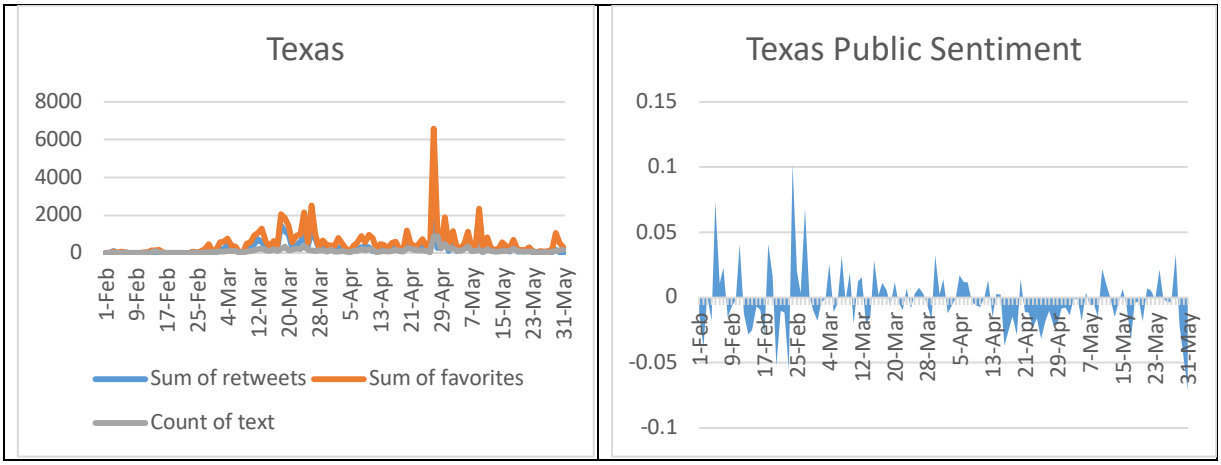


Figure 2: Social Media Data and Public Sentiment

From Figure 2, we can see that Kansas shows the most positive and stable public sentiment during this period. Since Kansas has lower confirmed COVID-19 cases and people didn't post many tweets on Twitter, the public sentiment looks favorite to the government agencies. The state of Ohio has similar confirmed cases by population as Kansas, but Ohio's public sentiment is more negative. New York and Texas have mixed public sentiment during this period. We will investigate the relationship between these social media factors and confirmed COVID-19 cases and death. New York also had several surges of tweets during this period. The state of Florida has more negative public sentiment after Feb. 28, 2020, than before. The reason is that the beginning of the COVID-19 outbreak was at the end of February in Florida. And we can see the tremendous number of tweets from February 28 to the end of May.

Table 3 presents the results of the two regression models. As we can see, the second model (R^2 = *.989*) fits better than the first model (R^2 = *.169*) with both daily new cases and death as independent variables. We found that the number of retweet and favorite do not affect the public sentiment at all. The volume of tweets and the number of daily new cases play an essential role in public sentiment. This can be interpreted as that the public sentiment will decrease 72.46 when the number of tweets increases 1. So, hypothesis 2 was supported. In model 2, the daily new cases will positively affect public sentiment. The daily new cases have a coefficient of 12.729 at a significant level of 0.1%. That indicates that the daily new cases align with the public sentiment. Hence, hypothesis 1 was supported, as well.

Table 3: Regression Models Results

Dependent Variable	**Public Sentiment (1)**	**Public Sentiment (2)**
(Constant)	.010**	7720.851***
	(.004)	(656.750)
Sum of Retweets	6.716E-6	-1.706
	(.000)	(1.248)
Sum of Favorites	3.775E-6	.716
	(.000)	(.582)
Volume of Tweets	-5.146E-5	-72.460***
	(.000)	(5.659)

Daily New Cases	-6.419E-6	12.729***
	(.000)	(.080)
Daily Death	1.118E-5	
	(.000)	
R-squared	.169	.989
Adjust R-squared	.132	.978

Coefficient Standard Error (in parentheses)
**. Significance at the 1% level
***. Significance at the 0.1% level

DISCUSSION AND CONCLUSIONS

Prior research pointed out the challenges and opportunities engendered by opinion-rich resources, such as online comments, blogs, social media platforms, and online forums (Pang & Lee, 2008). By investigating the value of those unstructured data, government agencies are able to gather public relationship data and improve their services and interaction with the public in a short time. In order to actively use information technology to seek and retrieve the data and to understand consumers' opinions, computational treatment of subjective texts like reviews, comments, and opinions has emerged in the past ten years (Pang & Lee, 2008). The proposed life cycle in Figure 1 can provide data visualization and immediate responses to negative comments and can reduce the risk of public crisis. Government agencies and policymakers can use an opinion-oriented information system to make decisions and to improve their services.

After comparing the six states in the U.S., we were surprised to see that the New York State Department of Health has fewer social media followers than the other states with a lower population. Although New York state has the highest number of confirmed COVID-19 cases as of date, fewer people follow the social media account of the Department of Health. There are a few possible explanations for this phenomenon. First, compared to the health department's website of the other states, including the state of Florida, Kansas, Texas, and California, the New York Department of Health website does not have a dedicated section for their social media accounts on their homepage. Besides, the links to their social media accounts are located at the very bottom of the homepage, which can be easily missed. Second, the New York State Department of Health publishes less information on their Twitter account regarding COVID-19, compared to the other states. An interesting comparison is with Ohio, which also does not have a dedicated social media section on the homepage of its website, similar to New York. However, the Ohio Department of Health publishes four times the number of tweets than that of New York. Ohio also has the second-highest number of followers and retweets, which suggests that the Ohio Department of Health has more interaction with the public regarding COVID-19. The sharp contrast between New York and Ohio in their Twitter activities, number of confirmed cases, as well as the designation by CDC as COVID-19 community transmission in defined areas for Ohio vs. widespread community transmission for New York (CDC, 2020), made us believe that Twitter, and in large part social media, played an essential role in the communication, prevention, and control of the COVID-19 disease in the state of Ohio.

As shown from the result of the second regression model, public sentiment has a direct correlation with the number of daily new cases and the volume of tweets. As suggested by plenty of existing research, social media can be an effective communication channel and platform for government agencies to communicate with the public and gather public opinions and sentimental feedback (M. Graham, 2013; M. W. Graham et al., 2015; Hofmann et al., 2013;

Mergel, 2013). Particularly in the case of against infectious diseases, the study and analysis of data from social media can provide valuable insights into a wide variety of aspects such as epidemiological study, public health informatics, public health policies, and communications (Wang et al., 2019). Our proposed social media posts life cycle model suggests government health agencies gather and analyze public opinions and sentiment from social media to help them make data-driven decisions and strategic planning for public communication and outreach based on the feedback.

LIMITATIONS AND FUTURE WORK

This study is limited in its scope in two aspects: 1) the limitation of data being collected, 2) the limitation of data source, and its' demographic representation.

Limitation of Data Collected

As mentioned previously, we have chosen only six states for data collection and analysis. Since the size, geographical location, population, and social-economic development level vary significantly from state to state in the U.S., the results and findings from this study may not represent and be applied to the rest of the states or the U.S. as a whole. In addition, we have only collected data from February 1 to May 31, 2020, shortly after the first reported case in Washington state in late January 2020. Thus, the public's opinion and sentiment toward COVID-19 before February 1, 2020, were not studied in this paper.

Limitation of Data Source and Its Demographic Representation

For this study, we have only collected data from a single social media platform - Twitter. Despite being one of the most popular social media platforms in the U.S., Twitter has some intrinsic characteristics, such as the limitation of the number of characters per post, that limit the in-depth thought and opinion sharing for the users compared to other platforms such as Facebook. Also, considering the majority of the Twitter users in the U.S. live in cities and suburban areas and are age 34 or younger (Sehl, 2020), the results and findings of this study may not represent the entire population.

Future Work

Due to the limitation of this research, we will continue and expand the research to all 50 states in the U.S. and even more. Some other factors should be considered when we analyze the social media public sentiment. For example, the demographical differences such as education level, race, age, job, and more should be considered in future work. Furthermore, a more in-depth statistical analysis should be performed. For social media analysis, content analysis with the application of topic modeling, weights, and the relationship of each model should be investigated. By the means of the social media analysis and with the official data, government agencies should have the specialist to collect these data and report to the decision-makers. In our proposed model, risk evaluation and management are the next steps to maintain a healthy public relationship.

REFERENCES

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of Medical Internet Research*, *22*(4), e19016. https://doi.org/10.2196/19016

Biswas, M. (2013). Health Organizations' Use of Social Media Tools during a Pandemic Situation: An H1N1 Flu Context. *Journal of New Communications Research*, *5*, 46–81.

CDC. (2020). *Previous U.S. COVID-19 Case Data.* Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/previouscases.html

Chew, C., & Eysenbach, G. (2010). Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLoS ONE*, *5*(11), e14118. https://doi.org/10.1371/journal.pone.0014118

Ding, H., & Zhang, J. (2010). Social Media and Participatory Risk Communication during the H1N1 Flu Epidemic: A Comparative Study of the United States and China. *Social Media*, 14.

Fung, I. C.-H., Fu, K.-W., Chan, C.-H., Chan, B. S. B., Cheung, C.-N., Abraham, T., & Tse, Z. T. H. (2016). Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. *Public Health Reports*, *131*(3), 461–473. https://doi.org/10.1177/003335491613100312

Fung, I. C.-H., Fu, K.-W., Ying, Y., Schaible, B., Hao, Y., Chan, C.-H., & Tse, Z. T.-H. (2013). Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infectious Diseases of Poverty*, *2*(1), 31. https://doi.org/10.1186/2049-9957-2-31

Graham, M. (2013). Government Public Relations and Social Media: An Analysis of the Perceptions and Trends of Social Media Use at the Local Government Level. *Public Relations Journal*, *7*(4), 21.

Graham, M. W., Avery, E. J., & Park, S. (2015). The role of social media in local government crisis communications. *Public Relations Review*, *41*(3), 386–394. https://doi.org/10.1016/j.pubrev.2015.02.001

Han, X., Wang, J., Zhang, M., & Wang, X. (2020). Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *International Journal of Environmental Research and Public Health*, *17*(8), 2788. https://doi.org/10.3390/ijerph17082788

HDX. (2020, June 11). *COVID-19 Pandemic.* HDX. https://data.humdata.org/event/covid-19

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, *115*(9), 1622–1636.

He, W., Tian, X., Chen, Y., & Chong, D. (2016). Actionable Social Media Competitive Analytics for Understanding Customer Experiences. *Journal of Computer Information Systems*, *56*(2), 145–155.

Hofmann, S., Beverungen, D., Räckers, M., & Becker, J. (2013). What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook. *Government Information Quarterly*, *30*(4), 387–396. https://doi.org/10.1016/j.giq.2013.05.013

Hossain, L., Kam, D., Kong, F., Wigand, R. T., & Bossomaier, T. (2016). Social media in Ebola outbreak. *Epidemiology and Infection*, *144*(10), 2136–2143. https://doi.org/10.1017/S095026881600039X

Hu, M., & Liu, B. (2004, August 22). *Mining and Summarizing Customer Reviews*. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, WA.

Kemp, S. (2020, April 23). *Digital 2020: April Global Statshot*. Datareportal. https://datareportal.com/reports/digital-2020-april-global-statshot

Laurent-Simpson, A., & Lo, C. C. (2019). Risk society online: Zika virus, social media and distrust in the Centers for Disease Control and Prevention. *Sociology of Health & Illness*, *41*(7), 1270–1288. https://doi.org/10.1111/1467-9566.12924

Li, C., Chen, L. J., Chen, X., Zhang, M., Pang, C. P., & Chen, H. (2020). Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance*, *25*(10). https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000199

Li, J., Xu, Q., Cuomo, R., Purushothaman, V., & Mackey, T. (2020). Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. *JMIR Public Health and Surveillance*, *6*(2), e18700. https://doi.org/10.2196/18700

Liu, B. F., & Kim, S. (2011). How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for U.S. health communicators. *Public Relations Review*, *37*(3), 233–244. https://doi.org/10.1016/j.pubrev.2011.03.005

Liu, B., Hu, M., & Cheng, J. S. (2005, May 10). *Opinion Observer: Analyzing and Comparing Opinions on the Web*. Chiba, Japan.

Mergel, I. (2013). Social media adoption and resulting tactics in the U.S. federal government. *Government Information Quarterly*, *30*(2), 123–130. https://doi.org/10.1016/j.giq.2012.12.004

Odlum, M., & Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? *American Journal of Infection Control*, *43*(6), 563–571. https://doi.org/10.1016/j.ajic.2015.02.023

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2(1-2)*, 1–135.

Park, H. W., Park, S., & Chong, M. (2020). Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. *Journal of Medical Internet Research*, *22*(5), e18897. https://doi.org/10.2196/18897

Roy, M., Moreau, N., Rousseau, C., Mercier, A., Wilson, A., & Atlani-Duault, L. (2020). Ebola and Localized Blame on Social Media: Analysis of Twitter and Facebook Conversations During the 2014–2015 Ebola Epidemic. *Culture, Medicine, and Psychiatry*, *44*(1), 56–79. https://doi.org/10.1007/s11013-019-09635-8

Sehl, K. (2020, May 28). *Top Twitter Demographics That Matter to Social Media Marketers*. Hootsuite. https://blog.hootsuite.com/twitter-demographics/

Seltzer, E. K., Horst-Martz, E., Lu, M., & Merchant, R. M. (2017). Public sentiment and discourse about Zika virus on Instagram. *Public Health*, *150*, 170–175. https://doi.org/10.1016/j.puhe.2017.07.015

Sesagiri Raamkumar, A., Tan, S. G., & Wee, H. L. (2020). Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison. *Journal of Medical Internet Research*, *22*(5), e19334. https://doi.org/10.2196/19334

Tang, L., Bie, B., Park, S.-E., & Zhi, D. (2018). Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *American Journal of Infection Control*, *46*(9), 962–972. https://doi.org/10.1016/j.ajic.2018.02.010

Tian, X., He, W., Tao, R., & Akula, V. (2016). *Mining Online Hotel Reviews: A Case Study from Hotels in China*. AMCIS 2016.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science & Medicine*, *240*, 112552. https://doi.org/10.1016/j.socscimed.2019.112552

Wirz, C. D., Xenos, M. A., Brossard, D., Scheufele, D., Chung, J. H., & Massarani, L. (2018). Rethinking Social Amplification of Risk: Social Media and Zika in Three Languages. *Risk Analysis*, *38*(12), 2599–2624. https://doi.org/10.1111/risa.13228

World Health Organization. (2020a). *WHO Coronavirus Disease (COVID-19) Dashboard*. World Health Organization. https://covid19.who.int

World Health Organization. (2020b). *Novel Coronavirus(2019-nCoV) Situation Report—11*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4

DECISION SCIENCES INSTITUTE

Hard and Soft Mitigation Strategies for Managing Supply Chain Risks: A Case of a Multinational Firm in Nigeria

Adenike Aderonke Moradeyo
Pan-Atlantic University
Email: nmoradeyo@pau.edu.ng

Marvel Ogah
Lagos Business School
Email: mogah@lbs.edu.ng

ABSTRACT

The purpose of this study is to analyse the performance effectiveness of the use of both hard and soft mitigation strategies as hybrid mitigation strategies for managing supply chain risks. This work makes use of the case study approach to investigate the issues of supply chain risks, the hard and soft mitigation strategies in use and the performance effectiveness in a multinational firm. The study confirmed the use of hybrid (hard and soft) mitigation strategies by firms, probable synergistic effect and performance effectiveness. A single case study was utilized which could limit the generalisability of the research findings. This work serves as a useful source of information about the forms of hard and soft mitigation strategies used by manufacturing firms and the importance of use of hybrid mitigation strategies. This paper provides a novel categorisation of mitigation strategies as hard and soft mitigation strategies and performance effectiveness.

KEYWORDS: Performance Effectiveness, Supply Chain Risk, Mitigation Strategies, Hard and Soft, and Hybrid

INTRODUCTION

Mand, Singh and Singh (2013:105) described a “supply chain to consist of suppliers, manufacturing centres, warehouses, transporters, distribution centres and retail outlets, as well as raw materials, work-in-progress inventory and finished products that flow between the facilities”. Supply chain management integrates all activities that involve information flow, the transformation of raw materials and the eventual transportation of finished products to end consumers including parties such as manufacturers, distributors and retailers (Gunasekaran, Lai and Cheng, 2008). According to Blackhurst et al. (2005), an effective supply chain management is a challenging task which can be made more difficult when supply chains face unexpected disruptions. Tang (2006a:453) defined “supply chain risk management as the use of mitigation strategies in managing supply chain risks which involves coordination and collaboration among the supply chain partners to ensure profitability and continuity”. Therefore, managing risks in supply chains have emerged as an important topic in supply chain management.

Mitigation strategies such as strategic stock, postponement, dynamic assortment planning, revenue management, silent product rollover, multiple suppliers, flexible transportation and contingent sourcing are used by manufacturing firms (Tang, 2006b; Stecke and Kumar, 2009; Sodhi and Tang, 2012). These mitigation strategies sometimes require technical know-how

before they can be implemented. The category of such mitigation strategies is called hard mitigation strategies in this study. Furthermore, there are some forms of strategies that do not require technical know-how but are rather a relationship-based and could be useful for the mitigation of supply chain disruptions. These strategies are known as soft mitigation strategies. For example, a manufacturing firm that purchases equipment for a credible supplier on credit which will be paid for later is a form of supplier relationship management strategy. This form of strategy is based on the notion that “our suppliers need to be helped for them to help us”.

The aforementioned strategy helps in building the loyalty of such a supplier. Hence, the issue of unavailability of supplies could be solved to a certain degree based on the fact that such a supplier will give a manufacturing company that renders help to it priority over others. This supplier relationship management strategy could help in the mitigation of the issue of stock-out and subsequently impact performance positively. Consequently, this study emphasizes the need to elaborate on relationship-based mitigation strategies. Therefore, this study tends to answer the following research questions;

- What are the hard and soft mitigation strategies employed by manufacturing firms for managing supply chain risks?
- What is the performance effectiveness of the use of both hard and soft mitigation strategies for managing supply chain risks?

LITERATURE REVIEW

Supply chain disruptions are described as unplanned and unanticipated events that disrupt the normal flow of goods and materials within supply chains (Hendricks and Singhal, 2003; Svensson, 2000). Some examples of supply chain disruptions highlighted in extant literatures include; unanticipated or volatile demand, insufficient or distorted information from customers about orders or demand quantity or mix, seasonality, volatility of fads, new product adoptions, short product lifecycle, lack of forecasting accuracy and over-inventory, supplier business risks, production capacity constraints in the supply market, supply cost risks, supply commitment risk, quality problems, technological changes, transportation delays, and product design changes (Zsidisin et al., 2000; Bode and Wagner, 2006; Cucchiella and Gastaldi, 2006; Tang and Tomlin, 2008; Olson and Dash Wu, 2010; Juttner, 2005; Okon, 2011; Soderbom and Teal, 2002; Iwarere, 2010; Malik, Teal and Baptist, 2006; Mentzer, 2004; Larossi, Mousley and Radwan, 2009; Snyder et al., 2016; Sarkar and Kumar, 2016; Dabhilkar et al., 2016; Simangunsong et al., 2016; Zhao et al., 2013). Therefore, supply chain risk management has gained prominence among practitioners and academics because of the increase in the trend of disruptions in modern supply chains (Bode and Wagner, 2006; Bode and Wagner, 2008). Supply chain risk management is a management process that supports supply chain partners and involves the use of diverse mitigation strategies to control, curb and reduce the impact of supply chain disruptions (Ellegaard, 2008; Norrman and Lindroth, 2004). Hence, mitigation strategies could be simply described as strategies used for controlling, curbing and reducing the impacts of supply chain disruptions (Norrman and Lindroth, 2004).

There have been several categorisations of supply chain mitigation strategies in extant literature such as robust strategies, proactive strategies, advance warning strategies, coping strategies, mitigation strategies for low-likelihood, high impact risks, mitigation strategies for high likelihood, low-impact risk, mitigation strategies for medium-likelihood, moderate impact risks (Stecke and Kumar, 2009; Tang, 2006b; Oke and Gopalakrishnan, 2009). These aforementioned classification systems of the mitigation strategies are somehow based on the functions that the mitigation strategies are required to perform but not on the technical or non-technical attributes of the mitigation strategies.

In the extant literature, there has been less emphasis on the need to group mitigation strategies based on the use of technical procedures or not. Hence, the classification of supply chain mitigation strategies applied in this study is relatively new to the area of supply chain risk management. This is because the classification system is based on whether or not technical know-how or relational attributes are required for carrying out the mitigation strategies. Also, by using the soft and hard mitigation strategies' classification, we want to know if there is a synergistic effect or interaction impact of the hard and soft mitigation strategies on performance. The classification system for the mitigation strategies in this study is based on the assumption that the integration of hard and soft mitigation strategies could create balance and synergistic effects on the performance of firms. According to Kirk (1995), the hybrid of scientific systems which involves the combination of both the human and technical systems give the best solutions. In practice, the implication of this could be that to effectively manage supply chain disruptions, mitigation strategies that are characterised by soft and hard attributes would offer optimal results. In essence, mitigation strategies characterised by soft and hard attributes are important in the area of supply chain risk management.

Soft mitigation strategies are strategies for managing supply chain disruptions or risks but that is relationship-oriented. They are soft mitigation strategies in the sense that they are used for building strong relationships and communication with other parties within the supply chain to manage disruptions. These kinds of relationships could be internally or externally-based in a firm. The internally-based relationships involve the interaction of stakeholders within the firm which involves working together as a team to create solutions to manage challenges faced in the firm. These include brainstorming, teamwork, strategic meetings, information or knowledge sharing, internal and external communication/ relationships, supplier management, monitoring, customer management, strategic alliances, collaborative planning, cooperative culture and joint problem solving (Royer, 2000; Hwacha, 2005; Kaplan et al., 2004; Huo, 2012; Aarikka-Stenroos and Jaakkola, 2012; Chen et al., 2013; Shi and Yu., 2013, Cantor et al., 2014; Gunasekaran, 2003). Externally based relationships include relationships with suppliers and other non-stakeholders. These mitigation strategies enhance the exposure to information, information sharing and facts about supply chain issues. It also involves keeping track of events and activities in the supply chain to ensure control. In fact, through soft mitigation strategies, other soft or hard mitigation strategies can evolve (Royer, 2000).

According to Barrat and Oke (2007), the concepts of visibility and information have been used interchangeably in extant literature. Also, Barrat and Oke (2007:1218) posited “information sharing as an activity whereas visibility is the potential outcome which leads to a more effective supply chain”. Information sharing is a soft mitigation strategy that facilitates the conveyance and transmission of information from one supply chain partner to the other internally and externally. There are also hard mitigation strategies that could be used to manage supply chain disruptions.

Hard mitigation strategies are strategies that require certain techniques to manage disruptions in a supply chain. These mitigation strategies are not relationship-oriented or based on interactions among the parties within the supply chain. These mitigation strategies are more tangible, do not involve interpersonal relationships and somewhat involve the use of techniques, procedures, models or defined technical know-how. These mitigation strategies require technical skills, methodology, processes and or procedures. They include revenue management, dynamic assortment planning, silent product rollover, postponement strategy, strategic stock, economic supply incentives, flexible transportation system adoption by a firm, flexible supply base, alternative sourcing arrangements and outsourced manufacturing

(Baryannis et al., 2019; Stecke and Kumar, 2009; Tang, 2006a, b; Oke and Gopalakrishnan, 2009; Chopra and Sodhi, 2004; Kleindorfer and Saad, 2005; Sawik, 2019; Sheffield and Rice, 2005; Samaddar and Nargundkar, 2010; Braunscheidel and Suresh, 2009; Olson and Wu, 2011; Scholten and Fynes, 2017; Chan et al., 2017; Jabbarzadeh et al., 2018; Vanalle et al., 2019).

THEORETICAL DEVELOPMENT/MODEL

The interaction approach of the contingency theory sees "fit as the interaction of pairs of organisation context and response variables which affects performance" (Sousa and Voss, 2008:706). Therefore, the premise on which this study is based is that the interaction of mitigation strategies and supply chain disruptions will impact on performance (Venkatraman, 1989; Drazin and Van de Van, 1985). This is because the contingency theory helps to explain the interaction between the disruption risk, mitigation strategies and in particular the impact on performance. According to Kirk (1995), the combinations of both relational and technical systems have the potential of giving optimum outcome. This idea could be used to elaborate on how the overlapping or interaction effects of the hard and soft mitigation strategies could improve operational performance when used to manage supply chain disruptions.

A study by Rahman and Bullock (2004) in the analysis of 261 manufacturing firms in Australia confirmed the impact of hard and soft total quality management (TQM) strategies on performance. Similarly, Calvo-Mora et al. (2013) assessed the roles of hard-soft TQM methods on business results. Ahire et al. (1996) also realised a synergistic effect of soft and hard strategies on performance. Zeng et al. (2015) emphasized the importance of hard and soft quality management on quality performance. Similarly, Escrig-Tena et al. (2018) empirically assessed the hard and soft dimensions of quality management on innovation management. All these studies stressed the importance of considering the soft and hard dimensions of factors and the contribution to performance. However, much emphasis is on quality performance in most of these previous studies with none on supply chain risk management. This study is not limited to quality performance but also evaluates the cost, flexibility and delivery performance implications.

The measure of performance used in this study includes the four cardinal measures of operational performance which are delivery, cost, quality and flexibility (Ferdows and De Meyer, 1990; Huo et al., 2014; Khanchanapong et al., 2014; Ketokivi and Schroeder, 2013; Prajogo et al., 2018; Santos Bento & Tontini, 2018). In essence, this study is based on the premise that the use of hard and soft mitigation strategies for managing supply chain risks will generate high impacts on the operational performance of firms.

METHODS

The case study design is used for this exploratory study. A single case is adopted for in-depth analysis of this study for theoretical generalization and robustness (Eisenhardt and Graebner, 2007; Eisenhardt, 1989; Miles et al., 1994; Siggelkow, 2007; Yin, 2003; Yin, 2009). The use of a pseudonym helps to protect the anonymity of the firm (Creswell, 2009). Information on risk sources, hard and soft strategies for managing risks and the performance effectiveness were obtained from the informants. Performance effectiveness implies the degree or extent to which the desired outcome of a performance variable is achieved. The major informants are the operations manager and supply chain manager of the firm. The data obtained were basically from the archival document and interview-based questionnaire/interview guide (Appendix 1).

Description of Case Unit and Analysis

MULT is under the food, beverage and tobacco manufacturing sub-sector. It is a multinational manufacturing firm that manufactures, packages, merchandises and distributes over 100 brands of non-alcoholic beverages. It is a world-class firm with 9 global standard manufacturing plants in different parts of Nigeria. The employee size is over 3,000 and the customer base is above 150 million in Nigeria alone. The supplier base is about 45,000 comprising of small independent and large international firms. This firm exists in several continents of the world, specifically 28 countries and it has been existing for over 100 years. This firm is suitable for this case study because it is a large firm with long years of existence, has a national and global presence, and has large customer and supplier base in Nigeria and the world at large. Hence, MULT has a highly standardized global supply chain which represents significantly the context and the phenomenon (Barratt et al., 2011) evaluated in this study.

The informants from firm MULT included the operations manager and supply chain manager. These informants were asked a series of questions related to the types of disruption risks experienced, mitigation strategies used by the firm and the operational performance effectiveness (Appendix 1). Also, archival data which showed important information about the firm was accessed from the annual reports (Eisenhardt, 1989). Tables 1-5 contain some of the forms of supply chain risks experienced by MULT, the hard and soft mitigation strategies, match of supply chain risks with mitigation strategies and the performance effectiveness respectively.

Table 1 Some Supply Chain Risks Experienced by MULT

Supply Chain Risks
Unexpected demand issues and stock-outs
Transportation issues
Misapplication of system procedures and rules
Erroneous information from customers about orders causing increased liability to the firm
Payment defaults by customers
Issues with suppliers such as poor quality of supplies and late delivery of supplies
Poor controls over systems
Capacity fluctuations in the supply market
Poor public power supply
Operational issues such as machine breakdowns

Source-Authors Computation

Table 2 Some Hard Mitigation Strategies Used by MULT

HARD MITIGATION STRATEGIES
Planning ahead of time and keeping of strategic stock to prevent stock-outs through forecasting
Postponement of products to accommodate customized orders.
Use of flexible transportation means
Sourcing for raw materials from multiple suppliers but keeping to standard registered suppliers in the suppliers' list, sample checks There are certain arrangements with other firms (outsourced firms) to produce some of MULT firm's products on a contract basis Managing customers' choices through the way the products are displayed
Some products are not officially advertised but introduced slowly into the market
Use of contingent suppliers in case of emergency need as the alternative source of raw materials
There are certain arrangements with other logistics service providers to deliver products to customers for MULT firm.
Revenue management through discounts and promotions offered to consumers
Outsourced logistics such as the use of third-party logistics service provider
Economic incentives are provided to gain credible, sustainable and loyal suppliers

Source-Authors Computation

Table 3 Some Soft Mitigation Strategies Used by MULT

SOFT MITIGATION STRATEGIES
Information sharing with other departments, suppliers and customers
Teamwork
Customer relationships
Supplier relationships such as giving reduced credit days, maintaining trust and long term relationships.
The collaborative culture of solving issues
Joint problem-solving approach
External and Internal communication
Information sharing
Strategic meetings tailored to solving peculiar issues. Mutual cooperation in departments and the firm as a whole

Source-Authors Computation

Table 4 Responses on the Forms of Mitigation Strategies for Managing Supply Chain Risks

Control and Operational risks	Demand risks	Power risks	Quality risks	Supply risks	Transportation risks
MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies	MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies	MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies	MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies	MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies	MULT uses both the relationship-based and interaction-focused and the technically-based approaches i.e. the use of both hard and soft mitigation strategies

Source-Authors Computation

Table 5 Performance Effectiveness of the Use of the Forms of Mitigation Strategies

Overall Weight	Performance Effectiveness			
	Cost performance	Delivery performance	Flexibility performance	Quality Performance
10	9	8	7	9
100%	90%	80%	70%	90%

Source-Authors Computation

RESULT AND DISCUSSION

This study has been able to highlight the different forms of hard and soft mitigation strategies for managing supply chain risks (Tables 1 and 4) experienced by manufacturing firms in Nigeria using the case of a multinational firm. The supply chain risks identified are supply, demand, transportation, quality, power, control & operational risks. Demand risks are supply chain risks that erupt as a result of issues in the downstream part of the supply chain such as unanticipated demand (Bode and Wagner, 2006; Chopra and Sodhi, 2004). Supply risks are those risks that emerge from the upstream part of the supply chain such as poor supplies and unavailability of raw materials in the supply market (Cucchiella and Gastaldi, 2006). Power risk is a form of risk that arises from a poor power supply such as erratic or fluctuations in power supply (Soderbom and Teal, 2002). Transportation risks are attributed to issues faced as a result of the movement of goods from one location to another such as transportation delays (Hendrick and Singhal, 2005). Control and operational risks are disruptions experienced by firms with their internal processes (Ghadge, Dani and Kalawsky, 2012; Norrman and Lindroth, 2004).

The mitigation strategies are grouped into two major categories; the hard and soft mitigation strategies. The hard mitigation strategies are comprised of a postponement strategy, multiple sourcing, strategic stocks, flexible transportation and others (Table 2). The soft mitigation strategies include teamwork, customer relationships, supplier relationships and others (Table 3). The different forms of mitigation strategies that are used to mitigate the different categories of supply chain risks are elaborated in Table 4. All the supply chain risks identified in the case study are managed by using both the hard and soft mitigation strategies (Table 4).

Out of the scale of 0 to10, the informants were asked to choose the level of effectiveness of the use of both the hard and soft mitigation strategies for managing the different supply chain risks identified in MULT (Table 5). A case of no level of performance effectiveness is rated 0 which corresponds to 0% and the highest level of performance effectiveness is 10 which corresponds to 100%. The average of ratings of the two informants' percentage weighted scales is used to compute each performance indicator (Table 5). The performance effectiveness of using both the hard and soft mitigation strategy to manage supply chain risks in quantitative terms range from 70%-90% (Table 5). This result shows that the performance effectiveness of using both the interaction-based and technical-based approaches are tenable and above average. Therefore, based on the case study, the integration of both hard and soft mitigation strategies are required to effectively manage supply chain risks to better impact the delivery, quality, cost and flexibility performance of firms.

Both cost and quality performance effectiveness of using both hard and soft mitigation strategies for managing supply chain risks are individually an average of 90%. The quality and cost performance effectiveness have the highest score of 90%. This indicates that the use of hard and soft mitigation strategies to manage quality-related risks is highly effective. This is followed by the delivery performance which is 80%. This also shows a good level of effectiveness but lower compared to the cost and quality performance. A plausible explanation for this could be that on-time delivery of goods could be hindered by poor roads and transportation networks (Larossi, Mousley and Radwan, 2009; Malik, Teal and Baptist, 2006). These transportation challenges are external and uncontrollable factors but could only be managed by firms. The flexibility performance effectiveness has the least percentage of 70% which is also above average. One of the reasons for the lower score of the flexibility performance effectiveness compared to others could be as a result of the overdependence of the firm on road transportation for the distribution of products in Nigeria. There is a poor road network and over-reliance on road transportation in Nigeria (Larossi, Mousley and Radwan, 2009; Malik, Teal and Baptist, 2006; Soderbom and Teal, 2002). An average Nigeria road is poorly maintained resulting in potholes and gullies. Hence, transportation of products from one location to another does not allow for much flexibility and this often results in transportation delays. The way out of this menace for most firms is by contracting out the inbound and outbound logistics to third-party logistics service providers (Bolloré Logistics Nigeria, 2017) to manage the transportation issues while the firms focus on their core competence. The dependence of the firm also on alternative sources of power supply such as the use of diesel generators is also a factor that could have accounted for the low score of the flexibility performance effectiveness. It is indicated from Table 4 that both the hard and soft mitigation strategies are used which comprise of the use of diesel generators and also maintaining a good relationship with the power holding company of Nigeria (PHCN) for regular public power supply respectively. However, in reality, the publicly generated power supply in Nigeria is unreliable and most firms still depend on privately generated sources of energy (Oseni, 2016).

CONCLUSION

In essence, this study has emphasized the importance of the use of both hard and soft mitigation strategies for managing supply chain disruptions. The use of these two categories of mitigation strategies has implications for both academics and managers. For the managers, this implies that the use of hard and soft mitigation strategies for managing supply chain disruptions could yield better result on operational performance because of the synergistic effects. For the academics, this implies that more quantitative studies which involve the use of survey method and regression analysis should be carried out to further test the interaction effect of the hard and soft mitigation strategies on individual supply chain risks and their performance implications. One of the limitations of this study includes the use of the self-reports of informants on the performance effectiveness of the hard-soft hybrid mitigation strategy. However, it was ensured that the informants selected are experts with broad knowledge in the area of supply chain management. Also, this study is focused on the food and beverage industry only and in Nigeria. Hence, the generalization of the findings to other industries in the manufacturing sector could be limited and to other countries as well. Hence, more industries should be explored for comparison and the research should be replicated in other countries in further studies. Also, the scope of supply chain risks emphasized is limited to those with a higher probability of occurrence which could impact day-to-day manufacturing operations. Future studies should extend the supply chain risks to other externally-oriented risks such as catastrophic risks and

other man-made disasters. Furthermore, the possibility of classifying risks as hard and soft supply chain risks could be explored in future studies.

Appendices

Interview Guide

1.Introduction questions related to company description, job position and title

2.What are the supply chain disruptions experienced by your firm?

3.What are the hard mitigation strategies used for managing these supply chain disruptions?

4.What are the soft mitigation strategies used for managing these supply chain disruptions?

5.Do you use both the hard and soft mitigation strategies for managing supply, demand, power, transportation and control & operational risks?

6.Please rank the performance effectiveness of using both hard and soft mitigation strategies for managing the five forms of supply chain risks identified in question 5 using 0-10 rating (with 0 and 10 corresponding to no effectiveness and highest level of effectiveness respectively) based on your experience.

References

Ahire, L.S., Golhar, D.Y. and Waller, M.A. (1996), "Development and validation of TQM implementation constructs", *Decision Sciences, Vol.* 27, pp.23–56.

Aarikka-Stenroos, L., and Jaakkola, E. (2012), "Value co-creation in knowledge intensive business services: A dyadic perspective on the joint problem solving process", *Industrial marketing management*, Vol. *41* No.1, pp. 15-26.

Barrat, M. and Oke, A. (2007), " Antecedents of Supply Chain Visiblity in Retail Supply Chains: A Resource-Based Theory Perspective", *Journal of Operations Management*, Vol.25, pp. 1217-1233.

Barratt, M., Choi, T.Y. and Li, M. (2011), "Qualitative case studies in operations management: trends, research outcomes, and future research implications", *Journal of Operations Management*, Vol. 29, pp. 329–342.

Baryannis, G., Validi, S., Dani, S., and Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. *International Journal of Production Research,* Vol. 57, pp. 2179-2202.

Blackhurst, J., Craighead, C.W., Elkins, D. and Handfield, R.B. (2005). "An Empirically derived agenda of critical research issues for managing supply-chain disruptions", *International Journal of Production Research*, Vol. 43, pp. 4067-4081.

Bode, C. and Wagner, S.M. (2006), " An empirical investigation into supply chain vulnerability", *Journal of Purchasing and Supply Management*, Vol. 12, pp.301-312.

Bode, C. and Wagner, S.M. (2008), "An empirical examination of supply chain performance along several dimensions of risk", *Journal of Business Logistics, Vol.29, pp. 307-325.*

Bolloré Logistics Nigeria to manage warehouse for Unilever. (2017, June 20). STAT Times. Retrieved from https://www.stattimes.com.

Braunscheidel, M.J. and Suresh, N.C. (2009), " The organisation antecedents of a firm's supply chain agility for risk mitigation and response", *Journal of Operations Management*, Vol. 27,pp. 119–140.

Calvo-Mora, A., Picón, A., Ruiz, C., and Cauzo, L. (2013), "The relationships between soft-hard TQM factors and key business results", *International Journal of Operations & Production Management*, Vol. *34*. No.1, pp.115-143.

Cantor, E. D., Blackhurst, J., Pan, M., and Crum, M. (2014), "Examining the role of stakeholder pressure and knowledge management on supply chain risk and demand responsiveness", *The*

International Journal of Logistics Management, Vol. 25, No. 1, pp. 202-223.

Chan, A. T., Ngai, E. W. and Moon, K. K. (2017), "The effects of strategic and manufacturing flexibilities and supply chain agility on firm performance in the fashion industry. European Journal of Operational Research", Vol. 259, No. 2, pp. 486-499.

Chen, J., Sohal, A. S. and Prajogo, D. I. (2013), "Supply chain operational risk mitigation: a collaborative approach", *International Journal of Production Research*, Vol. 51, No. 7, pp. 2186-2199.

Chopra, S. and Sodhi, M.S. (2004), " Managing risk to avoid supply chain breakdown", *Sloan Management Review*, Vol.4, pp. 63-71.

Creswell, J.W. (2009), Research Design. Qualitative, Quantitative and Mixed Methods Approaches. Third Edition. SAGE Publication Limited.

Cuchiella, F. and Gastaldi, M. (2006), " Risk Management in Supply Chain: A Real Option Approach". *Journal of Manufacturing Technology, Vol.* 17, pp.700-720.

Dabhilkar, M., Birkie, E.S. and Kaulio, M., 2016. "Supply-side resilience as practice bundles: a critical incident study", International Journal of Operations and Production Management, Vol. 36 No. 8, pp. 948-970.

Ellegaard, C. (2008), "Supply risk management in a small company perspective", *Supply Chain Management: An International Journal,* Vol. 13 No. 6, pp. 425-34.

Drazin, R. and Van de Ven, H. (1985), "Alternative Forms of Fit in Contingency Theory",*Administrative Science Quarterly*, Vol. 30, pp. 514-539.

Eisenhardt, K.M. (1989), "Building theories from case study research. Academy of Management", *Academy of Management Review*, Vol. 14, pp. 532-550.

Eisenhardt, K. M., and Graebner, M. E. (2007), "Theory building from cases: Opportunities and challenges", *Academy of Management Journal*, Vol. 50, No.1, pp.25-32.

Escrig-Tena, A. B., Segarra-Ciprés, M., García-Juan, B. and Beltrán-Martín, I. (2018), "The impact of hard and soft quality management and proactive behaviour in determining innovation performance", *International Journal of Production Economics*, Vol. *200*, pp.1-14.

Ferdows, K. and De Meyer, A. (1990), "Lasting improvements in manufacturing performance In search of a new theory", *Journal of Operations Management*, Vol. 9, pp. 168-184.

Ghadge, A., Dani, S. and Kalawsky, R. (2012), "Supply chain risk management: present and future scope", *International Journal of Logistics Management*, Vol. 23, No.3,pp. 313-339.

Gunasekaran, A., Lai, K. and Cheng, E. (2008), " Responsive supply chain: A competitive strategy in a networked economy", *Omega, Vol.* 36, pp.549 – 564

Gunasekaran, A., and McGaughey, R. E. (2003), "TQM is suppy chain management". *The TQM Magazine*, Vol. 15, No. 6, pp.361-363.

Hendricks, K.B. and Singhal, V.R. (2003), "The effect of supply chain glitches on shareholder wealth", *Journal of Operations Management*, Vol. 21, pp.501-522.

Huo, B. (2012), "The impact of supply chain integration on company performance: an organizational capability perspective", *Supply Chain Management: An International Journal*, Vol. 17, No. 6,pp. 596-610.

Huo, B., Qi, Y., Wang, Z. and Zhao, X. (2014). The impact of supply chain integration on firm performance: The moderating role of competitive strategy. *Supply Chain Management: An International Journa*l, Vol. 19, No. 4, pp.369-384.

Hwacha, V. (2005), "Canada's experience in developing a national disaster mitigation strategy: A deliberative dialogue approach", *Mitigation and Adaptation Strategies for Global Change*, Vol. *10*, No. 3, pp. 507-523.

Iwarere, H.T. (2010), "How far has the manufacturing industry achieve total quality management in Nigeria", *European Journal of Social Sciences*, London, United Kingdom, Vol. 15, No. 2, pp.34-43.

Jabbarzadeh, A., Haughton, M. and Pourmehdi, F. (2019), "A robust optimization model for efficient and green supply chain planning with postponement strategy", *International Journal of Production Economics*, Vol. 214, pp.266-283.

Juttner, U. (2005), " Supply chain risk management. Understanding the business requirements from a practitioner perspective", *International Journal of Logistic Management*, Vol. 16, pp. 120-141.

Kaplan, R. S., Kaplan, R. E., Norton, D. P., Norton, D. P. and Davenport, T. H. (2004), Strategy maps: Converting intangible assets into tangible outcomes. Harvard Business Press.

Ketokivi, M. A. and Schroeder, R. G. (2004). Manufacturing practices, strategic fit and performance: a routine based view. *International Journal of Operations and Production Management*, Vol. 24, pp. 171-191.

Khanchanapong, T., Prajogo, D., Sohal, A. S., Cooper, B. K., Yeung, A. C., and Cheng, T. C. E. (2014), "The unique and complementary effects of manufacturing technologies and lean practices on manufacturing operational performance"., *International Journal of Production Economics*, Vol. *153*, pp.191-203.

Kirk, D. (1995), "Hard and soft systems: a common paradigm for operations management", *International Journal of Contemporary Hospitality Management*, Vol. 7, pp.13 – 16.

Kleindorfer, P. R., and Saad, G. H. (2005), " Managing disruption Risks in Supply Chains", *Production and Operations Management*, Vol. 14, pp. 53-68.

Stecke, K.E. and Kumar, S. (2009), " Sources of Supply chain Disruptions, Factors That Breed Vulnerability and Mitigation strategies", *Journal of Marketing Channels, Vol.* 16, pp.193-226.

Larossi, G., Mousley, P. and Radwan, I. (2009), "An Assessment of the Investment Climate in Nigeria". 2009 The International Bank for Reconstruction and Development / The World Bank.

Malik, A., Teal F. and Baptist S. (2006), " *The Performance of Nigerian Manufacturing Firms: Report on the Nigerian Manufacturing Enterprise Survey 2004.* Centre for the Study of African Economies". University of Oxford. Oxford OX1 3UQ.UK.

Mand, J.S., Singh, C.D. and Singh, R. (2013). "Implementation of Critical Risk Factors in Supply Chain Management", *International Journal of Research and Business Strategy*, Vol 2, pp. 104- 120.

Mentzer, J.T. (2004), *Global Supply Chain Risk Management*. The University of Tennessee Department of Marketing and Logistics September 30, 2004.

Miles, M.B., Huberman, A.M., Huberman, M.A. and Huberman, M. (1994), Qualitative data analysis: An expanded sourcebook. Sage.

Norrman, A. and Lindroth, R. (2004). Categorization of supply chain risk and risk management. Supply chain risk, Vol.15, No. 2, pp. 14-27.

Oke, A. and Gopalakrishnan, M. (2009), " Managing disruptions in supply chains', *International Journal of Production Economics*, Vol. 118, pp.168-174.

Okon, S.B. (2011). *Supply chain and operations management*. First Edition.

Olson, D.L. and Wu, D.D. (2010). Enterprise risk management models (p. 15). Heidelberg: Springer.

Olson, D. and Wu, D. (2011), "Risk management models for supply chain: a scenario analysis of outsourcing to China", *Supply Chain Management: An International Journal*, Vol. 16 No. 6, pp. 401-8

Oseni, M. O. (2016), "Get rid of it: To what extent might improved reliability reduce self-generation in Nigeria?" *Energy policy*, Vol. 93, pp. 246-254.

Prajogo, D., Toy, J., Bhattacharya, A., Oke, A., and Cheng, T. C. E. (2018), " The relationships between information management, process management and operational performance: Internal and external contexts", *International Journal of Production Economics*, Vol. *199*, pp. 95-103.

Rahman, S. and Bullock, P. (2004), "Soft TQM, hard TQM, and organisational performance relationships: an empirical investigation"*, Omega, Vol.* 33, pp.73 – 83.

Royer, P. S. (2000), "Risk management: The undiscovered dimension of project management", *Project Management Journal*, Vol. *31*, No. 1, pp. 6-13.

Samaddar, S. and Nargundkar, S. (2010), " Analyzing Supply Chain Disruption Risk: A Decision Analysis Framework",*. Operations Management Education Review, Vol. 4, pp. 87-108.*

Santos Bento, G.D. and Tontini, G. (2018), "Developing an instrument to measure lean manufacturing maturity and its relationship with operational performance", Total Quality Management & Business Excellence, Vol. 29, No. (9-10), pp.977-995.

Sarkar, S. and Kumar S. (2016), "Demonstrating the Effect of Supply Chain Disruptions through an Online Beer Distribution Game", *Decision Sciences*, Vol. 14 No. 1, pp.25– 35.

Sawik, T. (2019). Two-period vs. multi-period model for supply chain disruption management. *International Journal of Production Research*, Vol. 57, pp. 4502-4518.

Sheffield, Y. and Rice, J.B. (2005). A supply chain view of the resilient enterprise. *MIT Sloan Management Review*, Vol.47, pp. 41–48.

Shi, M., and Yu, W. (2013), "Supply chain management and financial performance: literature review and future directions*", International Journal of Operations & Production Management*, Vol. 33, No. 10,pp. 1283-1317.

Siggelkow, N. (2007), "Persuasion with case studies", *Academy of Management Journal*, Vol. 50 No. 1, pp. 20-24.

Simangunsong, E., Hendry, L.C. and Stevenson, M. (2016), "Managing supply chain uncertainty with emerging ethical issues", *International Journal of Operations and Production Management*, Vol. 36 No. 10,pp. 1272-1307.

Söderbom, M. and Teal, F. (2002), " *The Performance of Nigerian Manufacturing Firms: Report on the Nigerian Manufacturing Enterprise Survey 2001",* United Nations Industrial Development Organisation and Centre for the study of African Economies. Department of Economics, University of Oxford.

Sodhi, M. S., and Tang, C. S. (2012), "Strategic approaches for mitigating supply chain risks. In Managing supply chain risk" (pp. 95-108). Springer, Boston, MA.

Snyder, L. V., Atan, Z., Peng, P., Rong, Y., Schmitt, A. J. and Sinsoysal, B. (2016), "OR/MS models for supply chain risks: A review", *IIE Transactions*, Vol. 48 No. 2, pp. 89- 109.

Souza, R. and Voss, C.A. (2008), " Contingency research in operations management practices", *Journal of Operations Management*, Vol. 26, pp. 697–713.

Svensson, G. (2000), "A conceptual framework for the analysis of vulnerability in supply chain", *International Journal of Physical Distribution and Logistics Management*, Vol. 30, pp.731- 749.

Tang, C.S. (2006a), " Perspectives in supply chain risk management", *International Journal of Production Economics*, Vol. 103, pp. 451–488.

Tang, C.S. (2006b), " Robust Strategies for Mitigating Supply Chain Disruptions", *International Journal of Logistic: Research and Applications*,Vol. 9, pp. 33-45.

Tang, C.S. and Tomlin, B. (2008), " The power of flexibility for mitigating supply chain risks",*International Journal of Production Economics*, Vol. 116, pp. 12-27.

Vanalle, R. M., Lucato, W. C., Ganga, G. M. D., and Alves Filho, A. G. (2019). Risk management in the automotive supply chain: an exploratory study in Brazil. *International Journal of Production Research*, 1-17.

Venkatraman, N. (1989), "The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence", *Academy of Management Review*, Vol. 14, pp. 423-444.

Yin, R.K. (2003), Case Study Research, Design and Methods, 3rd ed., Sage, Newbury Park, CA.

Yin, R. K. (2009), Case study research: Design and methods (4 ed.). Los Angeles, CA.

SageZeng, J., Phan, C. A., and Matsui, Y. (2015), "The impact of hard and soft quality management on quality and innovation performance: An empirical study", *International Journal of Production Economics*, Vol. *162*, pp. 216-226.

Zhao, L., Huo, B., Sun, L. and Zhao, X. (2013), "The impact of supply chain risk on supply chain integration and company performance: a global investigation", *Supply Chain Management: An International Journa*l, Vol. 18, No. 2, pp.115-131.

Zsidisin, G.A., Panelli, A. and Upton, R. (2000). Purchasing organisation involvements in risk assessment, contingency plans and risk management: an explanatory study. *International Journal of Supply Chain Management, Vol. 5,pp. 187-197.*

DECISION SCIENCES INSTITUTE
Institutional voids, supply chain risks and firm performance in Base of the Pyramid: A systematic review of the literature

Attique ur Rehman
PhD Scholar
Suleman Dawood School of Business
Lahore University of Management Sciences, Pakistan
Email: attique.rehman@lums.edu.pk

Muhammad Shakeel Sadiq Jajja
Associate Professor
Suleman Dawood School of Business
Lahore University of Management Sciences, Pakistan
Email: ssj@lums.edu.pk

ABSTRACT

Purpose: This article analyzes key supply chain risk (SCR) and risk management constructs in the Base of the Pyramid (BoP) literature, that characterize institutional voids.

Design/Methodology/Approach: A systematic review of BoP papers is presented based on content analysis of institutional voids, SCR and risk management and performance constructs.

Findings: Arguments of SCR and risk management are prevalent in BoP characterized by institutional voids.

Research and practical Implications: SCR and firm performance in BoP stream are linked that lead to theoretical development of the subject. Practitioners can manage supply chain risks in BoP related firms to enhance firm performance.

KEYWORDS: Supply chain risk, Base of the Pyramid, supply chain risk management, firm performance

INTRODUCTION

"Bottom of the Pyramid" (BoP) is the term coined by Prahalad and Hammond (2002), which was later called as "Base of the Pyramid", who focused on how multinational corporations can serve the underprivileged profitably as a source of competitive advantage. BoP is generally referred to as "the world's four billion consumers who live on $5 or less per day" (Fawcett & Waller, 2015). It further includes a subset of more than 2.6 billion people who live on $2.00 per day (Arnold & Valentin, 2013). Initially, the BoP research gained momentum by taking the BoP population in emerging economies as potential consumers of multinational corporations. Over time, this focus shifted towards producers' perspective (Gold, Hahn, & Seuring, 2013; Karnani, 2007a) who stressed on the importance of involvement of BoP population as potential producers to serve the needs of subsistence markets. BoP markets (including BoP consumers and BoP producers) are characterized by largely underdeveloped informal institutions (London, Anupindi, & Sheth, 2010), asymmetries in information, weak infrastructures, scarcity of resources (Madhubalan Viswanathan, Seth, Gau, & Chaturvedi, 2007), lack of communication systems and exclusivity of the needs of low-income consumers in subsistence markets (Prahalad & Hammond, 2002; Rivera-Santos, Rufin, & Kolk, 2012).

These attributes of BoP markets create voids in business processes, especially in supply chains (Parmigiani & Rivera-Santos, 2015; Rivera-Santos et al., 2012). Khanna and Palepu (1997) identified five types of institutional voids that exist in product markets, labor markets, capital markets, regulations, and contracts. Existence of these voids in supply chains (Parmigiani & Rivera-Santos, 2015; Wu & Jia, 2018) may expose the firms to increased risks. Supply chain risks are attributed to the probable events that affect firms' supply sources, internal operations and delivery mechanisms (Rangel, de Oliveira, & Leite, 2015). Presence of voids in BoP markets that lead to increased supply chain risks also affects firm performance (Mamabolo & Myres, 2019).

The creation of institutional voids creates challenges in supply chain mechanisms in emerging economies (Wu & Jia, 2018). To date, there has been limited research into what types of institutional voids lead to different supply chain risks that affect firm performance in the context of BoP markets. There is some work that discusses the challenges of supply chains in BoP markets (London et al., 2010; Parmigiani & Rivera-Santos, 2015; Seelos & Mair, 2007) but little work is available that categorizes key supply chain risks that exist in BoP markets due to the presence of institutional voids. Though supply chain risks and risk management have been researched widely in the context of developing countries (Davarzani, Zanjirani Farahani, & Rahmandad, 2015; Enyinda, Mbah, & Ogbuehi, 2010; Moradeyo, 2018; Tukamuhabwa, Stevenson, & Busby, 2017), this area is still under-researched in context of BoP markets. More than forty reviews have been published in the context of BoP, over the past fifteen years, focusing on innovation (Agarwal, Grottke, Mishra, & Brem, 2016; Pansera & Martinez, 2017), economic growth and competitive strategies (Kuada, 2016), sustainable supply chain management (Khalid et al., 2015), understanding consumer behavior (Subrahmanyan & Tomas Gomez-Arias, 2008) and business models and market creation in context of deep poverty (Seelos & Mair, 2007) but there is a lack of conceptualization at the interface of supply chain risk and risk management, institutional voids, and firm performance in BoP literature.

Through a systematic review of the literature we intend to answer the following key research question: “Which institutional voids, supply chain risks and risk management constructs are habitually discussed in BOP literature that affect firm performance?” The aim of the present paper is to enrich the theoretical basis of BoP research by analyzing the use of established supply chain risks, risk management and firm performance constructs (Ho, Zheng, Yildiz, & Talluri, 2015; Kaplan & Norton, 1999; Manuj & Mentzer, 2008; Rao & Goldsby, 2009; Ritchie & Marshall, 1993) in BoP-related publications that characterize institutional voids (Khanna & Palepu, 1997). By explicating the nature and influence of institutional voids and supply chain risks on firm performance in BoP context, we offer several contributions to the literature. Through frequency and contingency analysis of the literature, we expand and deepen the concept of institutional voids and supply chain risks by discussing the nature of each void and its relationship with different categories of supply chain risks. Second, we contribute to supply chain risk literature by mapping institutional voids (Khanna & Palepu, 1997) and supply chain risks towards BoP firm performance (Mamabolo & Myres, 2019). We argue that different institutional voids lead to exposure of BoP firms towards different supply chain risks that affect firm performance. Third, we contribute to BoP and subsistence market literature by explicitly connecting supply chain risk and risk management concepts with firm performance in BoP context. Fourth, this research can assist managers in understanding and managing supply chain risks by identifying institutional voids that exist in supply chains of BoP firms.

The structure of this paper is as follows. The next section provides the reader with a basic understanding of the terms used here. Section three elaborates research methodology that is employed to conduct a literature review. The fourth section presents the result of a systematic literature review. We categorize our findings of supply chain risk in BoP into frequency-related and contingency-related findings. Section five discusses the limitations of the study along with comprehensive discussion and a brief conclusion is provided in the end.

KEY TERMINOLOGIES AND CONCEPTUAL FRAMING

Supply chains in the base of the pyramid

The term *base of the pyramid* gained attention from the seminal work of Prahalad and Hammond (2002) that refers to those individuals that are excluded from what we call as ‘global capitalism’ (Arnold & Williams, 2012). Initially, the BoP research (referred to as BoP 1.0) called for the role of formal corporate sector in developed countries to engage in business activities in developing economies by *“selling to the poor”* (Simanis, Hart, & Duke, 2008). Karnani (2007a) critically questioned the capitalist role of multinational corporations in BoP markets that were *“doing well but not doing good”* and stated that the only way to alleviate poverty is to focus on the poor as producers rather than as a market of consumers. This new notion in the context of BoP was termed as BoP 2.0 *(“Business Co-venturing”)* by Simanis et al. (2008) that broadened the landscape of BoP research as it is an important development in BoP theory (Tate, Bals, & Marshall, 2019). Hart (2017) extended this phenomenon to BoP 3.0 focusing on the role of BoP firms for social development.

Supply chains involve all the key activities that are associated with the flow and transformation of goods from raw material to final product and distribution of these products through various

channels till they reach end consumer (Handfield & Nichols Jr, 1999; Mentzer et al., 2001). Activities within the supply chain include the procurement of raw material, production, distribution, inventory management, and marketing. According to the Supply Chain Operations Reference (SCOR) model, there are five key processes in supply chains: Plan, Source, Make, Deliver and Return (SCC, 2010). We put BoP firms as a focal firm in supply chain and how they manage supply chain activities and form relationships with other organizations to satisfy the end user. In the context of developed countries, focal firms can clearly identify sources of raw materials and establish contractual agreements with suppliers. They have sufficient finances available to execute production operations and establish distribution channels. Marketing activities are characterized by open communication with well-informed consumers and brand development.

Business activities in BoP are different from modern business landscapes in terms of business logic and supply chains. Because of different business context, applying traditional business logic does not yield results in BoP markets (Khalid et al., 2015). Supply chains in BoPs are different because of varying needs of BoP consumers (Rosca & Bendul, 2019), high market uncertainties (Parmigiani & Rivera-Santos, 2015) and lack of financial and other resources for BoP producers (Reficco & Gutiérrez, 2016). Thus BoP supply chains should be adapted to accommodate market changes (Lee, 2004; Zomorrodi, Fayezi, Lau, & McMurray, 2019). Similarly, BoP producers need to acquire key skills and capabilities to work as part of both local and global supply chains that may help in reducing uncertainties and increasing sustainable performance and economic development. Recently, researchers have called for more radical solutions for BoP markets in terms of business strategies and supply chain concepts (Schrader, Freimann, & Seuring, 2012; Zomorrodi et al., 2019). Ansari, Munir, and Gregg (2012) identified BoP as a nascent academic field of research and there exists a need to develop ingenious supply chains and innovative business solutions in BoP (Seuring & Khalid, 2019).

Institutional voids in the base of the pyramid

All supply chains are entrenched with a set of institutions which provide foundations for economic exchange, formal laws, and practices that reflect in transparency, protection of property rights and efficiency of markets (Li & Abiad, 1990; Mair & Marti, 2009). If these formal institutions are absent, weak or have limited enforcement power, gaps will be created that may affect market efficiencies. As BoP markets are secluded mainstream markets, such gaps in the business ecosystem become inevitable (Webb, Kistruck, Ireland, & Ketchen, 2010). Khanna and Palepu (1997) named these gaps as “institutional voids” and defined them as “lack of formal institutions that are necessary to support business operations”. Scholars have extended this concept and have concluded that BoP markets are affected by even more severe institutional voids (Mair, Marti, & Ventresca, 2012; Rivera-Santos et al., 2012). According to Parmigiani and Rivera-Santos (2015), these voids affect organizations at dyadic (between individual buyers and sellers) and network level (multi-tiers). Product markets and contracting voids occur at the dyadic level and capital markets, regulations, and labor markets voids ensue at the network level.

Product market void corresponds to the failure in product markets. With the focal firm in the center of the supply chain, this dyadic void exists in upstream (between firm and suppliers) or downstream (between firm and its customers). In upstream, lack of intermediaries and standards to help identify potential suppliers or inability of firms to provide an assessment on input quality, this void make it difficult for the firms to identify capable supplier (Parmigiani & Rivera-Santos, 2015). Likewise, lack of information and understanding of customer needs, firms may face difficulty in providing the customers with appropriate products. Attributed to the low literacy rate of the customers, it will be difficult for the firms to communicate with customers (Karnani, 2007b; Weidner, Rosa, & Viswanathan, 2010).

Labor market void is a network level void that makes it difficult for the firms to identify skilled employees (Madhu Viswanathan & Rosa, 2010). In subsistence markets, people rely on skills that they have developed over time without any formal education and training (A. Banerjee, Duflo, Glennerster, & Kinnan, 2015) but the requirement of these skills may vary based on the needs of firms. Most emerging markets suffer from a scarcity of well-trained labor (Khanna & Palepu, 1997). Some skills may be widely available in the market, but lack of channelization restricts the ability of the firms to identify required skillset (A. V. Banerjee & Duflo, 2007).

Capital market void is also a network level void that affects the whole market. Informal structure of firms and lack of access to microfinancing or other financing sources make it difficult for the firms to acquire required capital (A. Banerjee et al., 2015; Chu, 2007). Firms in subsistence markets face dearth of capital to sustain their operations and pursue growth (Khanna & Palepu, 1997). Though microfinance institutions have tried to close this void to a certain extent, still informal nature of BoP firms makes it difficult for them to acquire capital from the services offered by these institutions. The impact of capital market voids varies across BoP firms' activities that need to be financed. Activities that require less financing are likely to be impacted in a minimal way by capital market void whereas, the activities that require significant capital infusion will face scarcity of financing options (Parmigiani & Rivera-Santos, 2015).

Regulatory void is defined as unpredictable and changing rules and regulations that make it difficult for the firms to forecast future regulatory conditions (Khanna & Palepu, 1997). This is also a network level void that originates from the national or regional and local community level. National or regional level regulations are influenced by the government mainly that affects firms' operations but in subsistence markets, regulations vary at local community level in terms community and societal sphere, political and religious sphere (Mair et al., 2012). These regulations make it difficult for BoP firms to operate uniformly.

Contracting void is defined as lack of formal written contracts due to enforcement problems (Khanna & Palepu, 1997). Contracting void is dyadic in nature, as a contract is a mutual understanding and agreement between two parties. In BoP markets, informal institutions are prevalent as compared to formal institutions (London & Hart, 2004; Rivera-Santos et al., 2012; Webb et al., 2010) that lead to weak contracts that can be breached easily. Agreements in informal institutions do exist but these agreements are mainly at community or personal level, based on trust and embeddedness in social networks (Mair et al., 2012) and are not generalizable to more formal institutions.

Supply chain risks and risk management in base of the pyramid

Moving towards supply chain risk perspective, one of the first authors to define the term supply chain risk were March and Shapira (1987). They defined supply chain risk as “variations in the distribution of possible supply chain outcomes, their likelihood, and their subjective values.” Supply chain risk has been defined by researchers focusing on supply risk (Zsidisin, 2003), information flow, material flow and product flow risk (Jüttner, Peck, & Christopher, 2003; Peck, 2006) and micro and macro level risks (Ho et al., 2015).

Supply chain risk is a mature research stream being explored by the researchers through systematic and non-systematic literature reviews over the past decade (Fan & Stevenson, 2018) for over two decades. There are diverse views in the categorization of supply chain risks (M. S. Sodhi, Son, & Tang, 2012). Supply chain risks have been attributed to various sources like environmental factors, organizational factors (Jüttner et al., 2003), physical, financial, informational, relational and innovational factors (Cavinato, 2004), categories that pertain to systems, forecast, intellectual property, receivables, inventory and capacity risks (Chopra & Sodhi, 2004), supply, demand and contextual factors (M. Sodhi & Lee, 2007) and supply, process, demand, intellectual property, behavioral and social risks (Tang & Tomlin, 2008).

Drawing from various risk events found in the literature, Manuj and Mentzer (2008) identified interactions between operational, supply and demand risk in the context of “other risks” (figure 1). Some of the risk events have been identified under each heading. “Other risks” include the events that are uncontrollable and fall beyond the boundary of the organization currency fluctuation, wage rate shift and such events that may affect the security of firm’s supply chain (Manuj & Mentzer, 2008).

The frameworks by Manuj and Mentzer (2008) offer a comprehensive categorization of supply chain risks so we use these two frameworks to map supply chain risks in BoP related literature.

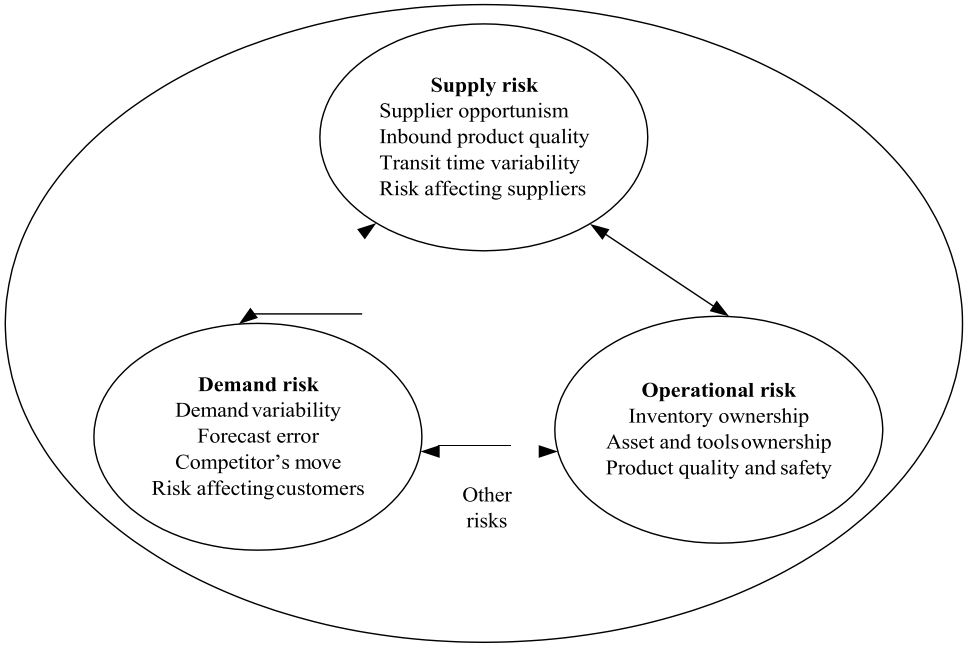


Figure 1: Types of risks by Manuj and Mentzer (2008)

Within supply chain risk literature, supply chain risk management has also gained importance (Fan & Stevenson, 2018). Supply chain risk management is aimed at identification, assessment, treatment, and monitoring of risks in supply chains. (Neiger, Rotaru, & Churilov, 2009). Jüttner et al. (2003) suggest four key elements of supply chain risk management i.e., assessing risk sources, identification of risk concept, tracking risk drivers and mitigating risk. Similarly, Kleindorfer and Saad (2005) identified three stages of risk management (1) specifying sources of risks, (2) assessment and (3) risk mitigation. Extending this definition, Fan and Stevenson (2018) provided a comprehensive definition of supply chain risk management that also caters for collaboration with supply chain members to reduce vulnerability in supply chains.

Supply chain risk and risk management have received little attention in the context of BoP (Khalid et al., 2015). Simanis and Duke (2014) identified that achieving the success of firms in Bottom of the Pyramid is dependent on firms' supply chains. Modern supply chains have not been designed to cater to the needs of the world's lowest-income consumers and producers (Fawcett & Waller, 2015; Simanis & Duke, 2014). In BoP markets, challenges in supply chains permeate especially in the downstream value-added process (Fawcett & Waller, 2015) i.e., product development, distribution, and customers' experience. Recently researchers have focused their efforts towards theory building at the intersection of supply chain and BoPs (Bendul, Rosca, & Pivovarova, 2017; Khalid et al., 2015; Seuring & Khalid, 2019). The concept of sustainable supply chains for BoP is prevalent in the literature (Bendul et al., 2017; Kuo, Shiang, Hanafi, & Chen, 2018; Rao, Nilakantan, Iyengar, & Lee, 2018) but supply chain risk and risk management has not yet been discussed in Base of the pyramid, though sustainability in supply chains and supply chain risks and risk management are part and parcel under the broad umbrella of supply chain management.

Institutional voids, supply chain risks and firm performance in the base of the pyramid

BoP literature views progressive business performance as a function of overcoming institutional voids and developing successful supply chains. Decision theorists argue that risk is not always the downside possibility but it may also result in something higher than expected, however, in supply chain management, it is the downside of risk that reflects in a business context that is characterized by institutional voids (Wagner & Bode, 2008). Institutional voids have conceptual commonalities with supply chain risks. Firms in BoP are more prone to have institutional voids thus are more exposed to supply chain risk. Amankwah-Amoah, Chen, Wang, Khan, and Chen (2019) developed a supply chain model that helps organizations overcome institutional voids and become ambidextrous to enhance performance. Supply chain risk has been discussed implicitly in BoP literature that is evident due to institutional voids found in subsistence organizations, but there is lack of extant literature that explicitly discusses supply chain risks and firm performance in subsistence markets.

Firm performance is characterized by supply chain risks and risk management. Firms that frequently get involved in complex supply chains and incur supply risks lead to negative performance (Wagner & Bode, 2008). Vast research has been conducted that discusses the impact of supply chain risks and risk management on firm financial performance (Chen, 2018; Shou, Hu, Kang, Li, & Park, 2018), operational performance (Afshar & Fazli, 2018; Shou et al., 2018) and buyer and supplier performance (Revilla & Saenz, 2017). In the context of BoP literature, firm performance has been measured largely on the basis of sustainability (Monzer, Rebs, Khalid, & Brandenburg, 2018; Seuring & Khalid, 2019). Mamabolo and Myres (2019) argued that balanced scorecard can be used as a tool to measure the performance of firms in the bottom of the pyramid. The central idea of the balanced scorecard is that firms performance is a function of four perspectives i.e., financial measures, customer focus, internal processes and innovation and learning (Kaplan & Norton, 1999). There exists a tension in BoP firms between generating social versus financial impact (Mamabolo & Myres, 2019). Firm performance at BoP level has been measured using sustainability indicators (like Triple Bottom Line reporting) (Monzer et al., 2018; Seuring & Khalid, 2019), but phenomena like innovation, learning, and customer are discussed more frequently in the literature at BoP level (Borchardt et al., 2018; Demssie, Wesselink, Biemans, & Mulder, 2019; S. Gupta, 2019; Onsongo, 2019; Rosca & Bendul, 2019). Apart from financial performance, the dynamics of BoP firms contribute more towards social, innovation, learning and customer performance. We use balanced scorecard model (Kaplan & Norton, 1999) to map firm performance in BoP and how that performance is characterized by supply chain risks and institutional voids.

We use established frameworks of institutional voids, supply chain risk, and risk management, and firm performance (Kaplan & Norton, 1999; Manuj & Mentzer, 2008; Prahalad & Hammond, 2002; Rao & Goldsby, 2009) to link existence of various types of institutional voids with supply chain risk and their relationship with performance in BoP firms.

METHODOLOGY

To address the aim of this paper, we applied a systematic literature review technique (Tranfield, Denyer, & Smart, 2003). This technique helps to systematically identify and collect relevant literature (Seuring & Khalid, 2019) thus avoiding ad hoc collection of randomly selected articles. Systematic reviews were developed in the field of medicine to integrate mammoth sets of available information and provide baseline data for decision making (Mulrow, 1994). It is a transparent approach to locate, appraise and synthesize the existing studies in a field of research to provide new insights on the phenomena (Petticrew, 2001). Since Tranfield et al. (2003) introduced this technique to the domain of management and organizational studies, it has become increasingly popular to bring together a diverse body of evidence so that it can reveal useful insights. Initially, a scoping study was carried out to identify the need for this review as mentioned in the introduction section. Using the recommendations provided by Tranfield et al. (2003) and Denyer and Tranfield (2009), a review protocol was established that include research question and search strategy of the literature and criteria for inclusion and exclusion. Denyer and Tranfield (2009) provided a five-step process for the systematic literature review: "(1) question formulation, (2) Locating studies, (3) Study selection/evaluation (4) analysis and (5) reporting results."

Question formulation: aligned with our objective of the study, the key question formulated for this systematic review was "What different types of supply chain risks and risk management strategies are habitually discussed in BoP literature in the presence of institutional voids that affect firm performance?"

Locating the studies: Bibliometric research has identified Elsevier's Scopus as the largest database of peer-reviewed literature (Jagtap, 2019; Mongeon & Paul-Hus, 2016). Scopus includes a wide range of journal articles, book chapters and conference publications (Jagtap, 2019). There exists a large overlap between search results of the web of science and Scopus (Gavel & Iselid, 2008), so we employ Scopus for searching the literature. Recent studies have also used Scopus as the only database to locate articles for literature review purpose (Baryannis, Validi, Dani, & Antoniou, 2019; Roy & Singh, 2017).

Study selection/evaluation: Two key phrases *"base of the pyramid"* and *"bottom of the pyramid"* were used to locate the studies as these terms are used interchangeably in BoP literature. Initially, the search returned 1024 articles spanning over more than 15 years. We restricted the search to the English language, peer-reviewed journals during the period 2004-2019 as after 2004 there was consistent growth in BoP literature. Further, we limited the search to the subject area of business management and accounting only. The refined search returned 219 articles. A further process of reading, analyzing and discriminating reduced the number to 94. We focused on all the three perspectives of BoPs i.e., BoP 1.0 that focuses on challenges of multinational corporations to target underprivileged (Prahalad & Hammond, 2002), BoP 2.0 that caters for the capacity building of subsistence firms to serve the needs of subsistence markets (Karnani, 2007a; Simanis et al., 2008) and BoP 3.0 that targets the role of subsistence firms in social development through innovation (Hart, 2017).

Analysis: Only those papers mentioning (implicitly or explicitly) one or more constructs of institutional voids, supply chain risks and risk management derived from Khanna and Palepu (1997), Manuj and Mentzer (2008), Rao and Goldsby (2009) and Kaplan and Norton (1999) were selected. In a subsequent process, we coded all 94 papers against the constructs mentioned. Only those papers were coded that were implicitly or explicitly mentioning the specific construct(s).

In order to ensure the validity and reliability of the overall process, interrater reliability was established by involving multiple researchers to code thirteen papers on a random basis. Beyond the pilot phase coding, the rest of the papers were coded by a single researcher. To ensure construct validity, the papers were coded keeping the original description of the construct in mind.

Following the completion of the coding phase, the frequencies of individual constructs were calculated. The frequencies relate to the number of occurrences of a particular construct as a percentage of the base sample. Further, a thorough understanding of the phenomena was developed using contingency analysis that measures paired relationships amongst various constructs without specifying causality. To ensure reliability, the contingencies were tested amongst those pairs of constructs that showed a frequency of at least 10%.

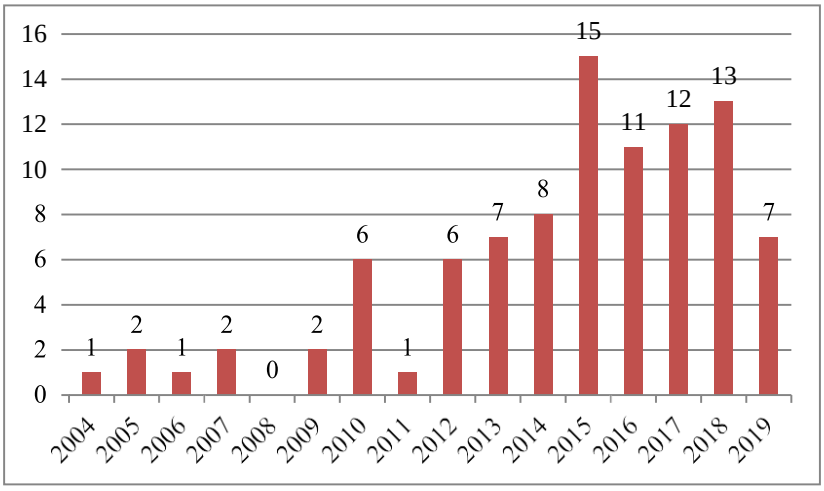


Figure 2: BoP papers that use specific constructs (n = 94)

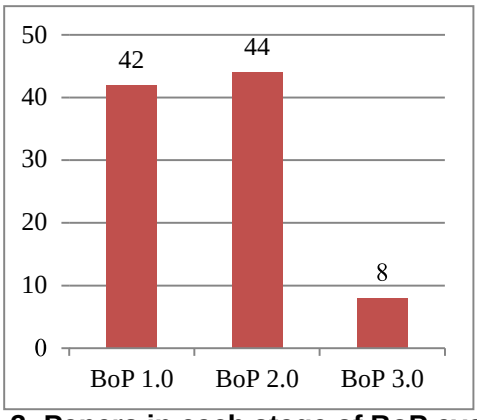


Figure 3: Papers in each stage of BoP evolution

RESULTS AND FINDINGS

BoP discussion has gained importance in the literature in recent years (Kolk, Rivera-Santos, & Rufín, 2014). A close analysis of selected articles in the domain of BoP indicates that this research stream has grown since 2010 (figure 2). Although BoP scholars have published across a range of business and management journals, certain journals stand out in this comprehensive list. BoP literature is discussed mostly in the Journal of Business Ethics (JBE) and Journal of Business Research (JBR). We move next to detailed results of construct-driven literature review.

We also bifurcated BoP literature on the basis of three perspectives at different points in time as BoP 1.0, BoP 2.0 and BoP 3.0 (figure 3). It is found that BoP 2.0 has gained importance over the last decade, focusing on enhanced capabilities of subsistence producers to serve the need of BoP markets innovatively.

Frequency analysis

We coded BoP papers against core constructs from the selected frameworks of institutional voids, supply chain risks and risk management and firm performance. The first step was to assess the frequencies of single constructs in BoP literature.

Institutional voids constructs: As mentioned earlier, BoP papers were coded against the institutional voids constructs developed by Khanna and Palepu (1997). BoP papers regularly refer to various institutional voids that are considered integral to business challenges in informal markets. These include product market voids (Bucher, Jäger, & Prado, 2016; A. D. Gupta, 2019; Onsongo, 2019; Siqueira, Mariano, & Moraes, 2014; Thun, 2018); labor market voids (Hens, 2012; Ramani & Mukherjee, 2014; Schuster & Holtbrügge, 2012); capital market voids (Armstrong, Ahsan, & Sundaramurthy, 2018; Bardy, Drew, & Kennedy, 2012; Lim, Han, & Ito, 2013; Perrot, 2017; Reynoso, Kandampully, Fan, & Paulose, 2015); regulatory voids (Agarwal, Brem, & Grottke, 2018; Bruton, Ahlstrom, & Si, 2015; Paras, Ekwall, & Pal, 2019; Rabino, 2015) and contracting voids (Chakrabarty & Bass, 2015; Hahn & Gold, 2014; London, Esper, Grogan-Kaylor, & Kistruck, 2014). It confirms that BoP literature is highly characterized by institutional

voids, especially product markets, capital markets and regulatory voids that show a frequency of 78.7%, 62.7% and 68.1% in total of 94 papers respectively. The base of the pyramid is characterized by poverty and social problems (Demssie et al., 2019) that lead to the creation of institutional voids in BoP markets. In other words, the informal nature of BoP markets creates gaps in business operations and restrict the abilities of firms, especially multinational corporations to perform at a certain level to exploit opportunities (A. D. Gupta, 2019). There exists a product market void due to non-affordability of the customers in BoP (Schaefers, Moser, & Narayanamurthy, 2018).

Supply chain risks and risk management constructs: Different types of voids are likely to affect different stages of supply chains (Parmigiani & Rivera-Santos, 2015) thus resulting in the emergence of different types of supply chain risks. We screen BoP literature for supply chain risks constructs. A close analysis of BoP papers revealed that various supply chain risks are habitually and implicitly discussed that can be traced to supply chain risk constructs established by Manuj and Mentzer (2008) and Rao and Goldsby (2009). Figure 4 summarizes the frequencies of supply chain risk constructs against the framework suggested by Manuj and Mentzer (2008). Under the three broad categories of supply chain risks, the frequencies of supply risks: supplier opportunism (18.1%), inbound product quality (15.9%), transit time variability (11.7%) and risk affecting suppliers (12.8%), demand risks: demand variability (37.2%), forecast error (7.4%), competitors' move (4.2%) and risk affecting customers (48.9%) and operational risks: inventory ownership (13.8%), asset and tools ownership (26.6%) and product quality and safety (32%) reveal that BoP literature is highly characterized by supply chain risks. Out of the mentioned frequencies, risk affecting customers and demand variability show strikingly higher frequencies. It supports the characterization of product market voids in subsistence market that lead to demand variability and risks affecting customers.

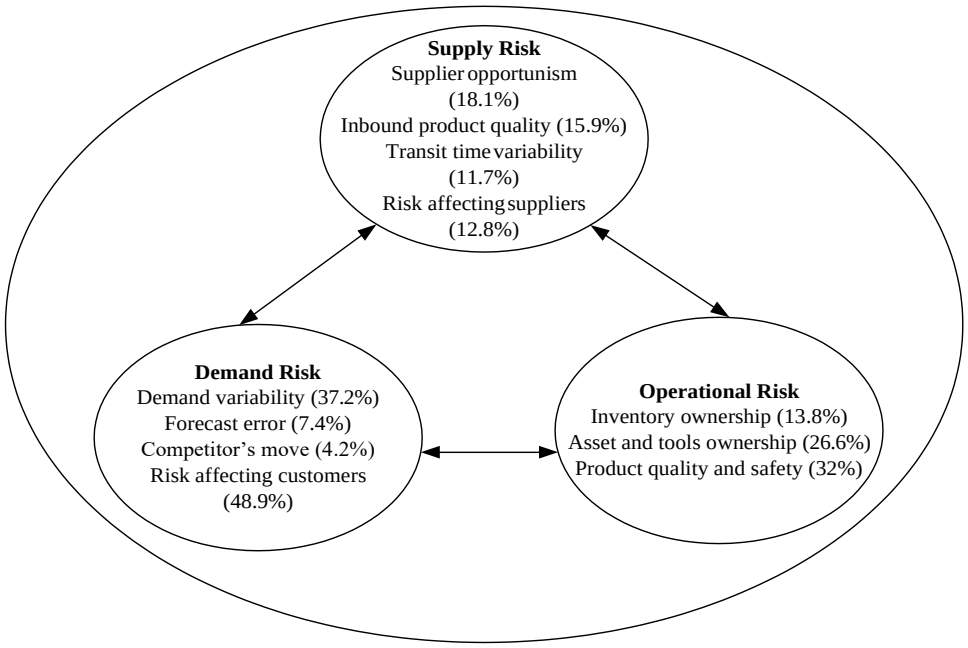


Figure 4: Frequencies of supply chain risk constructs in BoP literature

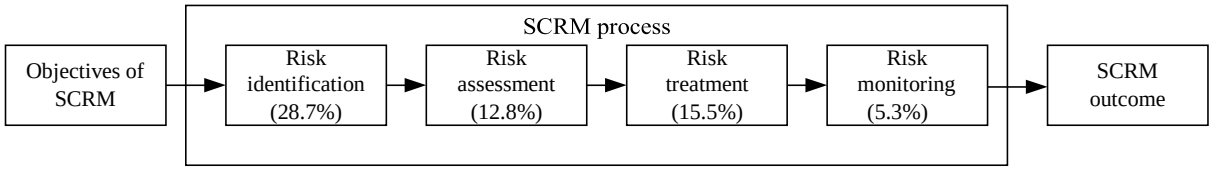


Figure 5: Frequencies of supply chain risk management constructs in BoP literature

BoP firm performance constructs:

Performance of firms in BoP context is highly characterized by innovation, customer needs satisfaction, financial stability and internal business perspective (Mamabolo & Myres, 2019). There are myriad approaches to measure firm performance and no unified model has been consented upon to determine the performance of a BoP firm like a social enterprise (Grieco, Michelini, & Iasevoli, 2015; Yang, Huang, & Lee, 2014). The key motivation behind using balance scorecard as a performance metric is the default nature of BoP firms that precisely focus on innovative and creative ways to satisfy the needs of base of the pyramid consumers that exist in informal markets in a weak institutional setting.

Our frequency analysis suggests that all the four dimensions of balance scorecard have been repeatedly used in BoP literature with the highest focus on innovation and learning perspective (80.1%) followed by customer perspective (49.0%). Financial perspective and internal business perspective are also characterized in BoP literature (40.4% and 42.5% respectively). The higher frequencies of performance attributed to innovation and customer perspective are aligned with the very nature of BoP. The underlying theme of BoP firms is to satisfy the needs of underprivileged innovatively, so balance scorecard precisely captures BoP firm performance mechanisms.

Contingency analysis

To fully comprehend the scholarly perspective on BoP, we performed contingency analysis on the constructs of institutional voids, supply chain risks, risk management and firm performance traced in BoP literature. Contingency analysis help infer the association between pairs of constructs. The direction of these associations is open for theoretical interpretation. Gold, Seuring, and Beske (2010) defined contingency analysis as “a positive association pattern between categories, i.e., it identifies pair of categories that occur relatively more frequently together in one paper than the product of their single probabilities would suggest.” A contingency is measured in terms of phi-coefficient (ϕ). A value of $\phi > 0.2$ suggests moderate association (Rea & Parker, 1992) and $\phi > 0.3$ suggests strong association between constructs (Fleiss, Levin, & Paik, 2013; Gold et al., 2010; Seuring & Khalid, 2019). The greater the value of ϕ, the stronger will be the association between constructs. The value of ϕ provides a degree of strength of association between pairs of constructs, so contingency analysis provides a scope of theoretical interpretation of observed contingencies between different pairs of constructs.

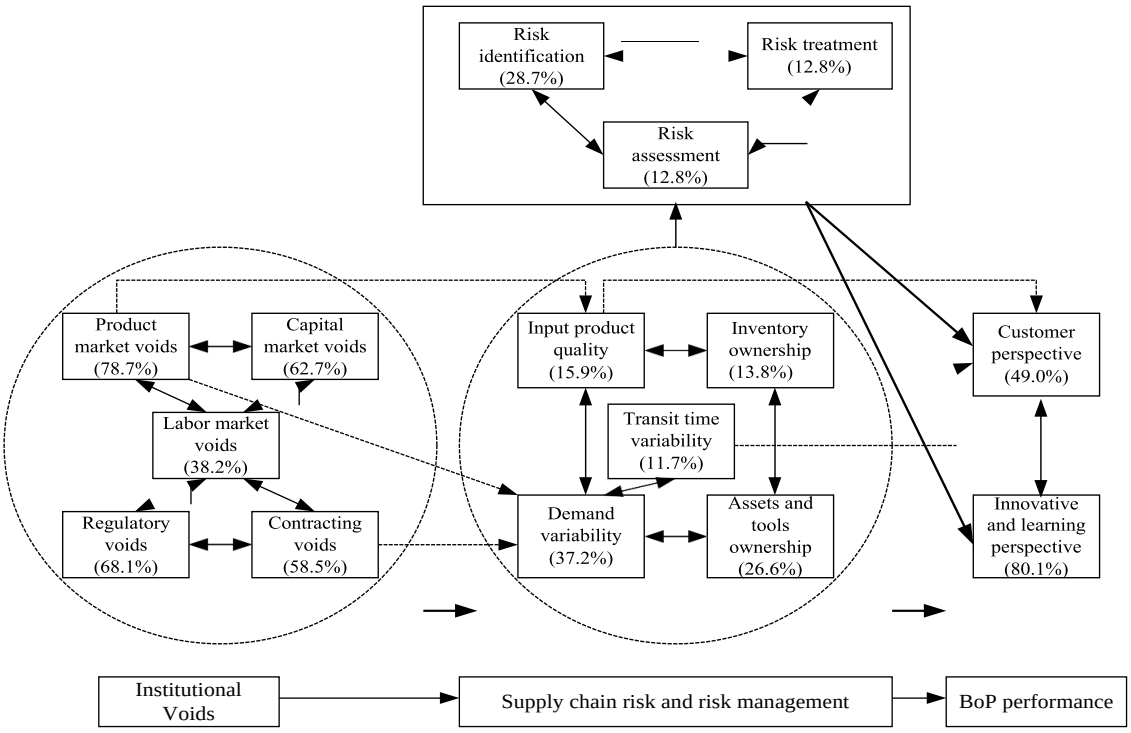


Figure 6: Contingencies amongst institutional void, supply chain risk and risk management and BoP performance constructs

Theorized relationship Contingent relationship Contingent association

DISCUSSION

This paper puts supply chain risk and risk management in the BoP context. By presenting sets of constructs from institutional voids, supply chain risk and risk management, and firm performance, this paper tries to bridge three research streams on the basis of a sound methodological analysis of the literature. The in-depth analysis of implicit and explicit use of supply chain risk and risk management constructs in BoP literature contributes to knowledge generation and opens horizons for future research. The present findings highlight those supply chain risk constructs that are relevant to BoP researchers and practitioners when addressing the unique challenges of this segment caused due to the existence of institutional voids.
BoP literature emphasizes the presence of institutional voids due to informal market systems (Parmigiani & Rivera-Santos, 2015) and these voids lead to exposure of these firms to supply chain risks. Identifying supply chain risk and managing these risks lead to enhanced firm performance (Wieland & Marcus Wallenburg, 2012) especially financial performance (Hendricks & Singhal, 2005), but in BoP context, there is more emphasis on innovative and customer-oriented performance (Mamabolo & Myres, 2019). BoP market demands novel and unique solutions to problems driven by informal market economies and institutions (Onsongo, 2019). The key obstacle concerning the performance of BoP firms is the process of supply chain risk management.

Though various categories of supply chain risks are discussed in BoP literature, a very limited number of articles discusses an explicit process of supply chain risk management. For instance, based on case studies on three multinational corporations, Schuster and Holtbrügge (2012) identified the constraints (risks) faced by these companies and the ways they overcome it successfully. Based on our earlier discussion of BoP 1.0, 2.0 and 3.0, it is identified that supply chain risk management is discussed more in BoP 1.0 literature (Ausrød, 2018; Perrot, 2017; Zomorrodi et al., 2019) focusing implicitly on risk identification, risk assessment and risk treatment by large corporations to overcome challenges in serving BoP markets. Due to the very nature of BoP 2.0, the micro firms that target the subsistence market operate with limited resources and under institutional voids (Davies & Torrents, 2017). Our frequency and contingency analysis revealed that in BoP 2.0 context, the key risks include policy uncertainty, product market risk, decision maker risk including knowledge/skills/biases and bounded rationality, and macro-economic, social and operating uncertainties. Interestingly, the frequency analysis showed that policy risks in discussed only in 13.8% of the BoP papers but its has a significantly high number of contingencies as compared to other risk categories. Policy uncertainty is amplified in the presence of institutional voids due to weak implications of legal and government regulations in the BoP context (Goyal, Sergi, & Jaiswal, 2016; Weerawardena & Mort, 2006). Similarly, input market risks only showed a frequency of 10% but has a significant number of contingencies with decision maker risks (informations seeking, knowledge/skills/biases and rules and procedures). Input market risks refer to the uncertainties of quality and quantity of input of BoP firms. These risks are naturally driven due to the lack of knowldege and access to information and unawareness about existing rules and regulations of procurement, especially for BoP producers. At BoP level, the established culture, beliefs, influence of monarchs, micro-level economic systems and customaries of the firms restrict their

abilities to forsee risks (Nakata & Antalis, 2015). For BoP firms to ehance their performance, these risks must be addressed adequately, but the BoP literature is inconclusive in overcoming these risks through risk management strategies.

Our analysis of BoP literature identified categories of institutional voids and supply chain risks vary in BoP 1.0, 2.0 and 3.0. Regulatory and contracting voids are more frequently discussed in BoP 1.0 context (Demssie et al., 2019; Gutiérrez & Vernis, 2016; Schuster & Holtbrügge, 2012; Thun, 2018) whereas product market, labor and capital market voids are more habitually discussed in BoP 2.0 context (Chowdhury, 2007; Ramachandran, Pant, & Pani, 2012; Webb, Morris, & Pillay, 2013). BoP 3.0 concept is relatively new and there is no specific focus on particular categories of institutional voids in BoP 3.0.

The relationship between supply chain risks, risk management, and firm performance is also theorized from the literature. Initially, the firm performance in BoP context is focused more on financial growth of large corporations by serving subsistence markets (Adegbile & Sarpong, 2018; Agarwal et al., 2018; Prado, Calderon, & Zúñiga, 2016; Reiner, Gold, & Hahn, 2015) but with the growth of literary debate in BoPs, other performance indicators like innovation performance and customer orientation also became focus of the academics in BoP literature (Hall, Matos, & Martin, 2014; Jose & Robert Buchanan, 2013; Ramachandran et al., 2012). Based on our contingency analysis the decision makers' risks are contingent with innovation and learning perspective of BoP firm performance. Similarly, social, liability and credit risks are contingent with the customer perspective.

Upstream supply chain members (BoP 2.0) recently referred to as "base of the chain" (Tate et al., 2019) are differentiated from downstream supply chain members (BoP 1.0) in terms of supply chain activities. More research has been conducted on downstream members of supply chains in BoP (Tate et al., 2019). A special issue of the International Journal of Physical Distribution and Logistics Management highlights the need for research to address both customers and suppliers in BoPs (Tate et al., 2019). Current BoP literature talks about inclusive business opportunities in BoP markets through enhancing the capabilities of micro-producers to serve subsistence markets but provides a limited perspective when it comes to defining the role of BoP actors in managing supply chain risks. Though earlier research discusses the challenges of scalability, cultural nuances and infrastructure in the context of BoP markets (M. S. Sodhi & Tang, 2016) but the scholars are yet to come up with a more holistic view of managing supply chain risks to enhance firm performance in entrepreneurial activities at BoP firm level. Companies in BoP face resource constraints (M. S. Sodhi & Tang, 2016) therefore BoP literature offer an opportunity to identify these constraints that lead to supply chain risk and design supply chains under these circumstances. BoP research emphasizes mainly on case-based empirical studies or conceptual work (Kolk et al., 2014) but there is a need for more survey-based research in BoP context to identify key supply chain risks and their impact on BoP performance.

The present study has some limitations. It might appear that the selection of articles in this literature review is very selective. The choice of articles to be included was aligned with the aim of this paper. The intention of this paper was to acquaint the reader with the key constructs that

are adopted from institutional voids, supply risk and risk management research to analyze their implication in BoPs and their performance. Not all the available BoP papers were aligned with the subject matter and did not correspond with the aim of this study. With regard supply chain risk and risk management constructs that prevail due to the presence of institutional voids, we are confident on our selection of four widely cited papers that were appropriate for an in-depth analysis of BoP literature. The selection of constructs invites future survey-based empirical research for a broader comparison of key supply chain risks influencing BoP firms. We have used well-cited frameworks that have addressed the issue of the validity of the constructs and also the involvement of multiple researchers in the analyzing results and compiling findings. We do acknowledge that the reliability of the study may be limited because of single-researcher approach in the coding process.

CONCLUSION

BoP is a fruitful and emerging context for research for both BoP markets (BoP 1.0) and BoP firms (BoP 2.0). The shift in BoP research from demand to supply side and taking BoP as potential producers (Karnani, 2007a), warrant more attention by the researchers to customize business practices that may serve the need of BoP firms (referred to as BoC) including supply chain dynamics. Supply chain risks have received wider attention in the context of developing countries but in the context of BoPs, this area is still under-researched.

We integrate three research streams i.e., institutional voids, supply chain risks and risk management and firm performance in the context of BoP literature to identify the gaps (voids) and key supply chain risks faced by BoPs and BoCs and risk management practices adopted that effect firm performance. Our frequency and contingency analysis suggest that supply chain risk and risk management is implicitly discussed in BoP literature but this aspect is widely overlooked in mainstream BoP research. We also concluded that the nature of institutional voids and risks in supply chain vary on the basis of BoP perspective (demand side or supply side). The insights from the current study can be used as a foundation to advance in supply chain risk management in BoP research.

BIBLIOGRAPHY

Adegbile, A., & Sarpong, D. (2018). Disruptive innovation at the base-of-the-pyramid: Opportunities, and challenges for multinationals in African emerging markets. *critical perspectives on international business, 14*(2/3), 111-138.

Afshar, Z. A., & Fazli, S. (2018). Investigating the effect of relational capital on supply chain risk and performance. *International Journal of Manufacturing Technology and Management, 32*(6), 517-532.

Agarwal, N., Brem, A., & Grottke, M. (2018). Towards a higher socio-economic impact through shared understanding of product requirements in emerging markets: The case of the Indian healthcare innovations. *Technological Forecasting and Social Change, 135*, 91-98.

Agarwal, N., Grottke, M., Mishra, S., & Brem, A. (2016). A systematic literature review of constraint-based innovations: State of the art and future perspectives. *IEEE Transactions on Engineering Management, 64*(1), 3-15.

Amankwah-Amoah, J., Chen, X., Wang, X., Khan, Z., & Chen, J. (2019). Overcoming institutional voids as a pathway to becoming ambidextrous: The case of China's Sichuan Telecom. *Long Range Planning*.

Ansari, S., Munir, K., & Gregg, T. (2012). Impact at the 'bottom of the pyramid': The role of social capital in capability development and community empowerment. *Journal of Management Studies, 49*(4), 813-842.

Armstrong, K., Ahsan, M., & Sundaramurthy, C. (2018). Microfinance ecosystem: How connectors, interactors, and institutionalizers co-create value. *Business Horizons, 61*(1), 147-155.

Arnold, D. G., & Valentin, A. (2013). Corporate social responsibility at the base of the pyramid. *Journal of Business Research, 66*(10), 1904-1914.

Arnold, D. G., & Williams, L. (2012). The paradox at the base of the pyramid: Environmental sustainability and market-based poverty alleviation. *International Journal of Technology Management, 60*(1/2), 44-59.

Ausrød, V. L. (2018). It takes two to tango: mobilizing strategic, ordinary, and weak resources at the base of the pyramid. *Journal of Strategic Marketing, 26*(8), 665-687.

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics, 7*(1), 22-53.

Banerjee, A. V., & Duflo, E. (2007). The economic lives of the poor. *Journal of economic perspectives, 21*(1), 141-168.

Bardy, R., Drew, S., & Kennedy, T. F. (2012). Foreign investment and ethics: How to contribute to social responsibility by doing business in less-developed countries. *Journal of Business Ethics, 106*(3), 267-282.

Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. *International Journal of Production Research, 57*(7), 2179-2202.

Bendul, J. C., Rosca, E., & Pivovarova, D. (2017). Sustainable supply chain models for base of the pyramid. *Journal of Cleaner Production, 162*, S107-S120.

Borchardt, M., Pereira, G., Viegas, C., Reolon, D., Xavier, Y., & Battaglia, D. (2018). Providing goods to the base of the pyramid: Opportunities for micro, small and medium-sized local producers. *Journal of Small Business Strategy, 28*(2), 80-89.

Bruton, G. D., Ahlstrom, D., & Si, S. (2015). Entrepreneurship, poverty, and Asia: Moving beyond subsistence entrepreneurship. *Asia Pacific Journal of Management, 32*(1), 1-22.

Bucher, S., Jäger, U., & Prado, A. M. (2016). Scaling private health care for the base of the pyramid: Expanding versus broadening service offerings in developing nations. *Journal of Business Research, 69*(2), 736-750.

Cavinato, J. L. (2004). Supply chain logistics risks: from the back room to the board room. *International Journal of Physical Distribution & Logistics Management, 34*(5), 383-387.

Chakrabarty, S., & Bass, A. E. (2015). Comparing virtue, consequentialist, and deontological ethics-based corporate social responsibility: Mitigating microfinance risk in institutional voids. *Journal of Business Ethics, 126*(3), 487-512.

Chen, H. L. (2018). Supply chain risk's impact on corporate financial performance. *International journal of operations & production management, 38*(3), 713-731.

Chopra, S., & Sodhi, M. (2004). Supply-chain breakdown. *MIT Sloan management review, 46*(1), 53-61.

Chowdhury, M. S. (2007). Overcoming entrepreneurship development constraints: the case of Bangladesh. *Journal of Enterprising Communities: People and Places in the Global Economy, 1*(3), 240-251.

Chu, M. (2007). Commercial Returns at the Base of the Pyramid. *Innovations: Technology, Governance, Globalization, 2*(1-2), 115-146.

Davarzani, H., Zanjirani Farahani, R., & Rahmandad, H. (2015). Understanding econo-political risks: impact of sanctions on an automotive supply chain. *International journal of operations & production management, 35*(11), 1567-1591.

Davies, I. A., & Torrents, A. (2017). Overcoming Institutional Voids in Subsistence Marketplaces: A Zimbabwean Entrepreneurial Case. *Journal of Macromarketing, 37*(3), 255-267.

Demssie, Y. N., Wesselink, R., Biemans, H. J., & Mulder, M. (2019). Think outside the European box: Identifying sustainability competencies for a base of the pyramid context. *Journal of Cleaner Production, 221*, 828-838.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review.

Enyinda, C. I., Mbah, C. H., & Ogbuehi, A. (2010). An empirical analysis of risk mitigation in the pharmaceutical industry supply chain: A developing-country perspective. *Thunderbird International Business Review, 52*(1), 45-54.

Fan, Y., & Stevenson, M. (2018). A review of supply chain risk management: definition, theory, and research agenda. *International Journal of Physical Distribution & Logistics Management, 48*(3), 205-230.

Fawcett, S. E., & Waller, M. A. (2015). Designing the supply chain for success at the bottom of the pyramid. *Journal of Business logistics, 36*(3), 233-239.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*: John Wiley & Sons.

Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online information review, 32*(1), 8-21.

Gold, S., Hahn, R., & Seuring, S. (2013). Sustainable supply chain management in "Base of the Pyramid" food projects—A path to triple bottom line approaches for multinationals? *International Business Review, 22*(5), 784-799.

Gold, S., Seuring, S., & Beske, P. (2010). Sustainable supply chain management and inter-organizational resources: a literature review. *Corporate Social Responsibility and Environmental Management, 17*(4), 230-245.

Goyal, S., Sergi, B. S., & Jaiswal, M. P. (2016). Understanding the challenges and strategic actions of social entrepreneurship at base of the pyramid. *Management Decision, 54*(2), 418-440.

Grieco, C., Michelini, L., & Iasevoli, G. (2015). Measuring value creation in social enterprises: A cluster analysis of social impact assessment models. *Nonprofit and voluntary sector quarterly, 44*(6), 1173-1193.

Gupta, A. D. (2019). Some Recent Cases in the Bottom of the Pyramid Concept: Lessons from India. *Global Business Review*, 0972150919829279.

Gupta, S. (2019). Understanding the feasibility and value of grassroots innovation. *Journal of the Academy of marketing Science*, 1-25.

Gutiérrez, R., & Vernis, A. (2016). Innovations to serve low-income citizens: When corporations leave their comfort zones. *Long Range Planning, 49*(3), 283-297.

Hahn, R., & Gold, S. (2014). Resources and governance in "base of the pyramid"-partnerships: Assessing collaborations between businesses and non-business actors. *Journal of Business Research, 67*(7), 1321-1333.

Hall, J., Matos, S. V., & Martin, M. J. (2014). Innovation pathways at the Base of the Pyramid: Establishing technological legitimacy through social attributes. *Technovation, 34*(5-6), 284-294.

Handfield, R., & Nichols Jr, E. (1999). Introduction to. *Supply Chain Management, Prentice Hall, Englewood Cliffs, NJ*.

Hart, S. L. (2017). Prologue: defining the path towards a BoP 3.0 *Base of the Pyramid 3.0* (pp. 9-12): Routledge.

Hendricks, K. B., & Singhal, V. R. (2005). An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and operations management, 14*(1), 35-52.

Hens, L. (2012). Overcoming institutional distance: Expansion to base-of-the-pyramid markets. *Journal of Business Research, 65*(12), 1692-1699.

Ho, W., Zheng, T., Yildiz, H., & Talluri, S. (2015). Supply chain risk management: a literature review. *International Journal of Production Research, 53*(16), 5031-5069.

Jagtap, S. (2019). Key guidelines for designing integrated solutions to support development of marginalised societies. *Journal of Cleaner Production*.

Jose, S., & Robert Buchanan, F. (2013). Marketing at the bottom of the pyramid: service quality sensitivity of captive microfinance borrowers. *Journal of Consumer Marketing, 30*(7), 573-582.

Jüttner, U., Peck, H., & Christopher, M. (2003). Supply chain risk management: outlining an agenda for future research. *International Journal of Logistics: Research and Applications, 6*(4), 197-210.

Kaplan, R., & Norton, D. (1999). The Balanced Scorecard for Public-Sector Organizations. Balanced Scorecard Report 1999. Boston: Harvard Business School Publishing, 20 [consultado el 15 de diciembre de 2015].

Karnani, A. (2007a). Doing well by doing good—case study:'Fair & Lovely'whitening cream. *Strategic management journal, 28*(13), 1351-1357.

Karnani, A. (2007b). The mirage of marketing to the bottom of the pyramid: How the private sector can help alleviate poverty. *California Management Review, 49*(4), 90-111.

Khalid, R. U., Seuring, S., Beske, P., Land, A., Yawar, S. A., & Wagner, R. (2015). Putting sustainable supply chain management into base of the pyramid research. *Supply Chain Management: An International Journal, 20*(6), 681-696.

Khanna, T., & Palepu, K. (1997). Why focused strategies may be wrong for emerging markets. *Harvard Business Review, 75*(4), 41-43.

Kleindorfer, P. R., & Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and operations management, 14*(1), 53-68.

Kolk, A., Rivera-Santos, M., & Rufín, C. (2014). Reviewing a decade of research on the "base/bottom of the pyramid"(BOP) concept. *Business & Society, 53*(3), 338-377.

Kuada, J. (2016). Marketing, economic growth, and competitive strategies of firms in Africa. *African Journal of Economic and Management Studies, 7*(1), 2-8.

Kuo, T., Shiang, W.-J., Hanafi, J., & Chen, S. (2018). Co-Development of Supply Chain in the BOP Markets. *Sustainability, 10*(4), 963.

Lee, H. L. (2004). The triple-A supply chain. *Harvard Business Review, 82*(10), 102-113.

Li, W., & Abiad, V. (1990). Institutions, institutional change, and economic performance.

Lim, C., Han, S., & Ito, H. (2013). Capability building through innovation for unserved lower end mega markets. *Technovation, 33*(12), 391-404.
London, T., Anupindi, R., & Sheth, S. (2010). Creating mutual value: Lessons learned from ventures serving base of the pyramid producers. *Journal of Business Research, 63*(6), 582-594.
London, T., Esper, H., Grogan-Kaylor, A., & Kistruck, G. M. (2014). Connecting poverty to purchase in informal markets. *Strategic Entrepreneurship Journal, 8*(1), 37-55.
London, T., & Hart, S. L. (2004). Reinventing strategies for emerging markets: beyond the transnational model. *Journal of international business studies, 35*(5), 350-370.
Mair, J., & Marti, I. (2009). Entrepreneurship in and around institutional voids: A case study from Bangladesh. *Journal of business Venturing, 24*(5), 419-435.
Mair, J., Marti, I., & Ventresca, M. J. (2012). Building inclusive markets in rural Bangladesh: How intermediaries work institutional voids. *Academy of Management journal, 55*(4), 819-850.
Mamabolo, A., & Myres, K. (2019). Performance Measurement in Emerging Market Social Enterprises using a Balanced Scorecard. *Journal of Social Entrepreneurship*, 1-23.
Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management, 38*(3), 192-223.
March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management science, 33*(11), 1404-1418.
Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business logistics, 22*(2), 1-25.
Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics, 106*(1), 213-228.
Monzer, D. A., Rebs, T., Khalid, R. U., & Brandenburg, M. (2018). Sustainable Supply Chain Management at the Base of Pyramid: A Literature Review *Social and Environmental Dimensions of Organizations and Supply Chains* (pp. 235-257): Springer.
Moradeyo, A. A. (2018). Rudimentary causes and impacts of supply chain risks in Sub-Saharan Africa. *Journal of Operations and Supply Chain Management, 11*(2), 16-31.
Mulrow, C. D. (1994). Systematic reviews: rationale for systematic reviews. *Bmj, 309*(6954), 597-599.
Nakata, C., & Antalis, E. (2015). Enhancing market exchanges at the base of the pyramid: The roles of national culture, nontraditional capitals, and transformative technologies. *International marketing review, 32*(6), 762-782.
Neiger, D., Rotaru, K., & Churilov, L. (2009). Supply chain risk identification with value-focused process engineering. *Journal of operations management, 27*(2), 154-168.
Onsongo, E. (2019). Institutional entrepreneurship and social innovation at the base of the pyramid: the case of M-Pesa in Kenya. *Industry and Innovation, 26*(4), 369-390.
Pansera, M., & Martinez, F. (2017). Innovation for development and poverty reduction: an integrative literature review. *Journal of Management Development, 36*(1), 2-13.
Paras, M. K., Ekwall, D., & Pal, R. (2019). Developing a framework for the performance evaluation of sorting and grading firms of used clothing. *Journal of Global Operations and Strategic Sourcing, 12*(1), 82-102.
Parmigiani, A., & Rivera-Santos, M. (2015). Sourcing for the base of the pyramid: Constructing supply chains to address voids in subsistence markets. *Journal of operations management, 33*, 60-70.
Peck, H. (2006). Reconciling supply chain vulnerability, risk and supply chain management. *International Journal of Logistics: Research and Applications, 9*(2), 127-142.
Perrot, F. (2017). Multinational Corporations' Strategies at the Base of the Pyramid: An Action Research Inquiry. *Journal of Business Ethics, 146*(1), 59-76.

Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *Bmj, 322*(7278), 98-101.

Prado, A. M., Calderon, D., & Zúñiga, R. (2016). Providing low-cost and high-quality medications to rural communities in developing countries: The case of Accion Medica Cristiana in Nicaragua. *Journal of Business Research, 69*(9), 3910-3922.

Prahalad, C. K., & Hammond, A. (2002). Serving the world's poor, profitably. *Harvard Business Review, 80*(9), 48-59.

Rabino, S. (2015). The bottom of the pyramid: an integrative approach. *International Journal of Emerging Markets, 10*(1), 2-15.

Ramachandran, J., Pant, A., & Pani, S. K. (2012). Building the BoP Producer Ecosystem: The Evolving Engagement of F abindia with I ndian Handloom Artisans. *Journal of product innovation management, 29*(1), 33-51.

Ramani, S. V., & Mukherjee, V. (2014). Can breakthrough innovations serve the poor (bop) and create reputational (CSR) value? Indian case studies. *Technovation, 34*(5-6), 295-305.

Rangel, D. A., de Oliveira, T. K., & Leite, M. S. A. (2015). Supply chain risk classification: discussion and proposal. *International Journal of Production Research, 53*(22), 6868-6887.

Rao, S., & Goldsby, T. J. (2009). Supply chain risks: a review and typology. *The international journal of logistics management, 20*(1), 97-123.

Rao, S., Nilakantan, R., Iyengar, D., & Lee, K. B. (2018). On the Viability of Fixing Leaky Supply Chains for the Poor Through Benefit Transfers: A Call for Joint Distribution. *Journal of Business logistics*.

Rea, L. M., & Parker, R. A. (1992). Designing and conducting survey research San Francisco. *CA: Jossey-Boss*.

Reficco, E., & Gutiérrez, R. (2016). Organizational ambidexterity and the elusive quest for successful implementation of BoP ventures. *Organization & Environment, 29*(4), 461-485.

Reiner, G., Gold, S., & Hahn, R. (2015). Wealth and health at the Base of the Pyramid: Modelling trade-offs and complementarities for fast moving dairy product case. *International journal of production economics, 170*, 413-421.

Revilla, E., & Saenz, M. J. (2017). The impact of risk management on the frequency of supply chain disruptions: A configurational approach. *International journal of operations & production management, 37*(5), 557-576.

Reynoso, J., Kandampully, J., Fan, X., & Paulose, H. (2015). Learning from socially driven service innovation in emerging economies. *Journal of Service Management, 26*(1), 156-176.

Ritchie, B., & Marshall, D. V. (1993). *Business risk management*: Chapman & Hall.

Rivera-Santos, M., Rufin, C., & Kolk, A. (2012). Bridging the institutional divide: Partnerships in subsistence markets. *Journal of Business Research, 65*(12), 1721-1727.

Rosca, E., & Bendul, J. C. (2019). Value chain integration of base of the pyramid consumers: An empirical study of drivers and performance outcomes. *International Business Review, 28*(1), 162-176.

Roy, V., & Singh, S. (2017). Mapping the business focus in sustainable production and consumption literature: Review and research framework. *Journal of Cleaner Production, 150*, 224-236.

SCC, S. C. C. (2010). Supply chain operations reference model SCOR version 10.0. *The supply chain council, Inc. SCOR: the supply chain reference ISBN 0-615-20259-4 (binder)*.

Schaefers, T., Moser, R., & Narayanamurthy, G. (2018). Access-based services for the base of the pyramid. *Journal of Service Research, 21*(4), 421-437.

Schrader, C., Freimann, J., & Seuring, S. (2012). Business strategy at the base of the pyramid. *Business Strategy and the environment, 21*(5), 281-298.

Schuster, T., & Holtbrügge, D. (2012). Market entry of multinational companies in markets at the bottom of the pyramid: A learning perspective. *International Business Review, 21*(5), 817-830.

Seelos, C., & Mair, J. (2007). Profitable business models and market creation in the context of deep poverty: A strategic view. *Academy of Management Perspectives, 21*(4), 49-63.

Seuring, S., & Khalid, R. (2019). Analyzing Base-of-the-Pyramid Research from a Supply Chain Perspective.

Shou, Y., Hu, W., Kang, M., Li, Y., & Park, Y. W. (2018). Risk management and firm performance: the moderating role of supplier integration. *Industrial Management & Data Systems, 118*(7), 1327-1344.

Simanis, E., & Duke, D. (2014). Profits at the bottom of the pyramid. *Harvard Business Review, 92*(10), 86-93.

Simanis, E., Hart, S., & Duke, D. (2008). The base of the pyramid protocol: Beyond "basic needs" business strategies. *Innovations: Technology, Governance, Globalization, 3*(1), 57-84.

Siqueira, A. C. O., Mariano, S. R., & Moraes, J. (2014). Supporting innovation ecosystems with microfinance: Evidence from Brazil and implications for social entrepreneurship. *Journal of Social Entrepreneurship, 5*(3), 318-338.

Sodhi, M., & Lee, S. (2007). An analysis of sources of risk in the consumer electronics industry. *Journal of the Operational Research Society, 58*(11), 1430-1439.

Sodhi, M. S., Son, B. G., & Tang, C. S. (2012). Researchers' perspectives on supply chain risk management. *Production and operations management, 21*(1), 1-13.

Sodhi, M. S., & Tang, C. S. (2016). Supply chain opportunities at the bottom of the pyramid. *Decision, 43*(2), 125-134.

Subrahmanyan, S., & Tomas Gomez-Arias, J. (2008). Integrated approach to understanding consumer behavior at bottom of pyramid. *Journal of Consumer Marketing, 25*(7), 402-412.

Tang, C., & Tomlin, B. (2008). The power of flexibility for mitigating supply chain risks. *International journal of production economics, 116*(1), 12-27.

Tate, W. L., Bals, L., & Marshall, D. (2019). Supply chain management at the base of the pyramid. *International Journal of Physical Distribution & Logistics Management, 49*(5), 438-450.

Thun, E. (2018). Innovation at the middle of the pyramid: State policy, market segmentation, and the Chinese automotive sector. *Technovation, 70*, 7-19.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management, 14*(3), 207-222.

Tukamuhabwa, B., Stevenson, M., & Busby, J. (2017). Supply chain resilience in a developing country context: a case study on the interconnectedness of threats, strategies and outcomes. *Supply Chain Management: An International Journal, 22*(6), 486-505.

Viswanathan, M., & Rosa, J. A. (2010). Understanding subsistence marketplaces: Toward sustainable consumption and commerce for a better world. *Journal of Business Research, 63*(6), 535-537.

Viswanathan, M., Seth, A., Gau, R., & Chaturvedi, A. (2007). *DOING WELL BY DOING GOOD: PURSUING COMMERCIAL SUCCESS BY INTERNALIZING SOCIAL GOOD IN SUBSISTENCE MARKETS.* Paper presented at the Academy of Management Proceedings.

Wagner, S. M., & Bode, C. (2008). An empirical examination of supply chain performance along several dimensions of risk. *Journal of Business logistics, 29*(1), 307-325.

Webb, J. W., Kistruck, G. M., Ireland, R. D., & Ketchen, J., David J. (2010). The entrepreneurship process in base of the pyramid markets: The case of multinational

enterprise/nongovernment organization alliances. *Entrepreneurship theory and practice, 34*(3), 555-581.

Webb, J. W., Morris, M. H., & Pillay, R. (2013). Microenterprise growth at the base of the pyramid: A resource-based perspective. *Journal of Developmental Entrepreneurship, 18*(04), 1350026.

Weerawardena, J., & Mort, G. S. (2006). Investigating social entrepreneurship: A multidimensional model. *Journal of world business, 41*(1), 21-35.

Weidner, K. L., Rosa, J. A., & Viswanathan, M. (2010). Marketing to subsistence consumers: Lessons from practice. *Journal of Business Research, 63*(6), 559-569.

Wieland, A., & Marcus Wallenburg, C. (2012). Dealing with supply chain risks: Linking risk management practices and strategies to performance. *International Journal of Physical Distribution & Logistics Management, 42*(10), 887-905.

Wu, Z., & Jia, F. (2018). Toward a theory of supply chain fields–understanding the institutional process of supply chain localization. *Journal of operations management, 58*, 27-41.

Yang, C.-L., Huang, R.-H., & Lee, Y.-C. (2014). Building a performance assessment model for social enterprises-views on social value creation. *Science Journal of Business and Management, 2*(1), 1-9.

Zomorrodi, M., Fayezi, S., Lau, K. H., & McMurray, A. (2019). Supply chain adaptations for the base-of-the-pyramid business: towards a theoretical model. *International Journal of Physical Distribution & Logistics Management*.

Zsidisin, G. A. (2003). A grounded definition of supply risk. *Journal of Purchasing and Supply Management, 9*(5-6), 217-224.

DECISION SCIENCES INSTITUTE

Simultanuous New Product Development: A Just-in-Time Manufacturing Perspective

(Full Paper Submission)

Mohammad Z. Meybodi
Indiana University Kokomo
mmeybodi@iuk.edu

ABSTRACT

The focus of this article is to demonstrate potential links between Just-in-Time (JIT) practices and simultaneous new product development (SNPD). Analysis and comparison of several critical elements show high degree of resemblances between the two methods. A number of hypotheses regarding similarities between the elements of the two methods were developed and tested. Survey data from a sample of manufacturing organizations strongly supports the hypotheses regarding similarities between the two methods. Statistical results also indicate compared with conventional companies, by utilizing SNPD, JIT manufacturing organizations are able to design and develop products faster with better quality, lower development cost, higher frequency, and lower manufacturing cost.

KEY WORDS: Product Development, Just-in-Time

INTRODUCTION

In a global market, innovation and speedy new product development is crucial for companies to gain competitive advantage. Creating new product ideas that are consistent with organizational strategy and moving these ideas through the stages of design, development, testing, and deployment has been the trade mark of successful world class organizations (Jacobs and Chase, 2020; Johansson and Safsten 2015; Ferioli et al. 2010; Roulet et al. 2010). Introducing new products to the market early has several strategic and tactical advantages. (Lofstrand, 2010; Kristav, 2016; Wen et.al. 2020; Cooper and Kleinschmidt, 1994; White, 1993).
Despite its well-known strategic role, for large number of manufacturing organizations innovation, design, and successful management of new product development has often been a major challenge. Long development time, prohibitive development and manufacturing costs, and questionable quality have been the common result for many of these organizations. The primary factor contributing to such unsuccessful result is perhaps the use conventional sequential method by these organizations (Morgan and Liker, 2006). However, manufacturing literature for the past three decades clearly shows that through their JIT manufacturing practices, some world class organizations such as Toyota have dominated competition not only in the area of manufacturing but also in the area of innovation, design, development, and quick commercialization of new technologies (Marisa et al. 2008; Heinzen and Hoflinger 2017; Ulrich and Eppinger, 2004; Michael, 2008; Unger and Eppinger 2009). Instead of traditional sequential approach, world class organizations utilized simultaneous new product development (SNPD) method. The focus of this article is to understand such contrast between the two type of organizations.

LITERATURE REVIEW

JIT has been a great force in the world of manufacturing since early 1990's. Some of the main benefits of a JIT such as lower inventory, quicker delivery, and lower cost have been well documented (Cook and Rogowski, 1996; Hobbs, 1994; Temponi and Pandya, 1995; Deshpande and Golhar, 1995; Billesbach, 1991; Handfield, 1993; Lawrence and Hottenstein, 1995; Golhar, Stamm, and Smith, 1990; McKay, et.al 2011; Moras and Dieck, 1992). In the simplest form, JIT requires maximizing value added production activities by removing unnecessary wastes. Identification and elimination of waste and respectful treatment of employee are the two fundamental principles of a JIT system (Hobbs, 1994; Womack and Jones, 2003). Elimination of waste is achieved by adopting practices such as continuous quality improvement, setup time reduction, utilizing flexible resources, group technology layout, and pull production system (Gargeya, and Thompson, 1994; Sohal, Ramsay, and Samson, 1993). Respectful treatment of people often means employee empowerment; it includes elements such as team work, fair compensation, employee training and new positive attitude toward suppliers (Sohal, Ramsay, and Samson, 1993). Unfortunately, since its beginning in mid 1980's, often a narrow view of JIT has been accepted and utilized by western manufacturers. Application of JIT to reduce inventory and increase deliveries is only a small fraction of the full potential benefits of a JIT system (Blackburn, 1991; Kristav, 2016). To take advantage of the full benefits of JIT, one needs to have a much broader view of JIT principles (Blackburn, 1991). Looking at JIT as a process of eliminating waste and respectful treatment of employee, its principles can be applied to other areas including service areas such as healthcare, education, government, and new product development, (Womack and Jones (2003). Application of JIT principles to new product development has great opportunity to shorten product development time, improve design quality, and reduce product development and manufacturing costs (Anand and Kodali, 2008). The company that originated famous JIT system, TPS, also developed Toyota Product Development System (TPDS). TPDS employs JIT principles and tools such as value stream mapping, Kanban, 5S system, and continuous improvement to eliminate waste from product development activities and bring quality products to market faster than their leading competition (Morgan and Liker, 2006; Ward, 2007). However, TPDS is a comprehensive strategy that involves various approaches to eliminate waste from new product development activities. The focus of this article is on special case of TPDS. The objective of the article is to answer the following questions:

1. Are there similarities between JIT and SNPD practices?
2. Are there differences between new product development performances for JIT companies using SNPD and conventional companies?

The remainder of the article is organized in the following manner: First, an overview of the differences between conventional sequential method and recent SNPD is presented. Second, the article compares and analyzes similarities between JIT and SNPD for a number of critical elements followed by a set of test of hypotheses on similarities between the elements. Third, the article tests product design performance for conventional sequential method and SNPD method. Research methodology, results, and conclusion are the final sections of the article.

SEQUENTIAL AND SIMULTANOUS NPD

NPD process is a sequence of inter-connected activities in which information regarding customer needs is translated into final product design. In a conventional method, NPD process typically involves phases such as idea generation and validation, preliminary design, final design and prototyping, and pilot production and ramp-up (Wheelwright and Clark, 1992; Jacobs and Chase, 2020). Traditionally, this design process is managed sequentially by personnel from various functions of the organization. A major drawback of this approach is that the output from one design feedback among sequential stage causes the process to require too many design changes which

causes the process to require longer development time which indeed causes the process to be too slow, too costly, and often of poor quality. The two elements of long delay and design changes during the delay creates a never-ending cycle where time delay causes design change and to accommodate design change it needs more time. The final result is that the designs are often rejected because the design is either outdated due to long development time or it is infeasible in term of manufacturing capability (Yamamoto and Abu Qudeiri, 2010; Blackburn, 1991; Ulrich and Eppinger, 2000).

Unlike conventional approach to NPD where functional units work sequentially and downstream functions are not involved until late in the process, SNPD requires early involvement of cross functional teams. It requires that designers, manufacturers, marketers, suppliers, and customers work jointly to design product and manufacturing process in parallel. The design team must truly understand the concept of simultaneous engineering in which activities of product and process design are performed in parallel and in a coordinated manner The objective is to integrate product design and process planning into a common activity (Albers and Braun, 2011; Liang, 2009; Anderson, 2008; Donnellon, 1993; Millson, Ranj, and Wilemon, 1992; Shunk, 1992). Application of SNPD under various manufacturing environments in order to shorten development time, improve quality, reduce risks, and reduce development cost is reported by these researchers (Anderson, 2008; Skalak, 2002; Kowang and Rasli, 2011; Lofstrand, 2010; Moges, 2009). Due to early cross-functional communication, SNPD approach enables an organization to be more innovative in terms of improving design quality, shortening development time, reducing design risks, and reducing development and manufacturing costs (Lynch et.al, 2016; Blackburn, 1991; Ulrich, and Eppinger, 2000; Arora and Mital, 2012; Katzy et.al, 2012; Zirger and Hartley, 1996).

COMPARISON OF JIT MANUFACTURING AND SNPD ELEMENTS

For the past three decades, there has been an extensive volume of research in the area of JIT. As a result, there is a set of generally accepted guidelines that organizations can follow to achieve manufacturing success. However, there has been limited research on the application of JIT principles to NPD and there is no comparable set of guidelines for successful management of NPD process. Recently, a number of world class companies have attempted to apply the principles of JIT manufacturing to NPD activities (De Waal and Knott, 2019). The company that started the most famous JIT system, Toyota Production System (TPS) is also started Toyota Product Development System (TPDS). TPDS employs JIT principles and enable the company to bring the highest quality products to market faster than their leading competition. Also, a number research on the application of JIT principles to NPD process has shown that achieving certain manufacturing process improvement such as reducing variation, reducing rework and yield loss, solving process bottlenecks, and managing capacity, can significantly reduce NPD times.

Similarities between JIT and SNPD for a number of critical elements are shown in Table 1, (Goffin et. al, 2019; Spencer and Guide, 1995). Following is a brief comparison and analysis of selected elements in Table 1:

Layout

Layout in JIT environment is often in the form of product focus and manufacturing cells. This type of layout is necessary because small lot size production requires that the layout to be compact and efficient to ensure smooth flow of materials and close communication between work stations. Unlike conventional manufacturing, where material is pushed

forward, the flow in a JIT environment is in two directions; material is pulled forward, but information flows backward to provide feedback on performance and material requirements.

SNPD, overlapping of a large number of activities requires a complete change in layout that facilitates communication and encourages team work. Instead of organizing by sequential functions, SNPD emphasizes on cross-functional integration and the formation of a design team. The design team sits together in one location, creating a type of project layout. A project layout creates an environment for frequent, two-way communication between team members, which encourages concurrent development of a product and its associated processes.

Lot Size

In contrast to conventional manufacturing, JIT manufacturing requires production of small lot-sizes. Production of small lot-sizes is possible by drastically reducing set-up times. It is well documented that production of small lot-sizes in JIT is closely associated with improved quality, reduced inventory, faster delivery, and more responsive to market demands.

Similar to JIT, SNPD also utilizes small lot-sizes; the only difference is that in JIT small lot sizes of goods are processed but SNPD requires small lot-sizes of information. That is, continuous two way communication in SNPD is similar to early release of small batches of information (Blackburn, 1991; White, 1993). With the early release of small batches of information, downstream constituents can begin working on different phases of the design while final design is evolving. The early release of information reduces uncertainty and encourages early detection of problems, which enables organizations to avoid costly, time-consuming changes.

Employee and Supplier Involvement

In JIT environment, management encourages employee involvement and team work. The responsibility for job scheduling and quality are often passed to the teams at the shop floor. Due to small lot size production, delegation of authority to the teams at the shop floor is essential for smooth production flow. Also, in JIT suppliers work closely with manufacturing organization to improve quality and shorten delivery time.

Similar to JIT, in SNPD the responsibility for scheduling of the activities pushed down to product development team at the lowest level. Passing responsibility down to the team is essential to achieve a high level of activity coordination and information sharing among team members. Also, in SNPD suppliers work closely with the design team to reduce development costs, shorten development time, and offer ideas toward improving the quality of the design.

Quality

In JIT and SNPD environments, organizations are often proactive and quality means getting it right the first time. In JIT, since batch sizes are small quality at the source and continuous quality improvement are the main foundations. Shop floor workers are empowered to become their own inspectors responsible for the quality of their output. In SNPD, because of the teamwork and two-way flow of information between team members, and utilization of quality improvement tool such as six sigma process quality problems are detected earlier and solved before they have a cumulative impact on the rest of the project (Chakravorty and Franza, 2009).

Technology

In a JIT system, technology is not viewed as a substitute, or shortcut to process improvement. Rather, technology has been utilized after process analysis and simplification has been performed. The role of technology in SNPD is also enormous; it requires that the design team with diverse expertise makes a large number of interrelated decisions regarding the form, fit, function, cost, quality, and other aspects of the design (Karagozoglu and Brown, 1993). This requires supply and processing of relevant information from multiple sources in a coordinated

manner. Successful organizations use technology in their NPD process similarly to the way they use technology in their JIT system. In SNPD, the design team utilizes appropriate technologies and tools at various stages of NPD process. Effective use of technologies and tools can dramatically shorten NPD time, reduce number of prototypes, cut costs, and improve quality of the design (McKay et al. 2011; Yamamoto and Abu Qudiri 2011; Roulet et.al (2010). The key to the success of technology in SNPD is building an effective design team with open cross-functional communication lines.

FACTOR HYPOTHESES

Comparison and analysis of elements in Table 1 show a high degree of similarities between JIT and SNPD. To study further, a set of twenty five hypotheses (H1-H25) that statistically test similarities between JIT and SNPD will be presented. The hypotheses are shown in Table 2. Each hypothesis in Table 2 consists of two parts- a and b. In part a, the test is conducted for JIT elements and the corresponding test for SNPD elements is conducted in part b. The last hypothesis examines the overall impact of JIT principles on SNPD.

Hypotheses (H1-H25):

There is a high degree of similarities between JIT manufacturing and SNPD elements.

PRDUCT DEVELOPMENT PERFORMANCES

The following dimensions of quality, time, competency, development cost, and manufacturing cost are used to measure the performance of NPD (Ulrich and Eppinger, 2000; Wheelwright and Clark, 1992):

Quality: Quality is ultimately reflected in the price customers are willing to pay, the market share, and the bottom line profit. In NPD, quality problems are often the results of incomplete information and miscommunication among various functions. Quality often means a minimal number of redesign or rework. In this article, number of design changes during the development process and early manufacturing phase is used as a measure of design quality.

Development time: Development time is the length of time between initial idea generation until new product is ready for introduction to the market. Shorter development time raises the competitive value of new product in terms of premium price, larger market share, and higher profit margin.

Development competency: Development competency is the ability of the organization to develop future products better, faster, and cheaper. Competent workforce and effective use of technologies are important elements of organizational NPD competency. Frequency of new product introduction to the market is used as a measure of development competency.

Development cost: This is the total cost from the early idea generation until the product is ready for manufacturing. For most organizations, development cost is usually a significant portion of the budget and must be considered in light of budget realities and the timing of budget allocations.

Manufacturing cost: Manufacturing cost includes initial investment on equipments and tools as well as the incremental cost of manufacturing the product. There is a close relationship between

manufacturing cost and the type of decisions made during the early design stage. Although early design decisions determine about 70 percent of future manufacturing cost, organizations often spend far too little time and resources during this stage (Huthwaite, B. 1991). To save future manufacturing cost, it is prudent for the companies to spend more time and resources during the early design phases of NPD process where critical design decisions are made.

PERFORMANCE HYPOTHESES

In the second set of hypotheses (H26-H30), the differences between NPD performances for JIT manufacturing companies and conventional companies are tested.

Hypotheses (H26-H30):

H26: By utilizing SNPD approach, JIT companies are able to design new products with fewer design changes than conventional companies(better quality).

H27: By utilizing SNPD approach, JIT companies are able to design new products faster than conventional companies.

H28: By utilizing SNPD approach, JIT companies are able to design new products more often than conventional companies.

H29: By utilizing SNPD approach, JIT companies are able to design new products with less development cost than conventional companies.

H30: By utilizing SNPD approach, JIT companies are able to design new products with less manufacturing cost than conventional companies.

RESEARCH METHODOLOGY

The target population for this study consisted of manufacturing firms in the states of Illinois, Indiana, Ohio, Michigan, and Wisconsin. A sample of manufacturing firms with more than 50 employees was chosen from manufacturers' directories of those states. The sample covers organizations in variety of industries ranging from fabricated metal, communication, electronics, automotive, toots, chemicals, rubber, and paper products. A comprehensive survey instrument based on examination of the literature and critical elements listed in Table 1 was developed. A panel of practitioners and researchers with experience in JIT and NPD was used to validate the survey. In addition to general organization and managerial profile items, the survey contained 50 items (25 paired) regarding similarities between JIT and SNPD elements. The twenty five paired questionnaire items are shown in Table 2. Also, the survey instrument contained a number of questionnaire items on NPD performances for JIT companies using SNPD and conventional companies. Out of 91 completed surveys received, 84 surveys were usable resulting in a response rate of 17%. Based on a number of questionnaire items on the principles of JIT practices, 33 organizations were grouped as JIT companies and 51 organizations were categorized as conventional companies.The survey data indicates that majority of respondents had various high level managerial positions from organization with less than 500 employees. Presidents and vice presidents accounted for 29% and plant managers accounted for 30% of the sample. About 35% of the sample had other managerial positions such as production managers, quality managers, and the remaining 6% were production line supervisors. In terms of manufacturing and NPD experience, about 28% of the respondents had between 10 to 20 years and 60% had more than 20 years of manufacturing experience. About 72% of the sample

had more than 10 years of JIT experience and close to 65% of the sample had more than 10 years of NPD experience.

RESEARCH RESULTS

As stated earlier, in the first set of hypotheses the objective was to examine similarities between JIT and SNPD for a set of paired elements shown in Table 2. For each item, the null hypothesis was that the mean response for JIT is equal to the mean response for SNPD. The differences between the mean responses for JIT and SNPD were compared using the statistical t-test. The respondents were asked to rate each element of Table 2 based on the degree of their agreement (1=strongly disagree, 5=strongly agree) to the question. Table 3 shows the result of similarities between JIT and SNPD.

As shown in Table 3, overall the respondents strongly agreed with the statements regarding similarities between JIT and SNPD elements. The mean ratings for about 70% of the elements for both JIT and SNPD are above 3.80. Specifically, out of twenty hypotheses, the respondents agreed that there is a high degree of similarities between JIT and SNPD for all except three hypotheses H4, H7, and H9.

For H4, the mean ratings for JIT and SNPD are respectively 4.34 and 3.81. This means although the respondents understood that short set-up and fast transition time are the main requirements of successful JIT and SNPD, the relationship between short set-up and JIT was much stronger. This This is a reasonable result because an average manufacturing manager has longer experience with JIT than SNPD. They clearly understood that successful JIT requires small lot-size and small lot-size requires short set-up time. However, due to their shorter experience with SNPD and because SNPD is primarily an information processing process, the links between small batches of information and fast transition time is not clear.

H7 hypothesizes the relationships between small lot-sizes and quality improvement for both JIT and SNPD. For this test, the mean ratings for JIT and SNPD are respectively 3.43 and 3.89. This indicates for an average manager it is easier to recognize the relationship between SNPD and quality improvement than the relationship between JIT and quality improvement. The higher rating for SNPD is perhaps due to continuous and two way communication among design team members, which encourages early detection of the design problem. The JIT result is also consistent with the literature because although total quality management and quality improvement are fundamental requirements of successful JIT, an average manufacturing manager has difficulty to understand this relationship. The relationships between small lot-size and reduced manufacturing cost in JIT and the relationship between small batches of information and reduced development cost in SNPD are examined in H9. The mean ratings for JIT and SNPD are respectively 3.58 and 3.94. For the same reasons as H7, this means for an average manager it is easier to understand this relationship in SNPD than JIT. The JIT result is interesting and also consistent with the literature because reduced manufacturing cost in JIT is primarily due to elimination of wastes, a fundamental principle of JIT, and an average manufacturing manager has difficulty to see this relationship. The overall impact of JIT principles on JIT and SNPD is examined in H24. It is obvious that the data supports the hypothesis as the mean ratings for JIT and SNPD are respectively 4.56 and 4.29 indicating strong agreement with the statements that the main principles of waste elimination and respectful treatment of people in JIT can also be applied in SNPD.

The last column of Table 3 shows correlation coefficients between JIT and corresponding SNPD elements. The correlation coefficients in Table 3 strongly support the above analysis. With the exception of three hypotheses H4, H7, and H9 other coefficients are greater than 0.60 indicating a high degree of linear association between JIT and SNPD elements.

The performance hypotheses (H26-H30) state that by utilizing SNPD approach, JIT companies are able to design new products with fewer design changes, faster, more often, with less development cost, and less manufacturing cost than conventional companies.

Table 4 provides useful statistical information regarding NPD performances for JIT and conventional companies. The average number of design changes for conventional and JIT companies are respectively 5.36 and 3.28, a quality improvement of 63%. The average development time for conventional and JIT companies are respectively 37.52 and 24.73 months, an improvement of 52%. For development competency, the average time between introduction of new products for conventional companies is 49.46 months and 32.72 months for JIT companies, an improvement of 51%. Table IV also indicates that JIT organizations enjoy a 45% reduction in NPD cost and 36% reduction in manufacturing cost. From the last column of Table 4, it is clear that the hypotheses are strongly supported by the data as the p-value for all five hypotheses is less than 0.005.

CONCLUSION

The focus of this article was to demonstrate possible links between JIT practices and SNPD. First, comparison and analysis of a number of elements showed remarkable similarities between JIT practices and SNPD. Second, a set of paired hypotheses was used to test similarities between JIT practices and SNPD elements. Statistical results clearly support the hypotheses regarding similarities between JIT and SNPD for majority of elements. Specifically, out of twenty four hypotheses, the respondents agreed that there is a high degree of similarities between JIT and SNPD for all but three hypotheses. The last pair of hypotheses that examines the overall impact of JIT principles is especially important. Statistical results strongly agreed that the main principles of waste elimination and respectful treatment of people in JIT is also applicable to SNPD. The correlation coefficients between JIT and SNPD elements also supported the same result. Third, statistical results also indicate that compared with conventional companies, JIT companies are able to develop new products with 63% better quality, 52% less development time, 45% less development cost, and 36% less manufacturing cost. Also frequency of new product introduction is 51% faster than conventional companies.

In summary, statistical results of this article clearly show strong links between JIT practices and SNPD. Managerial implication of the research is that successful implementation of JIT principles goes much beyond inventory reduction and frequent deliveries. For JIT organizations success in SNPD is the result of knowledge and technology transfer from their JIT system into their NPD process.

(Tables and references are available from the author upon request)

DECISION SCIENCES INSTITUTE
Capitalising Product Associations in a Supermarket Retail Environment

Michelle Cheong Lee Fong
Singapore Management University
Email: michcheong@smu.edu.sg

Chia Yong Qing
Singapore Management University
Email: yqchia.2017@mitb.smu.edu.sg

ABSTRACT

This paper explores methods to capitalise on retail companies' transactional databases, to mine meaningful product associations, and to design product placement strategies as a means to drive sales. We implemented three in-store initiatives based on our hypotheses – placing products with high associations together will induce an increase in sales of consequent; introducing an antecedent that is new to store will bring about a similar impact on sales of consequent based on established product association rules uncovered from other stores. Sales tracking over twelve weeks revealed that there were improvements in sales of consequents across all three initiatives performed in-store.

KEYWORDS: Association Mining, Market Basket Analysis, Shopper Marketing

INTRODUCTION

Market surveys revealed that the global online shopping market size is expected to hit 4 trillion by 2020 (Clement, 2020). Amidst heightened competition from e-commerce players, it is a challenge for traditional retailers to retain customer's patronage and market share. Increasingly, brick-and-mortar companies are turning to shopper marketing for innovative practices to market to their customers in-stores. (Kopalle, et al., 2009). A key factor exacerbating this trend is the availability and use of data by retailers to understand customer's profiles, their motivations for purchase, as well as the optimal channels to reach out to them.

This paper attempts to discover new ways to leverage consumers' purchasing patterns to trigger successful purchases at the point of sales. To do so, we partnered with a corporate sponsor, NTUC FairPrice, which is the largest retailer in Singapore with more than 50% of the market share in the physical retail segment. Despite being the market leader in its domain, FairPrice's supermarket business is not spared from disruption by online contenders such as Lazada's RedMart and Amazon Prime Now. With over 100,000 customer touch points daily, it is of FairPrice's interest to reinvent ways to tap on its rich transaction database to generate additional sales and value for its customers.

LITERATURE REVIEW

Shopper Marketing

Shopper marketing is defined as the act of planning and execution of marketing activities that influence a consumer along his shopping process (Shankar, 2011). There have been continuous developments in the field of shopper marketing due to a shift in environmental

factors that results in changes in shoppers' behaviours (Shankar, Inman, Mantrala, Kelley, & Rizley, 2011). One such shift is the increase in e-commerce adoption which has altered shoppers' search and browsing habits. While research has shown that most consumers prefer to browse for products online, more than half still prefer being in-store when making the final purchase decision (Briggs, 2016). Therefore, there is a potential upside for retailers to focus on marketing at the point of purchase (Löfgren, 2005).

In this paper, we referenced literature on innovations in merchandising which is a form of shopper marketing in the retail environment. While there has been extensive research on shoppers' navigation behaviours to create a positive shopper-centric environment, there is limited study on the use of consumers' purchasing patterns to design placement strategies that capitalise on synergies between products to trigger sales.

Association Rule Mining

We reviewed literature that explored association rule mining as a technique to identify the relationship between items in a large set of data. Particularly, we were interested in Market Basket Analysis, an application of association rules mining, to analyse customers' shopping patterns and formulate relationships amongst the various items placed in shoppers' basket (Gupta & Mamtora, 2014). The relationship between two product items can be represented as an association rule whereby X implies Y such that

$$X \Longrightarrow Y \quad \text{(X is an antecedent and Y is a consequent)}$$

Apriori is the most widely applied algorithm in the field of association rule mining where rules are discovered as the algorithm is passed repeatedly over a dataset and product pairs that are deemed to be infrequent are deleted (Yabing, 2013). Past literature studied different benchmarks to select rules which are useful and of interest to the user. Both statistical methods such as Support, Confidence, and Lift, as well as subjective methods, have been discussed.

Traditionally, high Support and Confidence values are used to evaluate the certainty of mined association rules. Support is defined as the percentage of transaction records that contain both products; for a pair of products X and Y, how often they are bought together in a transaction. Confidence measures the strength and accuracy of the rule, that is, how often do consumers who buy X also buy Y (Kotsiantis & Kanellopoulos, 2006). Lift takes this consideration one step further by measuring the ratio of Support divided by the probability that the antecedent and consequent occur together if the two products are independent. A lift value greater than 1 shows that the antecedent and consequent appear more frequently together than expected. Therefore, the larger the lift, the greater the association between the two products.

While rules with high Support shows that a product pair is often bought together, such pairings are generally well known and do not necessarily generate new insights (Coenen, Goulbourne, & Leng, 2004). Rules with low Support and high Confidence, on the other hand, are more likely to be counter-intuitive and interesting for business considerations. Beyond statistical measures, literature works such as (Geng & Hamilton, 2006) also suggested the use of a combination of objective, subjective and semantic measures to validate the association rules. Specifically, the user's domain knowledge is a good source of qualifier to rank the discovered rules according to various interestingness measures such as unexpectedness and conformity (Liu, Hsu, Chen, & Ma, 2000).

Though there have been extensive efforts to apply Market Basket Analysis to understand which products are commonly bought in combination with one another in a physical retail setting, most research works fell short on the practical application of the mined association rules for store merchandising purposes. Some applications include improving in-store layout (Surjandari & Seruni, 2010) and product category management (Griva, Bardaki, Pramatari, & Papakiriakopoulos, 2018) with limited focus on product placement. This paper attempts to bridge the gap in the study of product placement at the stock-keeping unit (SKU) level based on established association pairings and to evaluate the improvements before and after actual implementations. Specifically, we would like to investigate the impact of sales due to new product placements, and the addition of a new antecedent on the sales of its associated consequent in a retail store based on rules mined from other stores.

DATA SOURCE AND PREPARATION

Transactional data from receipts generated at four retail stores across the period of 1st Jun 2018 to 31st Aug 2018 was provided by the corporate sponsor. The four stores, selected based on their customer profile mix, were aligned to FairPrice's target consumer group and agreed upon during the initial discussions with the sponsor. The extracted dataset contains 9,729,581 rows of transactions. <TRAN_NO>, which reflects the transaction ID of a receipt, was repeated across the different stores and therefore not unique. To create a unique transaction ID, a new column <Unique_Trans_ID> was created by concatenating <Business_Date>, <Store>, <Reg_No>, <Tran_No>. Columns that are not relevant to our research have been streamlined, and the remaining dataset contains 17 columns as shown in Table 1. The pre-processed data file was then used to perform Market Basket Analysis using the association node function in SAS Enterprise Miner.

Table 1: Metadata Table Of 17 Data Fields

COLUMN	SAMPLE DATA	DESCRIPTION
UNIQUE_TRAN_ID	6/1/2018 12:00:00 AM31315368	Unique ID created by concatenating <Business_Date>, <Store>, <Reg_No>, <Tran_No>.
BUSINESS_DATE	6/1/2018 12:00:00 AM	Data and time of transaction in MM/DD/YYYY in 24-hour clock format.
STORE	313	Unique store code.
REG_NO	1	POS register counter number in store.
TRAN_NO	5368	Transaction number at POS counter.
LINE_NO	5	Sequence of scanned item in a transaction.
SKU	209280	Stock Keeping Unit (SKU) number, varying from 6-8 digits.
EAN	20183388610	International Article Number, varying from 11 to 13 digits.
ARTICLE_NAME	VALLEYCHEF CHK FRANKS 340G	Description of item.
DEPT	21	Unique department code ranging from 1-9.
DEPT_NAME	FROZEN FOOD	Description of department.
CLASS	21047	Unique ID of product class.
CLASS_NAME	FROZEN PROCESSED FD	Description of product class.

Table 1: Metadata Table Of 17 Data Fields		
COLUMN	SAMPLE DATA	DESCRIPTION
SUBCAT1	2104701	Unique ID of subcategory 1.
SUBCAT1_NAME	FROZEN MEAT PROCESSED FOOD	Description of subcategory 1.
SUBCAT2	210470101	Unique ID of subcategory 2.
SUBCAT2_NAME	FROZEN SAUSAGE/ FRAN	Description of subcategory 2.

THEORETICAL DEVELOPMENT

The research methodology was structured to uncover useful and interesting association rules and to design and test methods to capitalise on this information for sales-driving purposes. An overview of the approach is listed down below,

I. Market Basket Analysis of the pre-processed dataset was performed. A list of meaningful association rules was shortlisted based on Confidence and Lift value above a certain cutoff value. The list was then further streamlined based on the author's domain knowledge and assessment of the interestingness of the rules.

II. Research hypotheses were raised to study the impact of the presence of an antecedent on the sales performance of the consequent. These hypotheses were derived from the selected association mining rules discovered. In-store initiatives were then designed and implemented to measure the sales of selected consequent in the presence and absence of an antecedent, taking into account any macro and seasonal trends affecting the overall product category sales in the pilot store.

III. The proposed initiatives were implemented in a pilot store over 12 weeks. Sales tracking was performed, and the results and insights were evaluated and communicated to the corporate sponsors.

Selection Of Association Rules

In this paper, we were primarily concerned with rules which have low Support and relatively high Confidence and/or Lift values. We set the cutoff value for Support to be <0.1 and Confidence and/or Lift to be >10. This suggests that even though the two products were not purchased frequently, but whenever they were, it was very likely for them to be purchased together. Products that were on bulk purchase promotion, for example, buy 2 at a discounted price, were likely to have high Support, therefore they were not considered. The selected rules were also chosen to conform to the author's domain knowledge in the retail business to remain feasible for in-store applications. The three selected pairs of association rules are shown below,

Glad Cling Wrap 200ft ==> Glad Aluminum Foil 75sqft 1s
Support – 0.02
Confidence – 7.41
Lift – 37.07

Kim Poh Dish/wash liquid 5kg ==> Budget Kitchen Towel 6roll 60s
Support – 0.02
Confidence – 10.34
Lift – 11.05

Cirio Chopped Tomatoes 400g ==> Beef Minced kg
Support - 0.05
Confidence – 13.87
Lift – 11.86

DESIGN AND IMPLEMENTATION OF EXPERIMENT

Based on the criteria for the selection of rules, three promotional mechanisms were designed and implemented at a pilot store over a period of 12 weeks. The hypotheses tested are as follows,

- Hypothesis 1 – Placing products with high association together generates sales for consequents where X (antecedent) ==> Y (consequent).

- Hypothesis 2 – Introducing an antecedent (new product to store) will bring about an increase in sales of consequent based on established product associations from another store.

Table 2: In-store Initiatives Implemented In Test Store

S/N	ANTECEDENT	CONSEQUENT	HYPOTHESIS	IN-STORE INITIATIVES
1	Glad Cling Wrap 200ft	Glad Aluminum Foil 75sqft 1S	Hypothesis 1 – Placing products with high association together generates sales for consequents where antecedent ==> consequent.	

2	Kim Poh Dish/wash liquid 5kg	Budget Kitchen Towel 6Roll 60s	Hypothesis 1 – Placing products with high association together generates sales for consequents where antecedent ==> consequent.	
3	Cirio Chopped Tomatoes 400G	Minced Beef KG	Hypothesis 2 – Introducing a new antecedent to store will bring about an increase in sales of consequent based on established product associations derived from another store.	

Sales tracking of the above initiatives in FY2018/19 was performed over the duration of 12 weeks before implementation (Period 1) and 12 weeks after implementation (Period 2). We also compared the sales results during the same periods from 2017 to 2018 (denoted as FY2017/18). Such a comparison will take into account any sales improvement or decline due to Business As Usual (BAU) so that we can isolate any sales improvements of the consequents solely due to the experiments. The results of sales tracking and business insights derived are presented in the next section of the paper.

Table 3: Period Of Experiment		
	FY2017/2018	FY2018/2019
Period 1 (12 Weeks Before)	06 September 2017 ~ 28 November 2017	06 September 2018 ~ 28 November 2018
Launch Week	29 November 2017 ~ 05 December 2017	29 November 2018 ~ 05 December 2018
Period 2 (12 Weeks After)	06 December 2017 ~ 27 February 2018	06 December 2018 ~ 27 February 2019

RESULTS

Evaluation of Hypothesis 1

Experiment 1 - Glad Cling Wrap 200ft ==> Glad Aluminum Foil 75sqft 1s

Figure 1: Sales Of Consequent (Aluminum Foil)

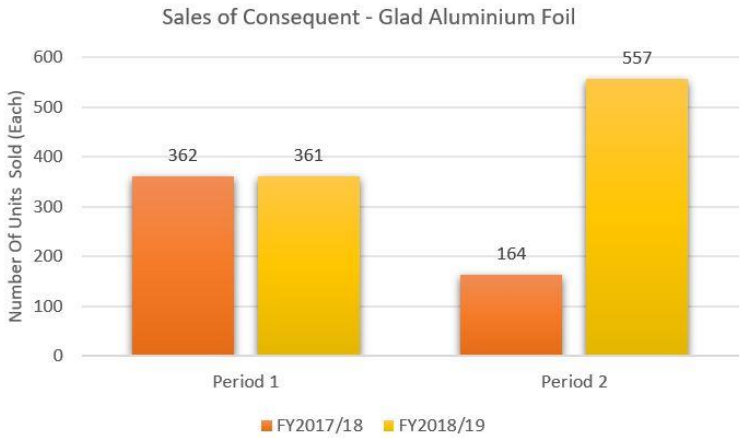


Figure 2: Category Sales For Party Needs

In Figure 1, the sales of consequent Glad Aluminum Foil 75sqft 1s decreased by 1 unit in Period 1 when we compared between FY2017/18 and FY2018/19, accounting for Business As Usual (BAU) negative growth from one year to the next. This BAU growth can be used to offset the expected BAU growth in Period 2 so that we can isolate the sales growth solely due to the implementation. In Period 2, we achieved an increase of 393 units when we compared between FY2017/18 and FY2018/19. With the BAU decline offset of 1 unit, we will achieve a net increase of 394 units. To put the comparison in context with Figure 2, we computed the BAU growth in sales dollar ($) for the Party Needs product category, where the consequent belongs to. The BAU sales dollar increased by $6070 in Period 1 when we compared between FY2017/18 and FY2018/19. In Period 2, we achieved only an increase of $4265 when we compared between FY2017/18 and FY2018/19, which signified that this product category performed worse in Period 2 than in Period 1, unable to achieve the expected BAU sales dollar increase. Such a

comparison further confirmed that while the party needs product category performed worse in Period 2 of FY2018/19, the consequent performed better than expected BAU, defying the expected downward trend in its product category.

Experiment 2 - Kim Poh Dish/wash liquid 5kg ==> Budget Kitchen Towel 6roll 60s

Figure 3: Sales Of Consequent (Kitchen Towel)

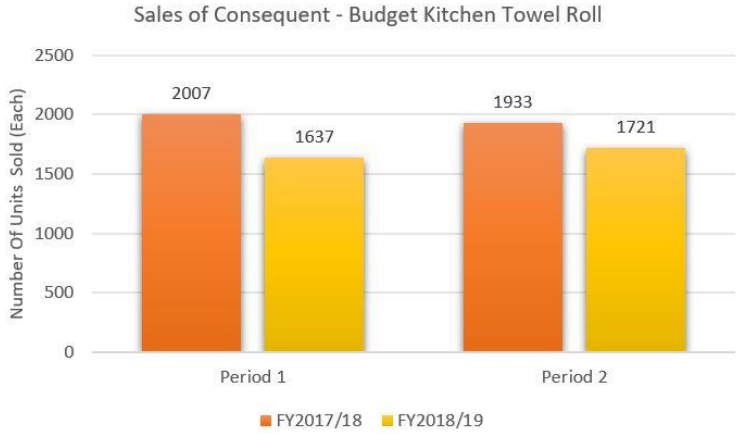
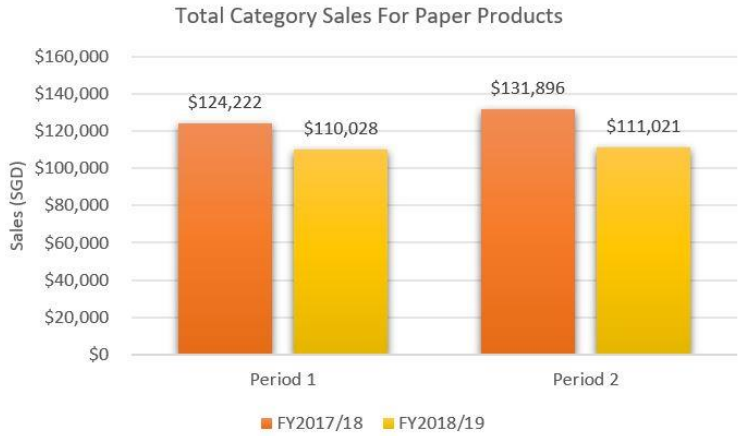


Figure 4: Category Sales For Paper Products



Evaluation of the sales of consequent Budget Kitchen Towel revealed similar positive results after implementation. In Figure 3, the BAU decline in the sales of the consequent in Period 1 between FY2017/18 and FY2018/19 was 370 units. In Period 2, the decline was reduced to 212 units. Offsetting the BAU decline, it was shown that the sales decline was reduced to a lower number due to the implementation, and thus achieved a net positive gain of 158 units. In Figure 4, an investigation of the Paper Products category which the consequent belongs to, showed an overall decline of $14,194 in sales dollar in Period 1 when we compared FY2017/18 with FY2018/19, representing the BAU sales dollar decline. In Period 2, the decline was increased to $20,875 between FY2017/18 and FY2018/19. Despite a higher sales dollar decline in the product category in Period 2, the consequent performed better than the expected BAU decline, slowing the rapid downward trend.

The results from both Experiments 1 and 2 provided insights to hypothesis 1, which suggested that product (antecedent) with a high association is likely to have a positive impact on the sales of the consequent when both products are placed next to each other in the sales display, despite an expected decline in sales and growth in their respective product categories.

Evaluation of Hypothesis 2

Experiment 3 - Cirio Chopped Tomatoes 400g ==> Beef Minced kg

Figure 5: Sales for Consequent (Minced Beef) vs. Total Category Sales For Chilled Meat

The antecedent Cirio Chopped Tomatoes 400g was a new brand of product introduced in the pilot store from the launch week and thereafter. Sales of the consequent Minced Beef was monitored in Period 2 against the same period last year without the presence of the antecedent. In comparison to last year, sales of consequent went up by $6,569, from $21,392 to $27,961 (30.7%) despite a decline in the overall category sales of Swiss Butchery concessionary category by $7013, from $114,929 to $107,916 (-6.1%).

The encouraging results from experiment 3 reinforced hypothesis 2 which states that an introduction of a new product (antecedent) may have a positive impact on the sales of the corresponding consequent based on established association rules from other stores, where traditionally, association rule implementations were often based on rules discovered within the same store. All in all, association rules are potential sources of information for designing in-store displays for synergy across different products, and to drive sales, which is especially important for retailers amidst declining growth.

CONCLUSION

This paper attempts to explore new methods using Market Basket Analysis to capitalise on an established supermarket retailer's database of transactional receipts as a proxy for consumers' purchase patterns. We sought to address two research hypotheses – whether the placement of antecedent near its corresponding consequent would bring about an increase in the sales of the consequent, and whether the introduction of a new product (antecedent) to store based on established association rules from other stores would have a similar positive impact on the sales of the consequent. Three association rules with low Support and high Confidence and/or Lift values were selected, and in-store initiatives were designed and implemented in a pilot store

over a period of 12 weeks. From the results, we observed an improvement in sales of consequents after implementation across all three initiatives, when compared to the 12 weeks before implementation, as well as when compared to the same period last year and, despite an overall poorer performance of their corresponding product categories.

ACKNOWLEDGMENT

The authors would like to thank NTUC FairPrice for the data used for the analysis presented in this paper.

REFERENCES

Briggs, F. (2016, July 6). *Shoppers Prefer To Make Final Purchase Decisions In-Store, Despite Researching Online, Study Shows*. Retrieved from Forbes: https://www.forbes.com/sites/fionabriggs/2016/07/06/shoppers-prefer-to-make-final-purchase-decisions-in-store-despite-researching-online-study-shows/#611637a05456

Clement, J. (2020, Aug 27). *Retail e-commerce sales worldwide from 2014 to 2023*. Retrieved from Statista : https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

Coenen, F., Goulbourne, G., & Leng, P. (2004). Tree Structures for Mining Association Rules. *Data Mining and Knowledge Discovery*, (pp. 25-51).

Geng, L., & Hamilton, H. J. (2006). Interestingness Measures for Data Mining: A Survey. *ACM Computing Surveys.* Association for Computing Machinery.

Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, (pp. 1-16).

Gupta, S., & Mamtora, R. (2014). A Survey on Association Rule Mining in Market Basket Analysis. *International Journal of Information and Computation Technology* (pp. 409-414). International Research Publications House.

Kopalle, P., Biswas, D., Chintagunta, P. K., Fan, J., Pauwels, K., Ratchford, B. T., & Sills, J. A. (2009). Retailer Pricing and Competitive Effects. *Journal of Retailing*, (pp. 56-70).

Kotsiantis, S., & Kanellopoulos, D. (2006). Association Rules Mining: A Recent Overview. *GESTS International Transactions on Computer Science and Engineering*, (pp. 71-82).

Liu, B., Hsu, W., Chen, S., & Ma, Y. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems and their Applications* (pp. 47-55). IEEE.

Löfgren, M. (2005). Winning at the First and Second Moments of Truth: An Exploratory Study. *" Managing Service Quality*, (pp. 102-115).

Shankar, V. (2011). *Shopper Marketing.* Marketing Science Institute.

Shankar, V., Inman, J. J., Mantrala, M., Kelley, E., & Rizley, R. (2011). Innovations in Shopper Marketing: Current Insights and. *Journal of Retailing*, (pp. 29-42).

Surjandari, I., & Seruni, A. C. (2010). Design of Product Placement Layout in Retail Shop. *Theory of Computing Systems*, (pp. 43-47).

Yabing, J. (2013). Research of an Improved Apriori Algorithm in Data Mining Association Rules. *International Journal of Computer and Communication Engineering.*

DECISION SCIENCES INSTITUTE
Deep Learning Based Expert Trace Recommender System

Ching-Chin Chern
National Taiwan University
Email: cchern@ntu.edu.tw

Shang-En Lee
National Taiwan University
Email: r06725015@ntu.edu.tw

Bo Hsiao
Chang Jung Christian University
Email: bhsiao@mail.cjcu.edu.tw

ABSTRACT

This study proposes a recommender system, EXTRA, with a recurrent neural network and expert weights as its core. By aggregating the users' predicted viewing sequences adjusted further by expert weights, an accurate recommendation list can be generated. The trace prediction model for users is trained by the historical records of the users' interactions with items, which is able to predict the item with which a user would want to interact next. The expert weight for a user is obtained from users' own information and items interacted by the user. From the experimental results, we can confirm that the performance of EXTRA is far better than that of content-based method and collaborative methods.

KEYWORDS: Recommender System, Machine Learning, Deep Learning, Recurrent Neural Network, Expert Decision, Sequential Data

INTRODUCTION

With the development of Internet, the volume of information online has increased to an extremely huge scale; for instance, there are millions of distinct catalog items on Amazon (Linden, et al., 2003). Browsing such large volume of content to find useful or desirable objects takes lots of time and effort. Thus, Recommender System (RS) is designed to provide suggestion of objects within a large collection of possible options that most possibly match a user's interest (Ricci, et al., 2010). Recommender Systems are now widely used in a variety of online services. Users can watch recommended movies on streaming platforms, check recommended goods on e-commerce platforms, and receive recommended news on online news media. In addition, about 60% of all video clicks from YouTube's home page are based on recommendations (Davidson, et al., 2010), and 80% of movies watched on Netflix are through recommendation (Gomez-Uribe and Hunt, 2015). Recommender Systems guide users to discover items meeting their interests that boosts the decision-making process. With Recommender Systems, not only users can save a lot of time and focus on the items they need, but also companies can provide better user experiences to retain their customers.

Recommender Systems use features extracted from users and items to generate recommendation results. Typically, the features are based on users' general data, items' content, and historical interactions between users and items. There are mainly three types of

recommender systems based on the input data: collaborative filtering, content-based, and hybrid (Zhang, et al., 2017). Collaborative filtering (CF) is a popular recommendation algorithm that has been widely used in several online platforms. The goal of CF-based recommender systems is providing suggested items to the user from items taken by those who have similar interests with this particular user. CF generally represents a user as a feature vector recording the interaction between the user and every item, and utilizing the matrix factorization method to generate recommendations (Balabanović and Shoham, 1997; Wang, et al., 2016; Linden, et al., 2003; Yun, et al., 2017). Content-based recommender system, however, provides recommendations for a user based on the content similarity of a recommended item and items that the user presents interest before. In order to do this, content-based RS will build a profile for each user by analyzing the content of items which interacted with the particular user in the past (Balabanović and Shoham, 1997; Yang, et al., 2017). Hybrid recommender system integrates two or more recommendation strategies to avoid shortcomings of using just one (Burke, 2007).

Both CF and content-based recommender systems do not consider the sequential relations of items. For CF recommender system, the item-user matrix only records the occurrences of the relations but not the sequences of relations. For a content-based system, it is apparent that the result of its recommendation is restricted to the content that a user has interacted in the past (Balabanović and Shoham, 1997). However, in the real world, users usually select items follow a specific order; for example, people would purchase screen protector after purchasing a mobile phone. In addition, researchers would be interested in a new released research that related to a past research they have read before. Hence, a recommender system considering sequential relations would be useful.

This study aims to build a recommender system in a different way that considers sequential relations between items. We want to make recommendations through aggregating users' trace histories and add expert weights for every user to distinguish the importance of his/her opinions. The proposed system should not only provide a more accurate recommendation results but also help users discovering fresh items. Moreover, the result of the recommendation should be useful for users and boost their decision-making progress.

PROBLEM DESCRIPTION

The purpose of this research is to build a recommender system aggregating users' trace histories with expert weights. In order to build a recommender system, we have to consider three elements: data, constructing techniques, and performance evaluation methods. The data used to build a recommender system depends on the type of recommendation we want to generate and the information that users communicate. The data usually contains the historical records of users' behavior. In this study, to consider the sequential relations between items, the historical records of users are the sequential histories of users.

To generate recommendations, the system has to process the input through feature selection techniques mentioned in previous chapters. In this research, the feature selection techniques have to learn features and make predictions based on users' behavioral data. After implementing the proposed method, we have to find effective measures and convincible baseline methods to evaluate the performance of the proposed method. The performance could be measured by the accuracy of recommendations, the error of predictions, or the feedback from the users.
The first issue related to the input data is how to define an expert in certain domain. In order to find experts from users, we have to find a method to obtain the experts. The solution of expert finding has been mentioned in previous chapter. Typically, to know a user's expertise, the

historical user data is needed because the data may show his/her interests in certain domain. Information retrieval methods, such as TF-IDF, language model, and Latent Dirichlet Allocation, can be used to analyze these data and obtain the score for each user that represents the user's degree of expertise on a certain question or topic. The expert finding model is designed differently based on the type of recommender systems since the attributes in the dataset and the purpose of expert finding may differ case by case. However, this study aims to develop a common technique to find experts for all types of recommender systems so that even if the attributes in the dataset and the purpose of expert finding differ, the expert finding technique will still be applicable.

The second issue is how to collect and extract experts' data, e.g., experts' domain knowledge or experts' historical behavioral records. Experts' domain knowledge data contains more high-level concept based on experts' background knowledge and experiences. For example, experts provide opinions or ratings about the preference of an agency based on experts' domain knowledge (Clinton and Lewis, 2017). Experts' historical behavioral records is experts' behavioral data, e.g., browsing record, purchase record, posting record, etc. For example, to investigate how an expert searches on the Internet, the searching records and the operating records are analyzed to understand the behavior of the expert (Tabatabai and Shore, 2005). The data can be obtained through a historical dataset, a survey of experts' opinions, personal interviews, etc. How to integrate the data from different sources is also an issue to be tackled, which includes synchronizing the format of the data from different sources and combining attributes from different sources to generate new attributes.

The third issue is how to collect users' data as the input of the system. To generate individualized recommendations for a user, we must find out the user's interest by extracting features that can represent the interest of users from their behavioral data. For example, the reviews written by a user on an e-commerce website show his/her preference of products; the ratings given by a user on IMDB represent his/her taste in movies. The data could be obtained from many different sources, e.g., users' browsing records, reviews, ratings, etc. Thus, the data integration problem mentioned above for experts' data is also an important issue for users' data.

The fourth issue is how to obtain items' data. In order to generate recommendations for users, the information of items is also required. For example, for a music recommender system, the singer and the release date of a song are required attributes. If there is no an existing dataset for items, a web crawler may be needed. The data integration problem mentioned above from experts' and users' is an important issue to be addressed here for the items, too.

The main challenge involved in the previous three issues is to choose appropriate attributes for each element so that the data is representative to provide enough information of experts, users, and items. The more representative data we have, the more accurate recommendations we can provide. Based on the kind of data and the type of the model, we can know what data is needed for the training and testing processes of the model. Data cleaning problem, e.g. dealing with missing-value data, correcting the erroneous data, and transforming data, is also a challenge in this study because of big data in nature.

The last issue is how to deal with the cold-start problem, which contains the user cold-start problem and the item cold-start problem. The user cold-start problem occurs when there is very little data for a new-coming user. We are not able to understand the interest of the user and generate the feature vector for the user. The item cold-start problem, on the other hand, occurs when none or few records exist for a new-releasing item. The item is not recommended if the

interaction between users and the item is unknown. Thus, to deal with these two problems, we have to find an alternative method or design another model.

After the data is prepared, the next step is utilizing these data to build up a recommender model and generate recommendations for users. The first issue related to the building techniques is to determine the outcome of the system. Before choosing the model, we have to confirm what kind of output is desired first. In other words, the label of each input data has to be defined. Typically, there are two types of outputs from a recommender system: rating and ranking. Rating prediction generates the missing entries of the user-item rating matrix, while ranking prediction produces a ranked list of items for users (Zhang, et al., 2017).

The second issue is to choose an appropriate model for the system. This problem is based on the previous issue to determine the output of the model first. This issue consists of three tasks: method selection, data transformation, and parameter tuning. For method selection, we aim to build a deep learning model for the recommender system in this study. Thus, the deep learning techniques mentioned in chapter 2 can be taken into consideration, which also depends on the format of input data. For example, if the input contains sequences of words, RNN model will be a suitable choice. CNN, on the other hand, may be the best choice if we have to deal with image data. To determine which model to use, the data need to be determined first.

After we select a method, the data collected from experts, users, and items have to be transformed into the input format of this model. Data transformation may include data processing and feature engineering. For example, document data has to be transformed into tf-idf vectors or document embedding vectors. For most of machine learning techniques, a dataset should be split into a training dataset and a testing dataset. In the training process, a validation dataset split from the training dataset is required to check if the model is overfitting. The testing dataset will be used in the evaluation step.

The next task is to adjust the parameters of the chosen model. For example, adjusting the learning rate in a machine-learning model can control how much we update the weights of the model. Units of each layer determines how many hidden nodes are contained. On the other hand, there are also many choices for the activation function used by each layer, such as Sigmoid, ReLU, TanH, and SELU. To figure out the best setting for the model, K-fold cross validation may be adopted. This task will take a lot of time because we need to test numerous different combinations of parameters.

The purpose of evaluation is to ensure the robustness and show the improvement of the proposed system compared to other methods. The first issue we face is to choose performance metrics of the system that we want to evaluate. Although there are many different metrics we can evaluate, we should choose metrics related to this research. This study focuses on the accuracy of the system but not the privacy. The accuracy is the most common metric used to evaluate a recommender system's performance, which can directly tell us whether the recommendations are accepted by the user or not. After choosing appropriate metrics, we have to decide which measures are used to evaluate the metrics. For example, there are many ways to measure the accuracy of a recommender system. We have to choose approaches based on the output of the system. If the system generates ratings for each item, the measurement will be 1 - Mean Absolute Error (MAE), where MAE can be calculated by the errors between numbers (Wang, et al., 2016). On the other hand, if the system provides recommended items directly, Precision at N, Recall at N, and Mean Reciprocal Rank (MRR) will be suitable choices (Ricci, et al., 2010; Said and

Bellogín, 2014). In sum, to solve this issue, we have to clarify the purpose of this research and determine the output of the proposed system.

The second issue is to decide which experimental setting is used. There are three types of experimental settings discussed in previous chapter. Different types of experimental settings are used for different purposes. For example, to test the novelty of the system, conducting a user study may be a suitable choice. However, to test the prediction accuracy of the system, offline experiments are enough. If we want the system to be test in a real environment, online evaluation should be conducted. Thus, in order to choose one or more experimental settings, we have to know the metrics we want to test and the scenario we expect the system to be tested in.

The last issue is how to find suitable baseline methods that can be applied in recommender systems. We have to find models that are most commonly used or strategies that are often applied to generate recommendations in certain domain. For example, top-10 hot items, which shows the top 10 popular items on the system. Although it is very simple, this strategy is commonly seen on e-commerce platforms. Based on the scenario we want to evaluate, content-based methods and collaborative filtering methods can also be options for baseline models. In sum, the baseline model should be popular and reasonable to users.

THE EXPERT TRACE RECOMMENDER SYSTEM (EXTRA-RS)

This study aims to build a recommender system adopting experts' decisions, which utilizes the historical traces of experts to make recommendations. We introduce an innovative recommender system, EXTRA, which provides recommendations for users who only have one history. It merges all past user experiences of an item to provide recommendations for users that currently interact with that item. The major steps of building EXTRA are: (1) Data Preparation, (2) Trace Model Building, (3) Expert Weight Calculation, (4) Recommender Model Building, and (5) Evaluation.

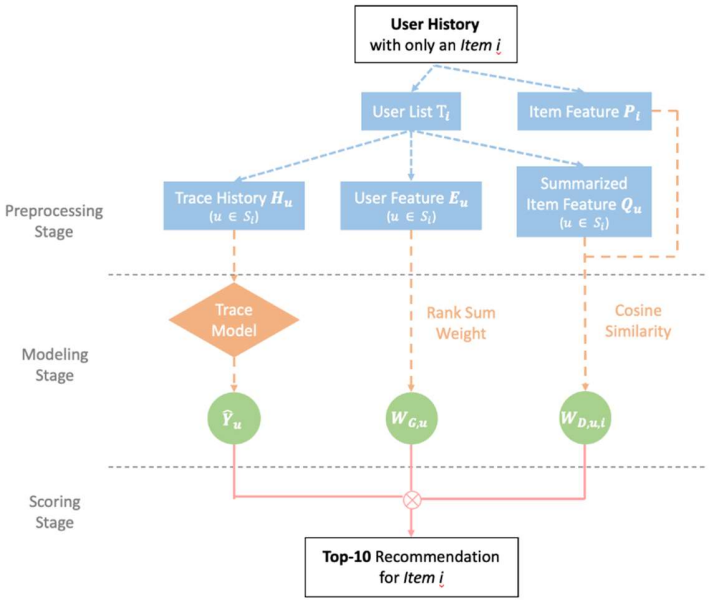


Figure 1: Overview of EXTRA

A recommender system basically consists of two roles: users and items. For example, on a shopping platform, users are the customers, and items are the products; on a forum, users are the authors, and items are the posts. We assume that there are M users and N items. In order to build EXTRA, five types of data should be collected from a given dataset: user lists, user feature vectors, item feature vectors, summarized item feature vectors, and trace history vectors.

A user list of an item stores users that have interacted with this item. For example, the buyer list of a product on a shopping platform or the replier list of a post on a forum. The users in the user list of an item are the target that we want to refer to, since they have related experiences of that item. Let Ti denote the user set for item i. User feature vector will be used to determine the expert weight for each user, which represents the characteristics of a user. For example, it can be the activeness of a user, such as the number of purchases on a shopping platform, the number of posts on a forum, or the number of login times; the general data of a user, such as age, gender, and occupation. The user feature vector for user u can be denoted as Eu, which has K dimensions representing K different attributes of each user.

Item feature vector should represent the characteristics of an item. For example, it can be the content term frequency–inverse document frequency (tf-idf) or Bag-of-Words (BOW) vector of a post on a forum. The item feature vector for item i can be denoted as Pi, which has D dimensions representing D different attributes of each item. The value of D can be adjusted based on the task. A summarized item feature vector for a user extracts his/her taste or expertise from all items that visited by this user. In order to learn what kind of items would meet a user's taste, we want to compare the user's summarized item feature vector with all the item feature vectors. Thus, to make them compatible, we generate the summarized item feature vector for user u by using the sum of all item feature vectors that user u has visited. Let Qu denote the summarized item feature vector for user u, which has the same dimension, M, as the item feature vector.

Trace history vector for a user describes the past interactions between that user and items that user visited before. Thus, it is a sequential behavioral record for a given user. For example, it could be a purchase sequence on a shopping platform, a browsing sequence on a forum, or a watching sequence on a video-streaming platform. Let Hu denote the trace history vector for user u, which contains the items' ID in order.

Current recommender models such as content-based models and collaborative-filtering models do not concern sequential relation between users and items. However, in some situations, taking this relation into recommender model is more appropriate. For example, shopping, searching information, and reading posts have sequential relation. In these situations, users' previous actions could influence the next action. Thus, we want to have EXTRA to learn sequential relation. To learn the sequential path of users, we need to build the trace model, which have the ability to predict the next item that interests a target user. For a user, we build a predictive model to generate probabilistic distribution of his/her next move rather than directly use his/her next move. In this way, we can avoid the data sparsity problem; for example, if user does not have the next interacted item, there would be no output of this user. In addition, directly using a user's next move may be biased, since we would only obtain an item for recommendations and have no information for other items. This model takes the historical item sequence of a user as input, and outputs the next most likely visited item for that user.

The best network fitting this trace model is Recurrent Neural Network (RNN), which is suitable for modeling sequential data. Unlike feedforward neural network, RNN has internal memories in its hidden states. We will suggest using Long Short-Term Memory (LSTM) or Gated Recurrent Unit

(GRU) to overcome the vanishing gradient problem. The input sequence of the trace model is an ID sequence of items with length d. The embedding layer transforms item's ID to a corresponding latent vector for the item. Then, RNN layers and full-connected layers will learn the relation between these vectors. The number of RNN and full-connected layers can be customized. The default dimension of embedding, RNN, and full-connected layers is 128, which can be adjusted depending on the task. To prevent overfitting, Dropout layers could be added, and the dropout rate can be customized. Finally, the output layer using the softmax function as the activation function will output a vector of which the length is the number of items. Let Su denote the score vector of user u. The value of the ith dimension in output score vector represents the possibility that item i is the next visited item. The loss function is the categorical cross-entropy between the output score vector and the answer one-hot vector.

In order to build the input and output data for the trace model, we have to process Hu for each user u. The input sequence of the trace model should be fixed length, and the output should be a number indicating an item's ID. Thus, we set a moving window with a size, d, and move through Hu for each user u. The window size, d, is the input length of the trace model, which means the input sequence is the sequence within the window. The default value of d is set to 5; however, this value can be adjusted depending on the situation. In addition, the item next to the window is the output of the trace model. The total moving times for each Hu are (Length of Hu – d). With the moving window, the input and output pairs can be generated from Hu for each user u. The pairs can be separated into a training set and a testing set. We use the training set to train the trace model with a proportion of training set as validation, and use the testing set to evaluate the overall performance of EXTRA.

Treating all users' trace in the same way is biased, since the importance of their experiences could differ from their expertise in that domain. Thus, in this part, we are going to find out how to obtain the expert weight for each user. We separate the expert weight into two parts: the expert general weight and the expert domain weight.

The expert general weight describes the characteristics and basic information of a user, which can be derived from user feature data, E. There are many ways to generate the expert general weight; here, we provide a method that can fit most situations. We choose the most important variable from user feature data; for example, the total purchase amount of a customer on a shopping platform, or the number of posts for a user on a forum. Then, for each user, we calculate the rank sum weight (RS) to be the expert general weight. A previous study has shown that RS method has better performance than other rank-order weighting methods (Saeid, et al., 2011).

The expert domain weight, on the other hand, describes the domain expertise of a user. For example, on a shopping platform, some people often buy 3C products; however, some seldom buy them. It is obvious that when we make a decision of buying 3C products, the former's experiences are more valuable than that of the latter. Thus, the expert domain weight can be calculated as the similarities between a user's past experience and each item; that is, the similarities between Qu for each user u and Pi for each item i. Let WD,u,i denote the expert domain weight of user u and item i, which is a numerical score that indicates how close is the relation between user u's experiences and item i.

To generate recommendations for an item, the last step is combining the prediction of the trace model, the expert general weight, and the expert domain weight. The prediction of the trace model is a list of vectors, which consists of the score vectors for all users in the user list. The score vector for a user indicates the possibility of each item that would be visited by that user. Thus, we could

use the sum of score vectors to generate the recommending priority for each item. However, this method gives all the score vectors the same weight; hence, the contribution of each user to his/her recommended items is the same. It is biased, since the importance of each user could be different. In order to deal with this problem, we multiply each score vector by the expert general weight and the expert domain weight. We calculate the inner-product of the score vectors, expert general weight, and expert domain weights to generate the final adjusted score vectors, S′i, where u is the user related to this input item. Then, we sort the sum of adjusted score vectors, S_{total}, by users to obtain top-10 scored items, which are the final recommendation for the input item.

COMPUTATIONAL ANALYSIS

In this section, EXTRA was implemented and tested with the data posted on Mobile01, the biggest life website and forum in Taiwan. The data is collected and provided by eLand, a Taiwanese cloud service company. The relative metrics of different experiments to test EXTRA are calculated and compared with the results obtained by the other three methods: content-based, collaborative filtering and top-view methods. In addition, we design several experiments with different scenarios to test the validity of EXTRA.

The original posts on Mobile01, collected by eLand from 2014 to 2016, consist of several popular discussion boards. We choose Investment board within 2015, including 196,966 entries, to demonstrate the performance of our model. Each entry is a post record from a forum's user, which can be either a main post or a reply. In this case, items are the main posts and users are the authors or commenters. There are 8,495 items and 14,273 users in this chosen dataset. EXTRA is applied to recommend posts that a certain user would want to "reply" to rather than "browse" only, since the dataset does not contain "browse" records for each user.

In order to build an EXTRA-based recommender system, EXTRA follows Subsection 4.1 to process origin data and generate the data that EXTRA needs, including user lists (T), user feature vectors (E), item feature vectors (P), summarized item feature vectors (Q), and trace history vectors (H). The user lists are stored as a dictionary with 8495 entries, and each entry represents the user list of an item. For user feature vectors, only the number of words is chosen as an attribute; thus, each user has a one-dimensional user feature vector. Item feature vector for each item is a 7,000-dimension BOW vector, consisting of the word counts in both main and reply of a post. Summarized item feature vector for each user is a 7000-dimension BOW vector that stores the word counts in all main and reply posts of the user. Trace history vector for each user is a sequence of post titles that posted or replied by the user. The length of each trace history vector could be different depending on the records of each user. After preparing these data, EXTRA is ready to be trained and provide the recommended list.

For testing the performance of EXTRA, we compare the results of EXTRA with three baseline models: content-based method, item-based collaborative filtering, and top-view method. Content-based method (CB) makes recommendations by comparing the content similarities between items. In this study, the content of an item is represented as a BOW vector. The recommended items for an input item are the items with the highest cosine similarity score to that input item. Item-based collaborative filtering (ICF) compares the user relations between items and find items that have similar user relations. In this study, the user relations are the "reply" records. The vector of an item stores the number of replies of each user; that is, the u-th element in this vector is the number of replies of user u to this item. Given an input item, the items with the highest cosine similarity score to that input item will be recommended. Top-view method (Top-V) simply picks top-10 popular items as the result of recommendation, which is commonly used in many content websites.

This method will provide exactly the same result for every testing set, since the top-10 popular items are fixed.

For this experiment, the testing results of EXTRA and three baseline methods are shown in Table 1 (The details are listed in the Appendix A). In order to test if there are differences between these four methods, we apply two-way randomized block ANOVA. However, the required condition of two-way randomized block ANOVA is violated because the variances of the results obtained from the four methods differ. Thus, instead of using two-way randomized block ANOVA, we apply Friedman test, a non-parametric version of two-way randomized block ANOVA. The null hypothesis of Friedman test is that the locations of all four populations are the same, while the alternative hypothesis is that at least two population locations differ.

The p-values of Friedman test for recall@10 and MRR@10 are 5.88×10-13 and 2.65×10-12, which infer that recall@10 and MRR@10 are different between at least two of the four methods. By comparing the rank sum in Table 2, we can infer that for both recall@10 and MRR@10, the top-performance models are EXTRA and ICF.

Table 1: Average performance of EXTRA and baseline models

	Metric	
Model	Recall@10	MRR@10
EXTRA	**0.1457**	**0.0685**
CB	0.0347	0.0150
ICF	0.0549	0.0247
Top-V	0.0443	0.0159

Furthermore, we apply matched-pair t-test to compare the performances of EXTRA and ICF. For both recall@10 and MRR@10, we apply Shapiro-Wilk normality test to check whether the differences of paired data are normally distributed. The p-value for recall@10 is 0.9995; the p-value for MRR@10 is 0.9990, which infers that the differences of data for both metrics are normally distributed. Thus, they both meet the required condition of matched-pair t-test. The null hypothesis of matched-pair t-test is the performances of EXTRA and ICF are the same; the alternative hypothesis is the performance of EXTRA is better than ICF. The p-values of matched-pair t-test for recall@10 and MRR@10 are 7.86×10-26 and 6.44×10-25. Hence, there are overwhelming evidences to infer that the performances of EXTRA are better than those of ICF in terms of both metrics.

Table 9: Rank sum of Friedman test

	Metric	
Model	Recall@10	MRR@10
EXTRA	80	80
CB	20	23
ICF	60	60
Top-V	40	37

To sum up, there are overwhelming evidences to prove that EXTRA performs better than the other three baseline models do on different topics and different platforms. From the experiments, EXTRA is proved to be suitable to build recommender systems for forum-like platforms. Furthermore, compared to other baseline methods, EXTRA performs better. Top-view method has the worst performance, since its recommendation is fixed for each input. Top-view items do attract users; however, most of the time, users would not keep reviewing top-view items. For content-based method, the drawback is that the method only recommends items with similar content but does not bring fresh things to users. EXTRA does not suffer from this problem, since the model structure does not fully rely on items' content. The performance of item-based collaborative filtering is the best among three baseline models. However, the performance of ICF is still far worse than EXTRA. The difference between EXTRA and ICF is that EXTRA considers sequential relations and the degree of users' expertise. In addition, the process of making recommendations by CB and ICF is much slower than by EXTRA, since they have to compare every item to find the most relevant ones, while EXTRA only need to make prediction based on the users in the item's user list.

CONCLUSIONS

In this study, we propose an item-based recommender system, EXTRA, to make recommendations by utilizing the historical traces of experts. This study develops a recommender system prototype based on EXTRA using the packages and the programming language on Python and this prototype can be run on Python console. This EXTRA-based recommender system is then applied on the data collected from Mobile01, the biggest life website and forum in Taiwan, for different scenarios based on the experiments we design to evaluate EXTRA. The results of experiments show that EXTRA has better performance compared to the other three popular baseline methods: content-based method, item-based collaborative filtering method, and top-view method. Moreover, EXTRA can make recommendations on different topics and different platforms. However, we recommend to apply EXTRA with at least thousands of trace histories and high main post ratio. In order to have better performance, the dataset applied EXTRA should have sequential relations between items, and users' opinions are valuable to each other. In summary, EXTRA provides a new aspect to make accurate recommendations for users. By utilizing trace histories and expert weights, the behaviors of users can be learned further. Moreover, EXTRA demonstrates an application of deep learning on recommender system by applying recurrent neural network (RNN).

ACKNOWLEDGEMENTS

This research was sponsored by the Ministry of Science and Technology in Taiwan, under project number MOST 108-2410-H-002-226.

REFERENCES

Balabanović, M. and Y. Shoham, "Fab: content-based, collaborative recommendation." Communications of the ACM, 1997. 40(3): p. 66-72.

Burke, R., "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction, 2002. 12(4): p. 331-370.

Clinton, J. D. and D. E. Lewis, “Expert Opinion, Agency Characteristics, and Agency Preferences.” Political Analysis, 2017. 16(1): p. 3-20.

Davidson, J., B. Liebald, J. Liu, P. Nandy, and T. V. Vleet, “The YouTube Video Recommendation System.” Proceedings of the fourth ACM Conference on Recommender Systems, 2010, p.293-296.

Gomez-Uribe, C. A. and N. Hunt, “The Netflix Recommender System.” ACM Transactions on Management Information Systems, 2015. 6(4): p. 1-19.

Linden, G., B. Smith, and J. York, “Amazon.com Recommendations.” IEEE Computer Society, 2003. 7(1): p. 76-80.

Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor, “Recommender Systems Handbook.” Springer, New York (2010).

Saeid, M., A. A. A. Ghani, and H. Selamat, “Rank-Order Weighting of Web Attributes for Website Evaluation.” The International Arab Journal of Information Technology, 2011. 8(1): p. 60-68.

Said, A. and A. Bellogín, “Comparative recommender system evaluation,” in Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14. 2014. p. 129-136.

Tabatabai, D. and B. M. Shore, “How experts and novices search the Web.” Library & Information Science Research, 2005. 27(2): p. 222-248.

Wang, E, H., J., M. Song, Q. Bi, and Y. Liu, “Incremental weighted bipartite algorithm for large-scale recommendation systems.” Turkish Journal of Electrical Engineering & Computer Sciences, 2016. 24: p. 448-463.

Yang, S., M. Korayem, K. AlJadda, T. Grainger, and S. Natarajan, “Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach.” Knowledge-Based Systems, 2017. 136: p. 37-45.

Yun, Y., D. Hooshyar, J. Jo, and H. Lim, “Developing a hybrid collaborative filtering recommendation system with opinion mining on purchase review.” Journal of Information Science, 2017. 44(3): p. 331-344.

Zhang, S., L. Yao, A. Sun, and Y. Tay, “Deep Learning based Recommender System: A Survey and New Perspectives.” ACM Computing Surveys, 2018. 1(1): p. 1-35.

DECISION SCIENCES INSTITUTE

Multi-Dimensional Observational Learning in Social Networks: Theory and Experimental Evidence

Liangfei Qiu
University of Florida
Email: liangfei.qiu@warrington.ufl.edu

Arunima Chhikara
University of Florida
Email: arunima.chhikara@warrington.ufl.edu

Asoo Vakharia
University of Florida
Email: asoo.vakharia@warrington.ufl.edu

ABSTRACT

The prevalence of consumers sharing their purchases on social media platforms (e.g. Instagram, and Pinterest) and future consumers using this information substantially impacts online retailing. We examine how product characteristics and the information provider type jointly impact the purchase decision in a social network setting. Utilizing an analytical observational learning framework of product differentiation and social ties and two experimental studies, we find that the effect of learning from strangers is stronger for vertically differentiated products than for horizontally differentiated products. However, the learning from friends does not depend on whether the underlying product is horizontally or vertically differentiated.

KEYWORDS: Multi-Dimensional Observational Learning, Social Ties, Product Differentiation

DECISION SCIENCES INSTITUTE
Beauty Contest and Social Value of Fintech: An Economic Analysis

Anurag Garg
University of Florida
Email: anurag.garg@warrington.ufl.edu

Liangfei Qiu
University of Florida
Email: liangfei.qiu@warrington.ufl.edu

Subhajyoti Bandyopadhyay
University of Florida
Email: shubho.bandyopadhyay@warrington.ufl.edu

ABSTRACT

With the advance of financial technology (Fintech), traders in financial markets are using the information available on social media to gauge investor sentiment and form higher-order beliefs. Following the insight of Keynes (1936) on financial markets being akin to a beauty contest, we develop an analytical model to analyze how higher-order beliefs, driven by the Fintech revolution, affect market efficiency and social welfare of investors. Since accounting disclosure is the main source of public information, our results highlight that the use of Fintech in financial trading can dramatically affect the optimal level of accounting disclosure (i.e., transparency) in the market.

KEYWORDS: first-order beliefs, higher-order beliefs, Fintech, public information, private information

DECISION SCIENCES INSTITUTE

Don’t let your left hand know the merciful deeds of your right hand? Blatant Benevolence and Social Capital Attainment on Social Network Sites (SNS)

Jiayuan Zhang
University of Rhode Island
jiayuanzhang@uri.edu

Koray Özpolat
University of Rhode Island
koray@uri.edu

Gulver Karamemis
Georgia Southern University
gkaramemis@georgiasouthern.edu

Dara Schniederjans
schniederjans@uri.edu

ABSTRACT

We apply a multi-method design to investigate the relationship between blatant benevolence and social capital attainment on SNSs. Our first study identifies the normal forms of blatant benevolence and the moderators of its impact on social capital attainment. Our second study empirically tests the hypothesized relationships. We provide evidence that blatant benevolence increases relational and structural social capital attainment, and the relationship changes under certain conditions. Not only does this manuscript make a theoretical contribution but also provides insight to SNS users to increase their social capital. Charity organizations can also make use of the results to promote prosocial behavior online.

KEYWORDS: Blatant Benevolence, Social Capital, Multi-method design

DECISION SCIENCES INSTITUTE

Binary Programming for Student Semester Scheduling

Hadi Farhangi
Savannah State University
Email: farhangih@savannahstate.edu

Jonathan Lambright
Savannah State University
Email: lambrij@savannahstate.edu

ABSTRACT

This work evaluates a program map and schedules students' semesters for the Global Logistics program at Savannah State University, where a program map is a semester by semester coursework of students. We develop a binary programming model that returns the program map for each semester considering all the restrictions such as prerequisites, semester offering, maximum number of courses per semester, and etc. We then consider the case of advising, in which a student has passed a few courses and wants to schedule the rest of his/her semesters. The analysis of our methods shows that our models can validate the program map and schedule the semester of students.

KEYWORDS: Program Map, Binary Programming, Semester Scheduling

INTRODUCTION AND LITERATURE REVIEW

Program maps and students advising are important tools available for administrators or faculty that can increase the success rate of students (Dills and Hernandez-Julian, 2008). This paper evaluates the program map and provides a decision support tool for university administrations to evaluate the program map and advise students more efficiently in their semester course scheduling. Note that by program map, we refer to the initial semester by semester schedule of the entire program of study for students. And, by advising, we refer to the scheduling of the remaining students' semesters. These tools are particularly important as many of students may overlook the on-time graduation and other students may violate prerequisite requirements; hence, advising students on course selection will help them to graduate on time. Advising, however, can become a burden. For example, advisors check for every course student passed every semester, prerequisites, availability of courses, and other requirements. To reduce the burden on advisors, this work makes the advising process easier, in which advisors can bring up the past and current courses of students and use the tools, introduced in this work, to schedule the remaining semesters considering all the degree requirements.

Mathematical programming, in general, and binary programming in particular, has been used to schedule various kinds of activities in a university. For example, Mehta (1981) uses graph coloring problem, a binary programming problem, to schedule course examination and Mooney et al. (1996) uses binary programming to schedule classes at Purdue university. In this study, however, we use binary programming to evaluate the program map and create an advising tool. Similar works can be found in the literature that use mathematical programming approaches for creating a program

map and scheduling the remaining semesters. For example, Kumar (2017) assign courses to semesters considering different time slots to schedule the remaining courses for each student, called individual program maps. They solved the proposed problem using Spreadsheet tools. This work differs from our study at which we account for all the unique requirements in the program, we disregard different available time-slots, and we solve the problem using Cplex solver. In addition, we consider all the prerequisites and availability of courses in our formulations.

The work of Winch and Yurkiewicz (2013, 2014) consider the assignment of courses for the next semester. Our study is different from these works in three points. First, we develop a program map. Second, we schedule the entire semesters for a given student, considering the remaining courses. Third, all the prerequisites are formulated mathematically in the presented models. It is worth mentioning that the work of (Feldman and Golumbic, 1990) consider the semester planning as a constraints satisfiability problem and create schedules considering different offerings of each course and priority restrictions. Our study considers all offerings of one course as a single course and puts the decision of selecting one of those offerings on students. In addition, we consider all requirements of the program such as prerequisites as constraints that must be feasible in the final recommended schedule.

This paper contributes to the literature by providing mathematical formulations that generate a feasible program map for the Logistics program that satisfies the unique requirements in the College of Business Administration, Savannah State University. The second contribution of the paper is the application of our method for advising students in this program. In the problem formulation section, we provide two formulations for the problem. First formulation is suitable to devise a program map. We compare the results with the current four-year map utilized by the program to evaluate the current degree map and check the correctness of our method. The second formulation is designed for advising the current students when the prerequisite violation is common among current and transfer students. Numerical analysis section provides a short numerical analysis of the models and conclusion section summarizes the research and hints on possible future directions of this research.

PROBLEM FORMULATION

Let $I = \{1, 2, ..., |I|\}$ be the set of courses ($|I|$ is the number of available courses) and let $T = \{1, 2, ..., |T|\}$ be the set of semesters ($|T|$ is the number of available semesters). We next define several subsets of these two sets. Let J be the set of courses that have prerequisite courses and $I_i, \forall i \in J$ be the set of courses that are prerequisite to the course $i \in J$. In addition, let K be a subset of I that represents courses with special double type prerequisites, let I_m be the set of course that must be taken, I_Q be the set of elective courses, I_{o1} and I_{o2} be two set of optional courses that students must select from, I_{AF} be the set of courses in area F of degree work, I_{42} be the set of low level courses that should be completed prior to the advance courses, I_F be the set of courses that are not being offered in the fall, I_S be the set of courses not being offered in the Spring, and I_U be the set of courses not being offered in the summer. Furthermore, let T_F be the set of fall semesters, T_S be the set of Spring semesters, T_{U1} be the set of first sections of summer semesters, T_{U2} be the set of second sections of summer semesters, and $T_U = T_{U1} \cup T_{U2}$ be the set of summer semesters.

We define the objective coefficient of course i in semester t by w_i^t. Moreover, N is the number of courses students prefer to take in either fall or Spring semesters, NU_1 is the number of courses students prefer to take in the first section of summer semesters, and NU_2 is the number of courses

students prefer to take in the second section of summer semesters. Finally, we define decision variables. Variable $x_i^t = 1$ if course i is assigned to semester t, otherwise $x_i^t = 0$. The summary of notation is shown in Table 1. Given these sets, parameters, and variable, we formulate the problem as follows:

$$\textbf{P}: \quad \min \quad \sum_{t\in T}\sum_{i\in I} w_i^t x_i^t \tag{1}$$

$$\text{s.t.} \quad \sum_{t\in T} x_i^t \le 1 \qquad \forall i \in I \tag{2}$$

$$\sum_{t\in T}\sum_{i\in I_m} x_i^t = 37 \tag{3}$$

$$\sum_{t\in T}\sum_{i\in I_Q} x_i^t = 1 \tag{4}$$

$$\sum_{t\in T}\sum_{i\in I_o} x_i^t = 2 \qquad I_o = I_{o1} \text{ or } I_o = I_{o2} \tag{5}$$

$$\sum_{i\in I} x_i^t \le N \qquad \forall t \in T_F \cup T_S \tag{6}$$

$$\sum_{i\in I} x_i^t \le NU_1 \qquad \forall t \in T_{U1} \tag{7}$$

$$\sum_{i\in I} x_i^t \le NU_2 \qquad \forall t \in T_{U2} \tag{8}$$

$$14x_j^{r+1} \le \sum_{t\in\{1,\dots,r\}}\sum_{i\in I_{42}} x_i^t \qquad \forall j \in k, \forall r \in \{1,2,\dots,|T|-1\} \tag{9}$$

$$6x_j^{r+1} \le \sum_{t\in\{1,\dots,r\}}\sum_{i\in I_{AF}} x_i^t \qquad \forall j \in k, \forall r \in \{1,2,\dots,|T|-1\} \tag{10}$$

$$|I_i|x_i^{r+1} \le \sum_{t\in\{1,\dots,r\}}\sum_{j\in I_i} x_j^t \qquad \forall i \in J, \forall r \in \{1,2,\dots,|T|-1\} \tag{11}$$

$$x_i^r \le 1 - \sum_{t\in\{r,\dots,|T|\}} x_j^t \qquad \forall i \in J, \forall j \in I_i, \forall r \in T \tag{12}$$

$$\sum_{t\in T_F}\sum_{i\in I_F} x_i^t = 0 \tag{13}$$

$$\sum_{t\in T_S}\sum_{i\in I_S} x_i^t = 0 \tag{14}$$

$$\sum_{t\in T_U}\sum_{i\in I_U} x_i^t = 0 \tag{15}$$

$$x_i^t \in \{0,1\} \qquad \forall i \in I, \forall t \in T \tag{16}$$

Equation (1) represents total weights of courses that should be minimized. Note that the minimization of this function will reduce the total number of semesters if we increase the weights w_i^t gradually over the semesters. Constraints (2) enforce the model to select each course at most once for the entire study. Constraint (3) enforces the selection of 37 mandatory courses. Constraint (4) enforces that one elective course to be included in the program map. Constraint (5) enforces the selection of one optional path. Constraints (6) limit the number of courses in each fall and Spring semesters to a maximum of N courses. Constraints (7) and (8) restrict the number of courses in section 1 and section 2 of the summer semesters to NU_1 and NU_2, respectively. Constraints (9) and (10) satisfy the prerequisite for three main courses in general business education. Constraints (11) enforce the prerequisite relation between courses. Particularly, if course

$i \in J$ is selected in semester $r+1$, then all prerequisites must be passed in previous semesters, i.e. $\sum_{t\in\{1,\dots,r\}} \sum_{j\in I_i} x_j^t \geq |I_i| x_i^{r+1}$. Constraints (12) guarantee that if course $i \in J$ has the set of prerequisites I_i, then all courses $j \in I_i$ cannot be taken at the same semester as $i \in J$ or in a semester after it. This means x_i^r must be less than equal to $1 - \sum_{t\in\{r,\dots,|T|\}} x_j^t$. Constraints (13) forbid the selection of courses in fall semesters if they are not offered in the fall, constraints (14) forbid the selection of courses in Spring semesters if they are not offered in the Spring, and constraints (15) forbid the selection of courses in summer semesters if they are not offered in the summer. Finally, constrains (16) define variables.

Problem **P** is a binary programming problem and in what comes next, we convert it into its compact form. Assume W is a vector of w_i^t values and X is the vector of decision variables x_i^t. In addition, we define A_1 as the coefficient matrix of inequality constraints and b_1 as right-hand-side of these constraints. Similarly, we define A_2 and b_2 for the equality constraints. Using this notation, we convert problem **P** to its compact form as follows:

$$\begin{aligned} \textbf{P}: \quad & \min \quad WX \\ & \text{s.t.} \quad A_1 X \leq b_1 \\ & \qquad A_2 X = b_2 \\ & \qquad X \in \{0,1\} \end{aligned}$$

An advisor usually advises at the start of each semester and plans the rest of semesters according to the passed/current courses, prerequisites, and requirements in constraints (2)-(16). In addition, students may violate some prerequisites which makes the problem infeasible. Problem **P** cannot be used to address these issues in its current form. First, we convert equality constraints (3)-(5) and (13)-(15) to inequality form. We modify constraint (3)-(5) as follows that guarantees at least 37, 1, and 2 courses will be selected, respectively:

$$\sum_{t\in T} \sum_{i\in I_m} x_i^t \geq 37 \tag{17}$$

$$\sum_{t\in T} \sum_{i\in I_Q} x_i^t \geq 1 \tag{18}$$

$$\sum_{t\in T} \sum_{i\in I_o} x_i^t \geq 2 \tag{19}$$

We also modify constraint (13)-(15) as follows that guarantees no courses will be selected in a semester that they are not being offered:

$$\sum_{t\in T_F} \sum_{i\in I_F} x_i^t \leq 0 \tag{20}$$

$$\sum_{t\in T_S} \sum_{i\in I_S} x_i^t \leq 0 \tag{21}$$

$$\sum_{t\in T_U} \sum_{i\in I_U} x_i^t \leq 0 \tag{22}$$

Using this modification, problem **P** can be written as follows, where A replaces A_1 and A_2; and b replaces b_1 and b_2:

$$\begin{aligned} \textbf{P}: \quad & \min \quad WX \\ & \text{s.t.} \quad AX \leq b \\ & \qquad X \in \{0,1\} \end{aligned}$$

Problem **P** is a binary programming problem and its solutions show a feasible program map. Now, we assume the current term of student is p, X_P shows the courses student passed/have and it is a known vector, X_R is the vector of the remaining variables of X, A_P shows the first p columns of A, A_R shows the remaining columns of A, and W_R shows the objective coefficients of the remaining variables, X_R. Given this, we modify problem **P** based on the unknown variables X_R as the following:

$$\begin{aligned} \widetilde{\textbf{P}}: \quad & \min \quad W_R X_R \\ & \text{s.t.} \quad A_R X_R \leq b - A_P X_P \\ & \qquad X_R \in \{0,1\} \end{aligned}$$

Solving problem $\widetilde{\textbf{P}}$ can schedule the remaining semesters of students given the passed/current courses they have: X_P. In the next section, we use both problems **P** and $\widetilde{\textbf{P}}$ to schedule students' semesters.

NUMERICAL ANALYSIS

In this section, we use the developed problems, i.e. **P** and $\widetilde{\textbf{P}}$, to schedule the program map of Global Logistics program at Savannah State University and create an advising tool. We solve problem **P** for the purpose of scheduling the program map and use problem $\widetilde{\textbf{P}}$ to create an advising tool. Both problems are implemented in Matlab 2016a and solved using Cplex ILOG 128. We set problem parameters as $|I| = 45$ and $|T| = 24$. Note that the total number of available courses is 85, however, we consider courses in the same category as a single course. Following this approach, we are able to reduce the number of courses to only $|I| = 45$ courses. For a maximum of 6 year college period, where we have fall and spring semesters and two short sections for summer semester each year; hence, $|T| = 24$. To create objective coefficients, we set $w_i^t = 0$, $\forall i \in I$, $\forall t \leq 16$, $w_i^{17} = 10$, $\forall i \in I$, $w_i^t = w_i^{t-1} + 10$, $\forall i \in I$, and $\forall t \geq 18$.

Problem **P** has $1,080$ variables and $2,466$ constraints. Cplex solves this problem within a few seconds. The purpose of solving this problem is to validate the current program map of the Global Logistics program. Based on a feasible solution of problem **P**, we recommend a slight modification of the current program map of the Global Logistics program.

Problem $\widetilde{\textbf{P}}$ on the other hand can schedule the remaining semesters of students and therefore can be used for advising. It is common for students to violate prerequisite requirements, especially if they change their major or if they transfer to the program. In this case, problem $\widetilde{\textbf{P}}$ renders infeasible. To overcome infeasibility, we slightly change X_P and constraints of the problem **P**. Recall from the problem formulations that p represent the current semester of students. If the current semester is fall, we set $p = 1$; if current semester is spring, we set $p = 2$; and if current semester is summer,

we set $p = 4$ (due to two sections in the summer). We move all passed/current courses to the first semester, i.e. the first $|I|$ elements of X_P may contain passed courses, where $|I| = 45$ is the number of all available courses. Reducing the number of semesters that students passed/have to $p = 1$ or 2 or 4 does not impact the feasibility of the problem if appropriately addressed. To address that, we only consider the case of $t > p$ for constraints (6), (7), (8), (13), (14), and (15). In addition, we eliminate all passed courses from sets J and K when generating constraints (9), (10), and (11). We do not need constraints (12) for courses in X_P. The rest of constraints do not need modifications. Note, one may reduce number of semesters as well (reducing $|T|$), which results in a lower number of variables. In our implementation, this reduction of semesters is not employed as the algorithm already solves the problem within a few seconds and variable reduction is not needed.

The aforementioned modifications change the matrix A and vector b; hence, problem $\widetilde{\mathbf{P}}$. To check the correctness of our model, we schedule the remaining semesters of five students using problem $\widetilde{\mathbf{P}}$, applying these modifications. Particularly, we use an Excel file to show the passed/current courses of students. Then, we import the data into Matlab as X_P and finally we successfully generate feasible schedules for the remainder of the semester. The resulting schedules satisfy the program requirements for these selected students and show the effectiveness of our model.

One can utilize the above procedure to make advising easier. For example, we can generate an Excel report of the passed/current courses from the student portal and run our codes on the report to assist advisors. Alternatively, one can embed the whole process within the student portal and generate the semester schedule online. Note that using problems **P**, $\widetilde{\mathbf{P}}$, and the above procedure is scalable and general. Hence, these models are applicable to all majors that are being offered at Savannah State University> any change in prerequisites can be reflected by changing sets J and I_i and additional requirements can be captured by adding constraints. Finally, these models can make the semester planning less cumbersome and more user friendly and they can help with the cost reduction if embedded in the student portal.

CONCLUSION AND FUTURE RESEARCH

This research proposes two problem formulations to model the generation of a feasible program map and an advising tool. To generate the program map, we use a binary programming problem, **P**, and we produce a recommendation to modify the current program map. To formulate the advising process we use another binary programming problem, $\widetilde{\mathbf{P}}$, to schedule feasible semesters for students. This problem is tested on the remainder coursework of five students and results show successful schedules. Finally, we suggest embedding our codes within the student portal to make the advising process easier to use, more user friendly, and less costly.

References

Dills, A. K. and Hernandez-Julian, R. (2008). Course scheduling and academic performance. *Economics of Education Review*, 27:646–654.

Feldman, R. and Golumbic, M. C. (1990). Optimization algorithms for student scheduling via constraint satisfiability. *The Computer Journal*, 33(4):356–364.

Kumar, R. (2017). A spreadsheet-based scheduling model to create individual graduation roadmaps. *Journal of Supply Chain and Operations Management*, 15:165–188.

Mehta, N. K. (1981). The application of a graph coloring method to an examination scheduling problem. *Interfaces*, 11(5):57–65.

Mooney, E. L., Rardin, R. L., and Parmenter, W. J. (1996). Large-scale classroom scheduling. *IIE transactions*, 28(5):369–378.

Winch, J. K. and Yurkiewicz, J. (2013). Student class scheduling with linear programming. In *Proceedings for the Northeast Region Decision Sciences Institute Annual Meeting*, pages 374–386.

Winch, J. K. and Yurkiewicz, J. (2014). Case article-class scheduling with linear programming. *INFORMS Transactions on Education*, 15:143–147.

Appendix: Notation

Table 1: Summary of Notation

List of Sets and Indices	
I	: the set of courses $I = \{1, 2, ..., \|I\|\}$
i, j, k	: indices used for courses, $i, j, k \in I$
K	: a subset of I that represents courses with special double type prerequisites
J	: the set of courses that have prerequisites
I_i	: the set of courses that are prerequisite for course $i \in J$
I_m	: the set of course that must be taken
I_Q	: the set of elective courses
I_o	: the set of optional courses that students select one; $I_o = I_{o1}$ or $I_o = I_{o2}$
I_{AF}	: the set of courses in area F of degree work
I_{42}	: the set of low level courses that should be completed prior to advance courses
I_F	: the set of courses not being offered in the fall
I_S	: the set of courses not being offered in the spring
I_U	: the set of courses not being offered in the summer
T	: the set of semesters $T = \{1, 2, ..., \|T\|\}$
T_F	: the set of fall semesters
T_S	: the set of spring semesters
T_{U1}	: the set of first sections of summer semesters
T_{U2}	: the set of second sections of summer semesters
T_U	: the set of summer semesters $T_U = T_{U1} \cup T_{U2}$
Parameters	
w_i^t	: the objective coefficient of course i in semester t
W_R	: the objective coefficient of a remaining course i in semester t
N	: the number of courses students prefer to take in either fall or spring semesters
NU_1	: the number of courses students prefer to take in the first section of summer semester
NU_2	: the number of courses students prefer to take in the second section of summer semester
Variables	
x_i^t	: $x_i^t = 1$ if course i is assigned to semester t, otherwise $x_i^t = 0$
X	: the set of courses students passed
X_R	: the vector of variables that shows the remaining courses of students

DECISION SCIENCES INSTITUTE

Optimizing demand forecasting methods considering bullwhip effects and total costs in a decentralized supply chain

Erika Tajima
Tokyo University of Science
Email: 7419522@ed.tus.ac.jp

Aya Ishigaki
Tokyo University of Science
Email: ishigaki@rs.tus.ac.jp

Ryuta Takashima
Tokyo University of Science
Email: takashima@rs.tus.ac.jp

Hajime Nishida
Canon IT Solutions Inc.
Email: nishida.hajime@canon-its.co.jp

ABSTRACT

In the supply chain, the bullwhip effect, in which order fluctuation increases from downstream to upstream, imposes a significant burden on corporate management. This study constructs an information decentralized supply chain model using a multi agent system. This model is used to analyze the strategies of each company to reduce the bullwhip effect and total cost, thereby increasing the supply chain benefits. This study conducts a simulation analysis under each condition, including external factors such as final demand and the presence of cooperation between companies. The trend of the best strategy is suggested to the company depending on the situation.

KEYWORDS: Supply chain management, Multi agent system, Demand forecast, Bullwhip effect, Beer game

INTRODUCTION

In recent years, competition between companies has intensified, and companies need to deliver produced products to customers within short delivery time. On the other hand, it is very difficult for business managers to efficiently produce and transport products in an unstable market due to diversification of customer needs. Therefore, supply chain management, which is not a logistics system limited to some companies, but is to build an integrated logistics system between multiple companies and improve management results has attracted attention. The supply chain is a series of flows from the time a product is manufactured until it reaches a customer. It is well known that there is a bullwhip effect in a supply chain in which order fluctuations increase from downstream to upstream. This phenomenon was observed by Forrester (1958) and found to increase costs and risks, reduce production efficiency, and put a heavy burden on corporate management (Lee et al., 1997a; Lee et al., 1997b; Wang & Disney, 2016). In addition, upstream companies in the supply chain are unable to accurately grasp market fluctuations because of the distorted information on final demand, making it extremely

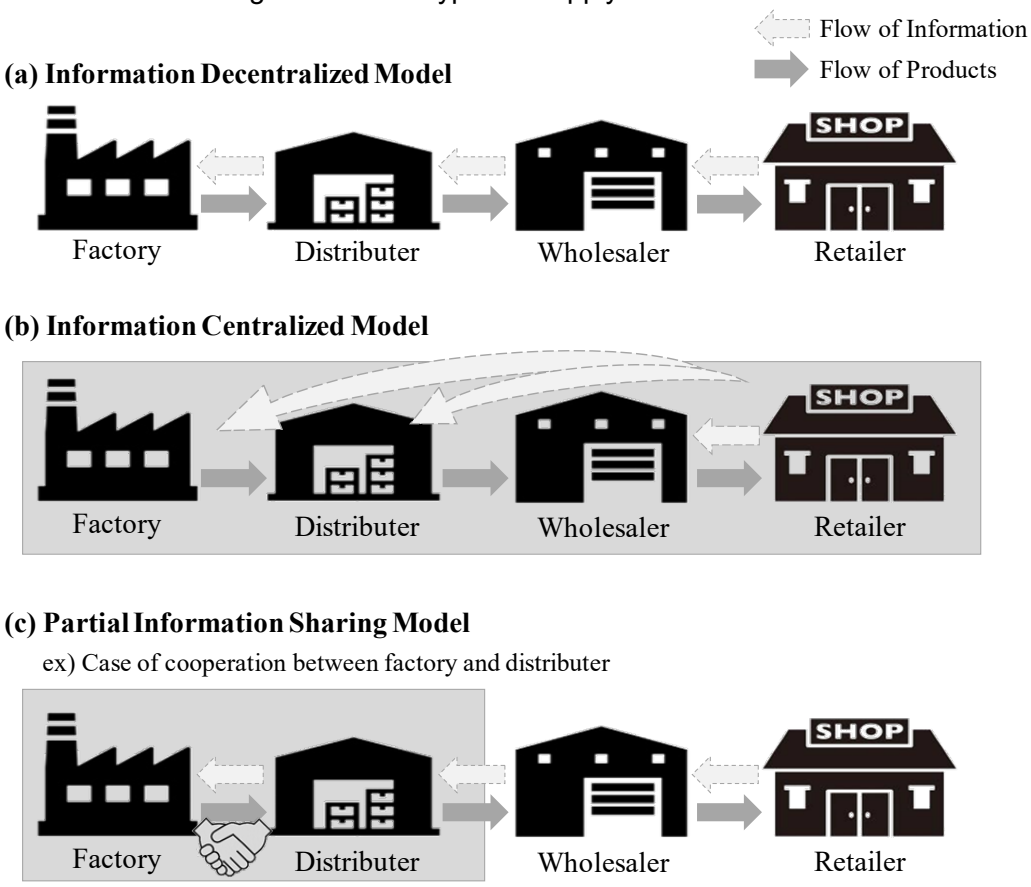
difficult for them to respond flexibly and agilely. In this way, the competitiveness of the entire supply chain is significantly reduced. Therefore, it is important to control the bullwhip effect in order to improve the competitiveness of the entire supply chain.

There are four main causes of the bullwhip effect: accuracy of demand forecast, lead time, self-centered behavior of each company, and demand variability (Chen et al., 1999; Chen et al., 2000a). The accuracy of demand forecast depends on the sharing of information between companies in the supply chain. Generally, there are two types of supply chain models: information decentralization, in which no information about final consumer demand is shared by companies other than the most downstream, and information centralization, in which final demand is shared by all companies. Many studies have shown that information centralization can reduce the bullwhip effect by controlling the supply chain in a top-down manner. On the other hand, many companies deal with multiple companies and multiple products simultaneously. As a result, the relationships between companies have become so complex that it is difficult to accurately share final demand information across the supply chain without delay. Therefore, this study considers an information decentralized model and a partial information sharing model where demand information is shared among some companies in the supply chain (Figure 1). This method is also effective in reality because information can be shared easily as long as there is an agreement between two companies. In addition, it is expected that the bullwhip effect caused by the partial-optimal behavior of each company can be reduced by the commonality of strategies through cooperation among some companies. In this study, a model is constructed for the case of cooperation such as information sharing among some companies in a multi-stage supply chain, and effective strategies to reduce the bullwhip effect are analyzed. It is known as methods to reduce the bullwhip effect that inventory policies, demand forecasting methods, and safety stock optimization, there are many studies on its contents. It is necessary to analyze whether these indicators are effective for the benefit of the entire supply chain. One of these analysis methods is known as the beer game. It's a simulation game invented by Jay Forrester, and it's a great way to learn about the value of information sharing and supply chain management. Furthermore, it is possible to evaluate the bullwhip effect and total cost for the different strategies of each company in the supply chain. Therefore, this study uses a beer game to analyze the effect of system structure on the decision making of each company. Several studies have examined supply chain management using beer games.

Zhang (2004) discussed the effect of demand forecasting methods on the bullwhip effect. Ponte et al. (2017) constructed a decentralized supply chain model using multi-agent system. Using this model, the optimal combination of demand forecast of each company to reduce the bullwhip effect was clarified. However, these studies only aim at reducing the bullwhip effect and do not investigate the total cost of the entire supply chain. Since the reduction of the bullwhip effect does not necessarily lead to the cost reduction of the entire supply chain, it is desirable to evaluate from two viewpoints: the reduction of the bullwhip effect and the total cost in the supply chain.

In this study, an information decentralized supply chain model and a partial information supply chain model were constructed using a multi-agent system and analysis of how changing strategies for each company will reduce the bullwhip effect and total cost in the supply chain. The trends of optimal strategies for changing the cooperation of each company are analyzed and the impact of cooperation between companies on supply chain performance is revealed.

Figure 1: Three types of supply chain models.



LITERATURE REVIEW

In order to reduce the bullwhip effect, demand forecasting, ordering system and safety stock optimization are effective. Croson et al. (2014) and Ribeiro et al. (2018) found that the uncertainty of the behavior of each company in the supply chain leads to order instability and causes the bullwhip effect. Thus, it is important to adjust the behavior of companies in the supply chain to reduce the impact of their independent, suboptimal behavior in order to reduce the bullwhip effect. Chen et al. (1999), Chen et al. (2000a), Chen et al. (2000b) and Simchi-Levi et al. (2008) formulated the bullwhip effect for a model in which each company independently sets the order processing and replenishment lead time in a two-stage supply chain. Bhattacharya & Bandyopadhyay (2011) showed the effect of changing the demand forecasting method used by the entire supply chain. Trapero et al. (2012) analyzed the effect of sharing demand information on the accuracy of demand forecasting in a two-stage supply chain model. Jaksic & Rusjan (2008) found that inventory policy has an effect on the occurrence of the bullwhip effect. These studies assume a two-stage supply chain that is less affected by the bullwhip effect. However, in the actual supply chain, a product often goes through several companies before it reaches the final consumer, so this study constructs a model for a multi-stage supply chain. Furthermore, few studies have considered the management policies and thinking of firms in the supply chain. Therefore, this study also considers what policies are beneficial to act on in accordance with the characteristics of the company.

Zhang (2004) discussed the effect of demand forecasting methods on the bullwhip effect. Ponte et al. (2017) used a multi-agent system to construct a decentralized supply chain model, reproducing a situation in which each company could choose its own demand forecasting method. This was used to identify the optimal combination of demand forecasts for each company to control the bullwhip effect. Croson et al. (2014) used a beer game to analyze the optimal inventory policy to control the bullwhip effect. Khan et al. (2019) used a beer game to analyze the effect of personality and experience of participants on the outcome of the game. These studies analyze the impact of the differences and characteristics of each firm's strategy in the supply chain on the bullwhip effect and do not investigate the impact on the total cost of the entire supply chain. Furthermore, communication between companies in the supply chain strengthens connections between them and improves the performance of the entire supply chain (Yoon et al., ,2011; Shnaiderman et al., 2014; Ganesh et al., 2014; Abdullah & Musa., 2014). No studies, to our knowledge, have analyzed the effect of cooperation among some companies in a multi-stage supply chain on the performance of the entire supply chain. Studies using beer games have focused on the success on the game and have not considered whether they are effective in the actual supply chain.
In this study, a multi-stage supply chain model for a beer game is constructed using a multi-agent system, and effective strategies are analyzed in a real situation. As a result, the management policy and thinking of the companies in the supply chain are considered to improve the performance of the entire supply chain. In addition, the effects of cooperation between some companies on the performance accuracy of the entire supply chain are analyzed. In this study, two measures of bullwhip effect and total cost are used to evaluate each strategy.

METHODOLOGY

A model in which products are traded in series by multiple companies is called a multi-stage model. This study focuses on a model in which orders are placed in four stages, such as factory →secondary wholesale → primary wholesale → retail → customer. Each company orders products using two types of demand forecasting methods (moving average forecasting, exponential smoothing forecasting, and simple forecasting using the proportional periodic ordering method). The ordering method for each company is a regular ordering method, and there is no restriction on the number of products that each company can possess (buffer restriction). This study experiments with the decentralized information model and the partial information sharing model. The definition of the model is given below.

Decentralized Information Model
In the information-distributed model, each player receives the amount of orders from a player one step downstream as the amount of demand, which is used as the basis for forecasting demand and setting the target inventory. In other words, they place an order only with reference to the demand information they receive.

Partial Information Sharing Model
In the partial information sharing model, two of the four players cooperate and share demand information. The upstream players among the cooperating players forecast demand using the demand received by the downstream players. Upstream players can forecast demand using information that is closer to final demand. The partial information sharing model is effective in reality because information can be shared easily as long as there is an agreement between two companies.

Company Behavior in This Model

In this study, agents can perform the following actions.

Reception State

Products are delivered from the upstream process and ordered from the downstream process. The delivered products are added to the available on-hand inventory.

Serving State

Orders and backlogs (products that have not yet been shipped in past orders) are shipped from on-hand inventory.

Updating State

On-hand inventory (net inventory) and work in process (order inventory) are updated. If there is an order backlog, the order backlog is updated.

Sourcing State

Product demand forecast is performed and the order-up-to point is determined. An order is issued to the upstream process.

In this study, each agent places an order using the regular order method. Each agent can freely change the demand forecasting method used to order products.

Mathematical Model

The actions of the players are formulated. Table 1 shows the model notation. First, the index parameters are explained. N indicates the total number of echelons in the supply chain, and n indicates the echelon in the supply chain. Here, $N = 4$ (factory: $n = 1$; distributer: $n = 2$; wholesaler: $n = 3$; retailer: $n = 4$).

Table 1: The model notation.

n	Echelon in the supply chain	N	Total number of echelons	t	Time perood
T	Time horizon of the test	I_n	On-hand inventory (net stock)	S_n	Shipment (units of products delivered)
B_n	Backlog	W_n	Work-in-progress (on-order inventory)	R_n	Receipts (units of products received)
D_n	Demand (purchase order received)	O_n	Purchase order issued	Y_n	Order-up-to point
SS_n	Safety stock	F_n	Demand forecast	l	Lead time
a	Order interval	m	Moving average periods	k	Proportional controller
f	Smoothing factor	BE_n	Bullwhip ratio	p	Inventory storage cost (per unit)
q	Stock-out cost (per unit)	C_n	Total cost		

Next, the mathematical formulation of the supply chain model is explained using the variables and parameters in Table 1. The on-hand inventory is calculated as shown in equation (1) by the cumulative difference between the amount of receipts from upstream and the amount of shipped

products downstream, in addition to the inventory on hand in the previous term. The shipment volume is calculated as shown in equation (2) by the demand in consideration of the backlog and the minimum value of the net stock after the products are delivered from the upstream. The order backlog is the difference between demand from downstream and shipment, and is expressed as equation (3). Work-in-progress inventory (on order inventory) is obtained as shown in equation (4) by accumulating the difference between the amount of purchase order issued in the previous term and the amount of shipped products in the current term.

$$I_n(t) = I_n(t-1) + R_n(t) - S_n(t) \quad (1)$$

$$S_n(t) = \min\{D_n(t) + B_n(t-1),\ \ I_n(t-1) + R_n(t)\} \quad (2)$$

$$B_n(t) = B_n(t-1) + D_n(t) - S_n(t) \quad (3)$$

$$W_n(t) = W_n(t-1) + O_n(t-1) - R_n(t) \quad (4)$$

Both on-hand inventory and backlog are always 0 or more. If the available inventory exceeds the shipment, the net inventory is positive and the order backlog is 0. On the other hand, if the available amount is not enough, the order backlog is positive and the inventory is 0. The amount of products received is the shipment volume of the company one upstream $\boldsymbol{l}$ period before, which is expressed as equation (5). However, the receipts of the factory is equal to the order quantity of the factory $\boldsymbol{l}$ term ago.

$$R_n(t) = S_{n-1}(t-l) \quad (5)$$

The demand is the order one step downstream in equation (6). However, in the partial information sharing model, the quantity demanded by the upstream players among the cooperating companies is the quantity demanded by the downstream players ($\boldsymbol{n} = \boldsymbol{k}$) in equation (7).

$$D_n(t) = O_{n+1}(t) \quad (6)$$

$$D_n(t) = D_k(t) \quad (7)$$

In the order-up-to policy, the order is placed to completely recover the difference between the order-up-to point and the amount of products, so the amount of products ordered is expressed as equation (8). The amount of order-up-to point considers demand forecast, safety inventory, and lead time as in equation (9).

$$O_n(t) = \max\{Y_n(t) - [I_n(t) - B_n(t) + W_n(t)], 0\} \quad (8)$$

$$\begin{aligned} Y_n(t) &= a * F_n(t) + SS_n + (l-1) * F_n(t) \\ &= l * F_n(t) + SS_n \end{aligned} \quad (9)$$

Two methods of demand forecasting, moving average forecasting and exponential smoothing, are used in this study. The method of using the past m average demands as forecast is called moving average forecast. The demand forecasting method in this case is as shown in equation (10), and the method of determining the amount of products ordered using equation (6) or (7), (8), and (10) is model 1.

$$F_n(t) = \frac{1}{m}\sum_{i=0}^{m-1} D_n(t-i) \tag{10}$$

Next, exponential smoothing prediction is explained. In this method, the order-up-to point is updated periodically by considering the difference between the two latest demands multiplied by the smoothing factor. The inventory of order-up-to point in this case is calculated as shown in equation (11), and the method of determining the order quantity using equation (6) or (7), (8) and (11) is model 2.

$$F_n(t) = F_n(t-1) + f * [D_n(t-1) - F_n(t-1)] \tag{11}$$

Performance Metrics

The bullwhip effect is known as an index for evaluating the performance in the supply chain (Disney et al., 2006). In this study, we use the bullwhip effect and the total cost in the supply chain as performance metrics. The Echelon n bullwhip effect is defined as the ratio of the variances of the order quantity and the demand quantity, and both the numerator and the denominator are adjusted as the average value as in equation (12) (Cannella et al., 2013). This is called the bullwhip ratio. Since the amount of products ordered of each company is the demand from one upstream, the total bullwhip ratio is expressed as the product of the four bullwhip ratios as shown in equation (13).

$$[BE_n]_{t=1}^{t=T} = \left[\frac{\frac{Var(O_n)}{Avg(O_n)}}{\frac{Var(D_n)}{Avg(D_n)}}\right]_{t=1}^{t=T} \tag{12}$$

$$\begin{aligned}[BE_{SC}]_{t=1}^{t=T} &= \left[\frac{\frac{Var(O_K)}{Avg(O_K)}}{\frac{Var(D_1)}{Avg(D_1)}}\right]_{t=1}^{t=T} \\ &= \prod_{n=1}^{n=K} [BE_n]_{t=1}^{t=T}\end{aligned} \tag{13}$$

The total cost of the supply chain is calculated as shown in equation (14) using the sum of the stock storage cost and stock-out cost of each company.

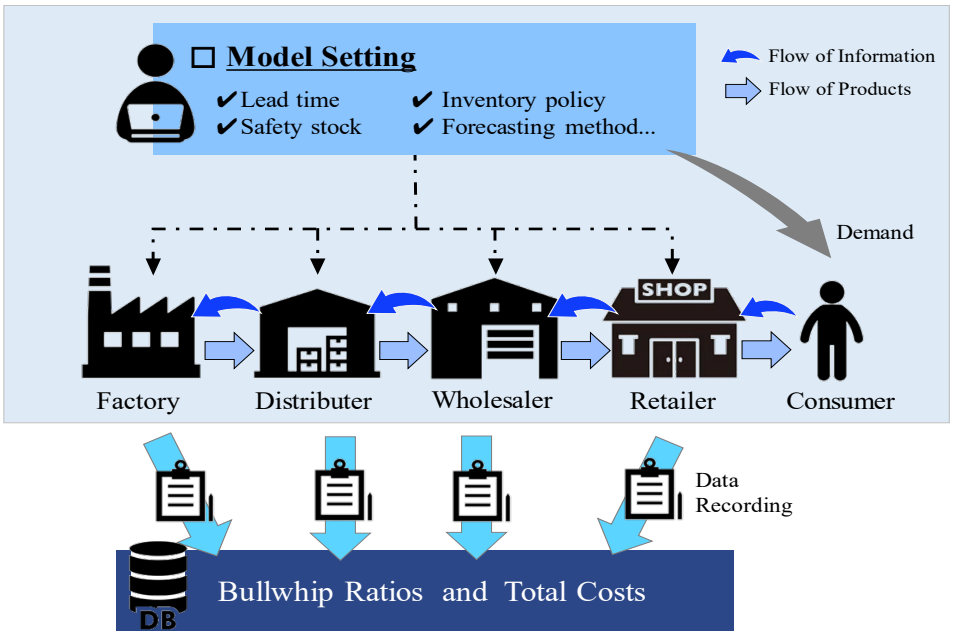
$$C_{sc} = p\sum_{i=1}^{N}\sum_{t=1}^{t=T} I_i(t) + q\sum_{i=1}^{N}\sum_{t=1}^{t=T} [D_i(t) - S_i(t)] \tag{14}$$

Multi Agent Model

In this study, a multi-stage supply chain model is constructed using a multi-agent system (Figure 2). Each player is represented as an agent and records data while acting according to the

procedures defined in section Mathematical Model. The simulation is performed for each scenario up to $T = 500$ periods.

Figure 2: System Diagram of this Model.



SIMULATION EXPERIMENT

Experiment Plan

In this simulation experiment, consumers' final demand is determined by three ranges of variation with a uniform distribution U(70, 90), U(60,100), and U(40, 120) with an average of 80. The lead time is set to l=2 and the order interval is set to a=1. In this study, the following two cases were analyzed. In Case 1, the decentralized information model is used, and in Case 2, the partial information sharing model is used.

Case 1: No information sharing among all players, but a common strategy among some players in the supply chain.
Case 2: Partial information sharing and common strategy among some players in the supply chain.

In this study, the four players in the supply chain were divided into upstream (factory and distributer), downstream (wholesaler and retailer), inside (distributer and wholesaler), and outside (factory and retailer), and the experiment was conducted assuming that each of them established a cooperative system.

Strategy Type
In this study, two strategy types are evaluated: stable and ascending. Let $p = 1$ yen per unit for inventory storage cost per day, and $q = 5$ yen per unit for out-of-stock cost per day. Out of stock carries not only the risk of lost opportunities but also the risk of losing the trust of customers. Therefore, it seems reasonable to use a higher cost than inventory storage costs to

evaluate them. In order to reduce the total cost, each player must behave as little as possible to avoid running out of stock, and thus overstocking. Therefore, the stable strategy in this study is to set a high safety stock to reduce the frequency of out-of-stocks, while the upward strategy is to set a low safety stock to reduce the excess stock.
In addition, two demand forecasting methods are used in this study: moving average forecasting and exponential smoothing forecasting. Moving-average forecasting can be considered as a stability-oriented strategy because it does not get distracted by short-term changes in demand and makes calm judgments about the situation. The exponential smoothing forecast can be considered as an upward-looking strategy because it does not overlook even small changes in demand and can be adapted flexibly to the situation while taking risks. Thus, stable inventory management is achieved by using moving average forecasting for stable strategies and exponential smoothing forecasting for upward strategies to allow for a flexible response. This study wants to evaluate the superiority of the strategy as a purely strategic trend, not as a result of the numerical value of the parameters. Therefore, the moving average period m, the exponential smoothing coefficient f and the safety stock were determined as follows. An experiment was conducted assuming that all players in the supply chain used the same demand forecasting method and safety stock. From the results, the parameter with the closest bullwhip ratio was determined for each scenario (Table 2).

Table 2: The two types of strategies in this study.

		Stable Strategy	Upward Strategy
Demand Forecasting Method		Moving Average Forecasting	Exponential Smoothing Forecasting
m or f	U(70,90)	10	0.36
	U(60,100)	12	0.34
	U(40,120)	16	0.32
Safety Stock	U(70,90)	40	5
	U(60,100)	60	10
	U(40,120)	80	20

Result

Effects of Commonality of Strategy Between Companies

The simulation results for Case 1 are shown in Table 3. The overall average is the average of the results for each player's independent strategies. In Case 1, it was found that regardless of the range of variation in final demand, the bullwhip effect and total costs could be reduced by using specific strategies through cooperation among the players compared to the overall average. In other words, it can contribute to the benefit of the entire supply chain by cooperating among the players than if each player independently formulates a partially optimal strategy. When upstream players collaborated, an upward strategy reduced the total cost of the entire supply chain (Figure 3a). The advantage of exponential smoothing forecasts is that they are a short-term strategy and always compare the forecasted and actual values of the previous period to correct the trajectory of the demand forecast, so they can react quickly to small changes. Therefore, it is likely that upstream players could reduce excess inventory by using exponential smoothing predictions and low safety stocks to manage their inventory by identifying the amount

of inventory needed while reacting quickly to small changes. On the other hand, when downstream players collaborate, a stable strategy reduces the total cost of the entire supply chain (Figure 3b). Moving-average forecasting is a long-term strategy, which has the advantage of less susceptibility to recent large demand fluctuations and more stable ordering. Therefore, the use of moving average forecasts and high safety stocks by downstream players could make calm decisions without being distracted by fluctuations in consumer demand, and this could help to control the source of order variability and reduce shortages throughout the supply chain. Furthermore, when the inside players were cooperative, the use of a stable strategy reduced the bullwhip effect and total costs (Figure 4a). This factor may be due to the use of moving average forecasting and high safety stock by the inside players, which stabilized the core of the supply chain and led to the stability of the overall inventory management. When the outside players were cooperative, the use of upward-looking strategies reduced the bullwhip effect and total costs (Figure 4b). The use of exponential-smoothing forecasts by outside players allows retailers to instantly reflect recent demand trends within the supply chain. In factories where inventory control is considered to be very difficult, this system can quickly respond to demand from downstream. These factors are considered to be the reasons. Furthermore, the low safety stock at retailer may have reduced excess orders and led to stability in inventory management for upstream players.
Based on the above, it can be said that the performance of the entire supply chain can be improved when the companies in the supply chain share strategies with other companies according to their respective characteristics.

Table 3: Bullwhip ratios and total costs in each scenario (Case 1).

Demand of Customer		Overall Average	Upstream		Downstream		Inside		Outside	
			Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy
[70,90]	Bullwhip rat	9.50	9.77	9.62	9.63	9.21	**9.47**	9.56	9.94	**9.29**
	Total cost	3,705,101	4,673,536	**3,204,733**	**3,473,002**	3,649,581	**3,196,309**	3,833,810	4,925,780	**2,996,054**
[60,100]	Bullwhip rat	9.39	9.67	9.64	9.41	9.43	**8.87**	9.83	10.07	**9.10**
	Total cost	4,107,992	4,799,098	**3,682,930**	**3,188,239**	5,049,016	**2,745,055**	4,897,037	4,788,899	**3,381,526**
[40,120]	Bullwhip rat	9.34	9.45	9.95	9.49	9.31	**8.84**	9.81	9.49	**8.84**
	Total cost	4,303,183	4,818,731	**4,212,524**	**4,172,787**	4,590,482	**3,921,694**	5,035,168	4,627,009	**3,626,587**

Figure 3: Cooperation between upstream or downstream.

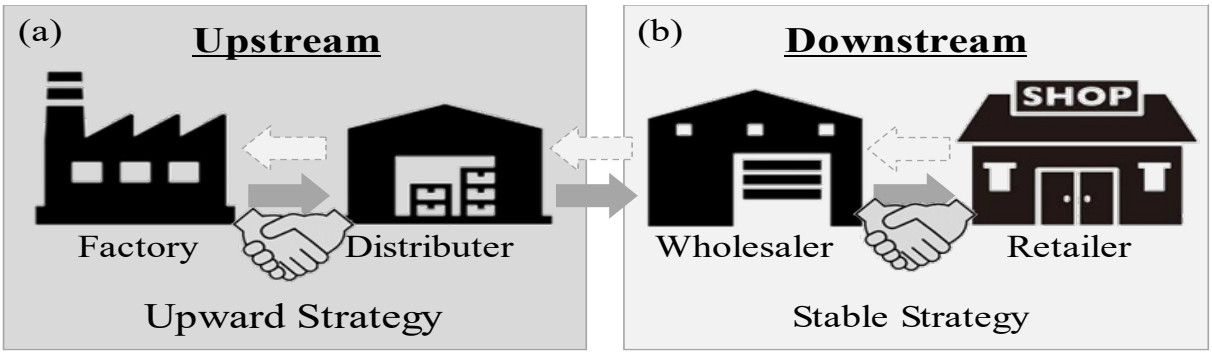
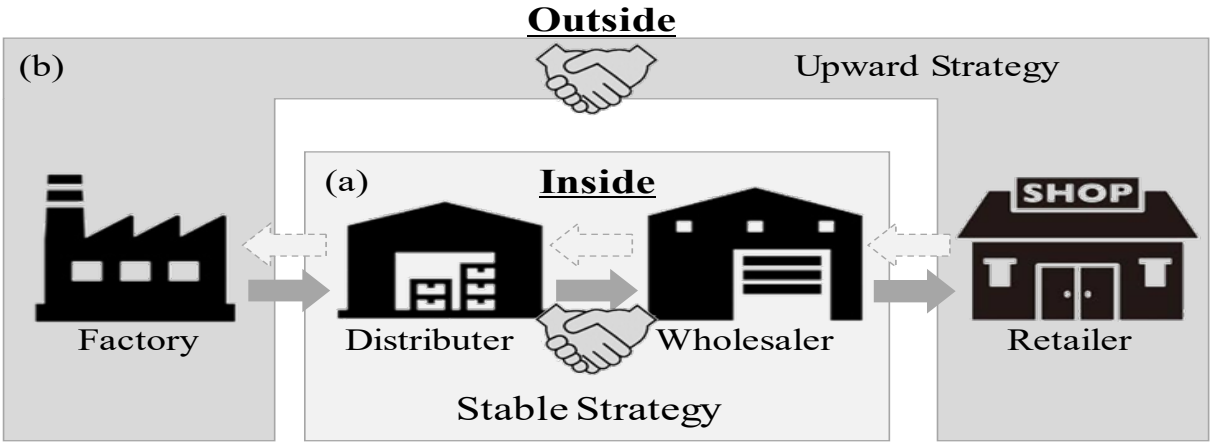


Figure 4: Cooperation between inside or outside.



Effects of Partial Information Sharing and Commonality of Strategy

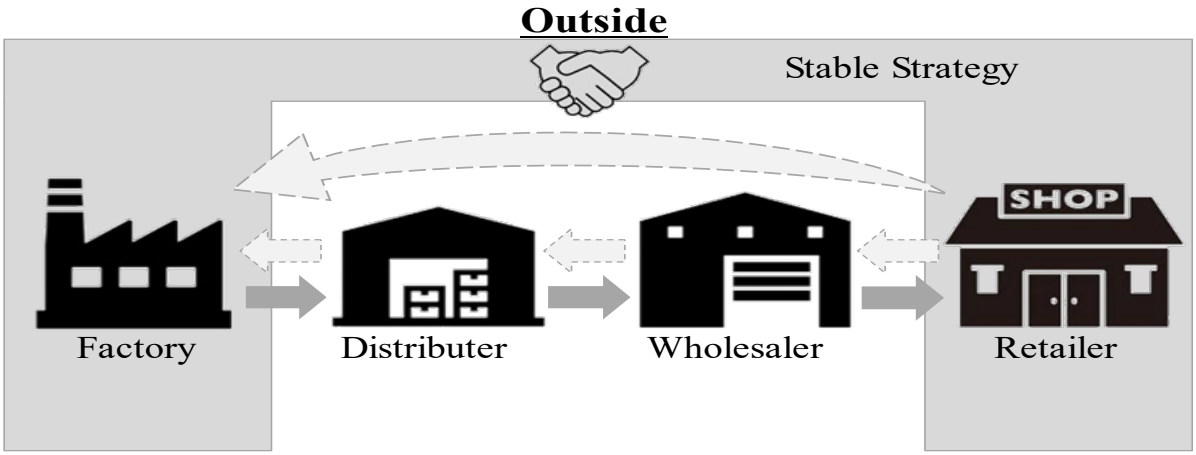
The simulation results for Case 2 are shown in Table 4. The simulation results show that information sharing among some players in the supply chain is effective in improving the performance of the entire supply chain.
The results show that the same strategy as in Case 1 can be used to reduce the bullwhip effect and total cost when cooperation is established between upstream, downstream and inside players. It was found that the optimal strategy for cooperation while sharing demand information among the outside players depended on the magnitude of variation in final demand.
The same results as in case 1 are obtained when final demand follows a uniform distribution U(70,90) with relatively low variability. On the contrary, in the case of large demand variability, such as uniformly distributed U(60,100) and U(40,120), the outside players were able to reduce the bullwhip effect and total costs by using a stable strategy (Figure 5). The use of upward-looking strategies by the outside players has the effect of allowing retailers to instantly incorporate demand trends throughout the supply chain. In addition, it has the advantage of allowing factories, which are considered to be very difficult to manage inventory, to respond quickly to downstream situations. When information sharing is not possible or the demand variability is small, an upward-looking strategy is useful. However, the use of exponential smoothing forecasting when the range of demand variability is large may cause variability in the order quantity due to temporary demand fluctuations. At this time, the ability to share information allowed factories, which were considered to be very difficult to manage, to take in changes in demand quickly and to have more flexibility in inventory management. Therefore, stable orders among the outside players were prioritized, and the use of moving average forecasting and high safety stock contributed to improving the performance of the entire supply chain.

Determining the Optimal Strategy

Table 4: Bullwhip ratios and total costs in each scenario (Case 2).

Demand of Customer		Overall Average	Upstream		Downstream		Inside		Outside	
			Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy	Stable Strategy	Upward Strategy
[70,90]	Bullwhip rat	5.03	4.02	**3.82**	**3.83**	5.34	**5.00**	9.49	4.36	**4.17**
	Total cost	81,024	81,093	**79,079**	**76,806**	78,902	**80,680**	83,216	82,359	**79,827**
[60,100]	Bullwhip rat	4.51	3.92	**3.70**	**3.39**	4.94	**3.95**	8.98	**3.04**	4.36
	Total cost	115,262	129,779	**104,984**	**104,359**	116,326	**110,713**	129,062	**111,678**	119,141
[40,120]	Bullwhip rat	4.09	3.48	**3.28**	**3.23**	4.48	**3.62**	7.48	**2.91**	4.05
	Total cost	234,758	232,838	**206,357**	**203,391**	237,529	**222,756**	283,037	**221,303**	244,373

Figure 5: Cooperation between outside (Case 2).



Consideration

The results of the simulation experiments show that commonality of strategy and information sharing among some companies in the supply chain can reduce the bullwhip effect and total cost of the entire supply chain. It is important for each company to use appropriate strategies according to the characteristics of the company, the status of cooperation with other companies, and the demand situation. The stable strategy in this study uses moving average forecasting and high safety stock, which tends to limit the risk of order instability and out of stock. Therefore, companies that are optimal in using stable strategies need to strengthen their stability-oriented strategies and train their human resources. On the contrary, the upward strategy in this study uses exponential smoothing forecasts and low safety stocks, which tends to reduce excess inventories while having the risk of order instability by responding quickly to sudden demand fluctuations. Therefore, it is important for companies that are best suited to use an upward-looking strategy to reinforce the upward-looking strategy and to train human resources accordingly.
From the above, it is necessary for each company in the supply chain to have appropriate strategies and human resource development according to its own position and cooperation with other companies.

CONCLUSION

In this study, a multi-stage supply chain model was constructed using a multi-agent system and the impact of different strategies of each company on the entire supply chain was evaluated. The effect of cooperation, common strategy and information sharing among some companies in the supply chain was analyzed. In the simulation experiment, characteristics of management policies and strategies are categorized into stable and upward strategies.
The results of the simulations show that the bullwhip effect and total costs are reduced by establishing cooperation among some companies as compared to the case where each company develops its own strategy independently. Furthermore, the results of this study can be used to contribute to the improvement of the performance of the entire supply chain by planning appropriate strategies and training human resources according to their own characteristics and cooperation with other companies.
In this study, the experiments are conducted under specific conditions, such as final demand. In the future, the differences in optimal strategies under various conditions will be analyzed by conducting experiments under multiple conditions. Furthermore, it is possible to evaluate human behavior by simulating a beer game. Therefore, it is necessary to conduct a demonstration experiment using a beer game in the future.

ACKNOWLEDGE

This research was partially supported by KAKENHI, Grant-in-Aid for Scientific Research (C), 16K01262 from 2015.

REFERENCES

Abdullah, Z. & Musa, R. (2014). The Effect of Trust and Information Sharing on Relationship Commitment in Supply Chain Management. Procedia - Social and Behavioral Sciences, 130(15), 266-272.

Bhattacharya, R., & Bandyopadhyay, S. (2011). A review of the causes of bullwhip effect in a supply chain. The International Journal of Advanced Manufacturing Technology, 54(9-12), 1245-1261.

Cannella, S., Barbosa-Póvoa, A. P., Framinan, J. M., & Relvas, S. (2013). Metrics for bullwhip effect analysis. Journal of the Operational Research Society, 64(1), 1-16.

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (1999). The bullwhip effect: Managerial insights on the impact of forecasting and information on variability in a supply chain. In Quantitative models for supply chain management, 417-439.

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000a). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. Management science, 46(3), 436-443.

Chen, F., Ryan, J. K., & Simchi-Levi, D. (2000b). The impact of exponential smoothing forecasts on the bullwhip effect. Naval Research Logistics (NRL), 47(4), 269-286.

Croson. R., Donohue. K., Katok. E., & Sterman. J. (2014). Order stability in supply chains: coordination risk and the role of coordination stock, Prod Oper Manag 23(2), 176-196.

Disney, S. M., Farasyn, I., Lambrecht, M., Towill, D. R., & Van de Velde, W. (2006). Taming the bullwhip effect whilst watching customer service in a single supply chain echelon. European Journal of Operational Research, 173(1), 151-172.

Forrester, J. W. (1958). Industrial Dynamics. A major breakthrough for decision makers. Harvard business review, 36(4), 37-66.

Ganesh, M., Raghunathan, S., & Rajendran, C. (2014). The value of information sharing in a multi-product, multi-level supply chain: Impact of product substitution, demand correlation, and partial information sharing. Decision Support Systems, 58, 79-94.

Jakšič, M., & Rusjan, B. (2008). The effect of replenishment policies on the bullwhip effect: A transfer function approach. European Journal of Operational Research, 184(3), 946-961.

Khan, M. H., Ahmed, S., & Hussain, D. (2019). Analysis of bullwhip effect: a behavioral approach. In Supply Chain Forum: An International Journal, Taylor & Francis, 20(4), 310-331.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997a). The bullwhip effect in supply chains. Sloan management review, 38, 93-102.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997b). Information distortion in a supply chain: The bullwhip effect. Management science, 43(4), 546-558.

Ponte, B., Sierra, E., de la Fuente, D., & Lozano, J. (2017). Exploring the interaction of inventory policies across the supply chain: An agent-based approach. Computers & Operations Research, 78, 335-348.

Ribeiro. J. P., & Barbosa-Povoa A. (2018). Supply Chain Resilience: Definitions and quantitative modelling approaches – A literature review, Computers & Industrial Engineering, 115, 109-122.

Shnaiderman, M., & El Ouardighi, F. (2014). The impact of partial information sharing in a two-echelon supply chain. Operations Research Letters, 42(3), 234-237.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E., & Shankar, R. (2008). Designing and managing the supply chain: concepts, strategies and case studies. Tata McGraw-Hill Education.

Trapero, J. R., Kourentzes, N., & Fildes, R. (2012). Impact of information exchange on supplier forecasting performance. Omega, 40(6), 738-747.

Wang, X., & Disney, S. M. (2016). The bullwhip effect: Progress, trends and directions. European Journal of Operational Research, 250(3), 691-701.

Yoon, S. W., Matsui, M., Yamada, T., & Nof, S. Y. (2011). Analysis of effectiveness and benefits of collaboration modes with information-and knowledge-sharing. Journal of Intelligent Manufacturing, 22(1), 101-112.

Zhang, X. (2004). The impact of forecasting methods on the bullwhip effect. International journal of production economics, 88(1), 15-27.

DECISION SCIENCES INSTITUTE

User Satisfaction and Information Technology Continuance Behavior

Md Rasel Al Mamun
University of North Texas
Email: MdRasel.Mamun@unt.edu

William D. Senn
Emporia State University
Email: wsenn@emporia.edu

Russell A. Torres
University of North Texas
Email: Russell.Torres@unt.edu

Victor R. Prybutok
University of North Texas
Email: prybutok@unt.edu

Daniel A. Peak
University of North Texas
Email: Daniel.Peak@unt.edu

ABSTRACT

User satisfaction has proven to be a strong antecedent of IT continuance intention. Synthesizing the extant satisfaction and IT continuance literature thoroughly, the authors create a structural model and test the model in the context of IT to investigate the impact of dual satisfaction in terms of cognitive and emotional satisfaction on IT continuance behavior. The findings suggest that cognitive satisfaction has a significant effect on emotional satisfaction and overall satisfaction, and emotional and overall satisfaction have a strong impact on continuance intention. This research confirms the salience of both cognitive and emotional satisfaction and their correlation in IT continuance.

KEYWORDS: Continuance intention; User satisfaction; Cognitive satisfaction; Emotional satisfaction; Post-adoption behavior.

DECISION SCIENCES INSTITUTE
Learning Analytics as a Community Design Approach to Complex Problems
With Application to the Urban Mobility System

Tristan Stull
Department of Management Science and Information Systems
College of Management
University of Massachusetts Boston

Telephone: (001) 781 710 2368 Email: tristan.stull001@umb.edu

ABSTRACT

This presents a design approach for the Urban Mobility System as a complex problem. The approach leverages data and analytics in search of *sociotechnically optimal* solutions. Sociotechnical optimality is defined in solutions that are efficient, effective and socially congruent. The design model is closely coupled with decision and data models detailing how disaggregate travel behavior and aspirations are expressed against candidate solutions to create mobility plans. The approach is encapsulated in a design-science research framework comprising meta-designs for a process and analytical artifacts. A system of experiments is proposed using agent-based simulation and data mining. A case study is used to explore the approach.

KEYWORDS: Sociotechnical Optimization. Decision Science. Design-Science Research. Vector Optimization. Data Science. Learning Analytics. Urban Mobility.

DECISION SCIENCES INSTITUTE
Communications Among Statistics and Business Journals:
Citation Analysis, Text Analytics, and Network Analysis

Mary Whiteside
University of Texas at Arlington
Email: whiteside@uta.edu

Mark Eakin
University of Texas at Arlington
Email: eakin@uta.edu

Sridhar Nerur
University of Texas at Arlington
Email: snerur@exchange.uta.edu

Qiang Ruan
University of Texas at Arlington
Email: qiang.ruan@mavs.uta.edu

ABSTRACT

Do research trends in business and statistics predict or even reflect the emergence of analytics in business practice and programs in the 21st century? This paper explores the answer to this question by applying MDS maps, network analysis and log-multiplicative models. This research finds that since the emergence of business analytics, knowledge sharing between statistics and academic business disciplines has increased but not substantially.

KEYWORDS: Interdisciplinary Communication, Intradisciplinary Communication, Citation Analysis, Text Analytics, Network Analysis

DECISION SCIENCES INSTITUTE

Using Lean to Improve the Customer Experience in Call Centers: A Meta-Analysis Approach

John Maleyeff
Metropolitan College, Boston University
E-mail: maleyeff@bu.edu

Danqi Lu
Metropolitan College, Boston University
E-mail: danqilu@bu.edu

Canan G. Corlu
Metropolitan College, Boston University
E-mail: canan@bu.edu

ABSTRACT

Call centers have become an important aspect of business for manufacturing and service organizations. Lean methods have been used by call center managers mainly to reduce labor costs. This article uses a meta-analysis to: (1) study the relationship between performance metrics and customer satisfaction, (2) create a process map of a generic call center and illustrate their wasteful activities, and (3) uncover typical solutions that have been reported to help increase call center productivity. A construct based on queuing theory was used to argue that previous efforts could have been used to improve the customer experience without incurring additional costs.

KEYWORDS: Lean, Call center, Meta-analysis, Queuing, Customer experience

INTRODUCTION

The decreased lifespan of products and services, coupled with their increased technical complexity, has precipitated the growth of call enters as an important business process for many firms. Measured in marketing size, the telemarketing and call center industry represents 25.1 billion USD in revenue in 2020 and is expected to grow 2.3 percent annually (IBISWorld, 2020). Because a substantial portent of this industry is outsourced, many business leaders are unfamiliar with the nuances of call center operations. The focus of management efforts tends to be on cost minimization, in particular minimizing cost per call. This goal is typically achieved by maximizing the utilization of call center staff and by making efforts to reduce call handling times.

This article combines a meta-analysis of Lean call center operations management efforts with an analysis of call center operations from a queueing theory perspective. Its aim is not only to show how Lean methods have been successful in improving call center productivity, but to illustrate how these same methods can also be effective in reducing call waiting times without adding additional labor costs. In fact, they can play a dual role of improving productivity along with improving the customer experience. The results are important for call center managers because they show that the myopic goal of maximizing productivity can hamper the call center's ability to service customers effectively.

This article is organized as follows. First, relevant literature that forms the basis of the meta-analysis will be described. Then, a construct is created that uses published work to show the correlations among the methods of Lean, performance metric results, and business impacts. This section is followed by the development of a meta call center process flow map with non-value-added activities that have been highlighted in the cited work. A simulation is used to illustrate the non-linear relationship between call center labor utilization and customer waiting times, highlighting where benefits of reducing handling time can disproportionally improve customer service. Examples of Lean solutions highlighted in the cited literature will be used to illustrate how call centers have found effective ways to reduce call handling times. Finally, results are summarized by discussing important implications for call center managers and researchers.

LITERATURE REVIEW

This research integrates prior work on the management of call center performance and the use of Lean methods to make operational improvements. It cites published literature including the work discussed in this section.

Call Center Operations

Call centers play an important role in establishing a long-lasting relationship with the customer. Seventy-five percent of the consumers surveyed worldwide expressed that they would do business with a company due to a great call center experience (Genesys Global Consumer Survey, 2007). A more recent version of this survey also shows that more than half of the consumers will stop doing business with companies that provide poor customer experience. Maintaining high customer satisfaction is important because of its positive relationship with customer loyalty (Hallowell, 1996; Amoako et al., 2016; Gronholdt et al., 2000; Lia et al., 2014; Heskett & Schlesinger, 1994). In turn, loyalty is a strong indicator of how customers act in the future by impacting the likelihood of customers re-engaging with the same company (Edvardsson et al., 2000).

The determination of call center operations' performance has generated a common list of performance measures, such as waiting time (or queue time), abandonment rate (the percentage of customers who hang up before talking to operators), and average talk time (or handling time). These and other metrics have been found to be critical in call center management because they have a relationship with customer satisfaction (Anton, 1997; Cleveland & Mayben, 1997). However, not all of these metrics are significantly correlated with customer perceptions of quality. Feinberg et al. (2000) used data from 514 call centers (including 15 identifiable separate industries) to analyze the relationship between some of those operational metrics and caller satisfaction. The results showed that there is positive correlation between customer satisfaction and: (a) the percentage of calls closed on the first contact, (b) the call abandonment rate, and (c) customer waiting time. Feinberg et al. (2002) confirmed these correlations by analyzing data from 992 companies. Another study showed the impact of waiting time on customer satisfaction to be quite strong (Pruyn & Smidts, 1998).

All call centers face the challenge of balancing efficiency and increasing customer satisfaction, although many call centers have difficulty achieving both goals (Labach, 2010). In a call center environment, a possible conflict occurs when managers encourage call center employees to patiently resolve customer complaints, but uphold strict operational performance evaluations

using metrics such as handling time per call (Tuten & Neidermeyer, 2004). In fact, an emphasis on invasive performing monitoring can increase stress among call center employees (Spigg & Jackson, 2006). Sarel and Marmorstein (1998) performed an empirical study to point out that the majority of customers neither expect nor condone delays, and the assessments of service quality will be affected when delays occur frequently. Other researches have shown that as waiting time increases, customer satisfaction decreases (Garcia et al., 2012; Peevers et al., 2009; Davis & Volmann, 1990; Bielen & Demoulin, 2007). Thus, reducing waiting time plays a key role in improving customer satisfaction.

The method commonly applied to improve customer satisfaction is performance monitoring, such as call listening, call recording and feedback on performance and call quality (Holman et al., 2007; Piercy & Rich, 2009). However, theories about monitoring are controversial. Monitoring does promote job performance to a certain extent (Aiello & Kolb, 1995), but it invades employee privacy and decreases job satisfaction (Haley et al., 2012; Piturro, 1989). Also, the effect of monitoring depends on how the employee perceives management policy. Haley et al. (2012) reported that positive perceptions of monitoring systems are likely to reduce call centers employees' turnover intentions. Conversely, negative perceptions of monitoring systems are likely to weaken employee exchanges with their organizations to the point where they consider leaving their organizations, causing the company to recruit an inexperienced new hire. Call center turnover costs are high because replacing employees involves more than just advertising for positions, interviewing candidates, and providing initial training (Hillmer et al., 2004). On average, replacing one agent equals 16% of the gross annual earnings of a call center worker – that is, the simple replacement costs of one worker equals about two months of a typical worker's pay. If lost productivity is taken into account, replacing one worker equals between three and four months of a typical worker's pay (Holman et al., 2007). Therefore, improving customer satisfaction without adding substantial costs becomes a crucial issue for all call centers.

Lean in Call Centers

Lean production was first developed by Taiichi Ohno and Shigeo Shingo at Toyota Motor Manufacturing in Japan (Ohno, 1998). Their approach contrasted with the economies of scale approach that was typified by the production of large batches of products. Lean seeks to minimize time spent on activities that do not directly add value for customers, which reduces the need for batching, decreases production lead times, and enables better responsiveness to customer demand. Despite taking its roots from a production environment, Lean has gradually evolved into service industries mainly due to the growing competitiveness in the service sector (George, 2003; Antony et al., 2007). Service industries where Lean has been applied include banking (Rizkya et al., 2018), education (Emiliani, 2004), healthcare (Bhat et al., 2014; Kovach & Borikar, 2018), and environmental consulting (Ball & Maleyeff, 2003). Today, Lean management is a total business approach designed to identify and eliminate forms of waste in the process of producing goods, services, or combinations of both (Ball & Maleyeff, 2003).

Womack and Jones (1996) discuss the role that Lean plays in the service industry and, by analyzing 50 companies implementing Lean principles, they outlined five steps as a guideline for managers. Lean makes use of process maps that display, from start to finish, how a process creates products and services. The process map is analyzed to identify activities that do not add value for customers, called non-value-added, or wasteful, activities. Although Shingo categorized wasteful activities for manufacturing operations (Ohno, 1998), Maleyeff (2006)

created a list of categories relevant to service processes: (1) mistakes, (2) delays, (3) reviews, (4) movement, (5) duplication, (6) resource inefficiencies, (7) processing inefficiencies. To eliminate time spent on wasteful activities, Lean practitioners collaborate with teams of workers to create standard work processes that employ poka-yoke (mistake proofing), visual controls, and a method of workplace organization called 5S.

For large call centers, small reductions in handling times can reduce costs through reductions in labor requirements (Saif, 2016). The application of Lean can optimize operations and improve both employee and customer satisfaction. Piercy and Rich (2009) applied Lean methods in three call centers within the financial sector where wasteful activities were identified and removed from the process, such as customers exiting the system due to long wait times to speak to an operator. In addition, the elimination of dysfunctional performance measures and rigid worker controls generated positive gains in workplace morale, reducing staff absenteeism and turnover. In another study, Al-Akwaa and Dunahay (2015) found that Lean is efficient in removing waste in technical support call centers and helps operators solve customer problems on the first call which enhanced customer satisfaction. Laureani & Douglas (2010) studied a large vehicle leasing and renting industry call center where the successful application of Lean principles streamlined operations, reduced operators' turnover, and increased customer satisfaction. Similarly, through the use of Lean Six Sigma principles, McAdam et al. (2009) showed that the reduction of handling time, abandonment rate, and failure related calls led to higher service quality and better satisfaction of demand. Here, the costs and utilization of staff decreased because work being handled by 26 advisers during the implementation was equivalent to the work of 30 advisors in the past. Others have shown similar improvements in performance using Lean methods (Bhat & Jnanesh, 2014; Weaver et al., 2013).

THEORETICAL DEVELOPMENT

Call center operations can be modeled as a classic queuing system, and analyzed using either an analytical queueing model or a Monte Carlo simulation (MCS). Queueing models tend to be mathematical abstractions that are useful when the real system adheres to the restrictive assumptions of the model. They are easy to use as long as a computational tool is provided to perform the somewhat complex calculations. A MCS is robust and therefore can be used when the system is too complex for analytical modeling. However, MCS often requires long code development, long processing times, and their results are affected by random variation.

In this work, call center operations will be analyzed using a MCS with the following assumptions:

(a) Customers will arrive one-at-a-time at random times that are independent across customers at a constant rate. As such, the number of arrivals will follow a Poisson process and the time between arrivals will follow an exponential distribution.

(b) Customers will be served by a set of parallel servers, such that all servers can handle any call. They will wait in a common queue when all servers are busy and taken off the queue on a first-come first-served basis.

(c) Each customer is served by one server and the service time follows a gamma distribution. The gamma distribution is flexible because it includes a shape parameter that determines its level of skew, and a scale parameter that determines its magnitude.

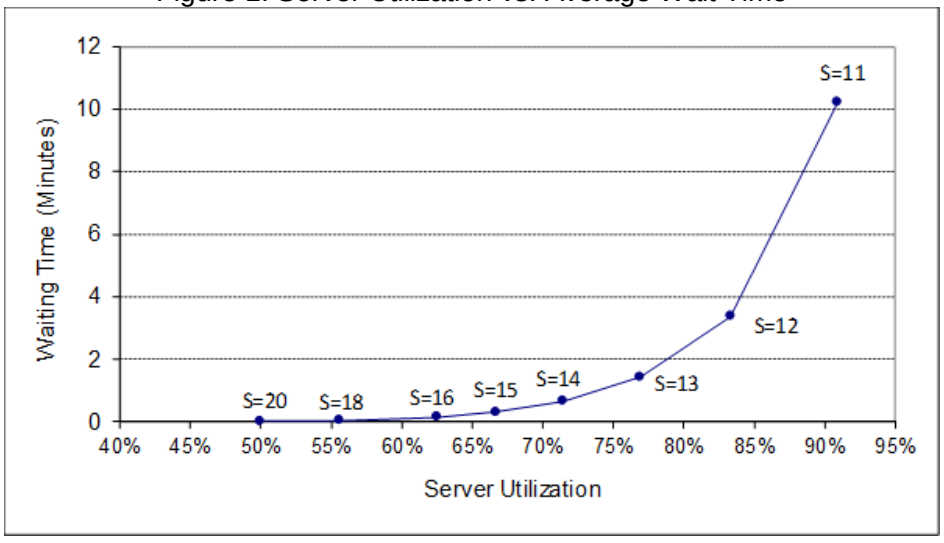
(d) Customers exit the system after service is complete.

The gamma distribution is useful for modeling service time variation. When the shape parameter is 1.0, the gamma distribution is identical to the exponential distribution. In this case, M/M/S analytical queuing model assumptions will apply. However, it is not common for service time distributions to follow an exponential pattern of variation. Usually, the shape parameter exceeds 1.0 and the service time distribution to be right-skewed, where the amount of skew decreases as the shape parameter increases. Coefficients of variation between 25% and 75% are common for service times that follow a gamma distribution (Millhiser & Veral, 2019).

The effectiveness of a queuing system is typically measured based on two perspectives. From the customer's perspective, effectiveness is generally measured by either waiting time in queue or total time in system. From the service provider's perspective, effectiveness is generally measured by the planned server utilization (or similarly, service provider's idle time which is also called the capacity buffer size), which is directly proportional to labor costs. The reconciling of these two perspectives – the customers' desire for quick service and the firm's need to lower costs – can be a challenge because one perspective (i.e., server cost) is financially quantifiable while the other perspective (i.e., customer service) is rarely financially quantifiable. When the cost of customers' waiting can be quantified financially, the optimal server utilization can be easily calculated. However, this case is rare.

An approach for determining the optimal server utilization is based on the non-linear mathematical relationship between the server utilization and the average customer waiting time. This relationship will follow a so-called *hockey-stick* function. As an illustration, consider a call center where customer calls arrive at a rate of 40 per hour and the average service time is 15 minutes. Using the M/M/S model assumptions (i.e., parallel servers, Poisson arrivals, and exponential service times), the function shown in Figure 1 *w*ould exist as the number of servers (S) is changed from 20 to 11 (e.g., when S=13, the server utilization will be 77% and the average wait time will be about 1.5 minutes). This hockey-stick pattern is found in any queuing system with either random arrivals or random service, regardless of the service time distribution.

Figure 1: Server Utilization vs. Average Wait Time



Because of the difficulty in quantifying costs associated with customers' waiting, the approach used here is generic and mathematical, based on finding the "knee" of the waiting time versus server utilization curve. The problem of finding the knee of a non-linear curve has been discussed (Salvador & Chan, 2004), and several methods are available. In the context of internet congestion, Kleinrock (2018) suggested that a power function be calculated for each potential decision option (Equation 1) where Ws is the average time in the system, ρ is the server utilization, and μ is the average service rate. The power function is the ratio of "good" (because one goal is high server utilization) and "bad" (because another goal is low service turnaround times). Its numerator will vary from 0 to 1 and its denominator, called the normalized average response time, will typically vary from 1 (when there is no waiting) to about 2. The optimal decision is based on the point of the hockey stick where the power is maximized. This value is referred to as the "knee" of the hockey-stick function.

$$Power = \frac{\rho}{\mu Ws} \quad (1)$$

If we assume that the arrival rate and service times will not change, the only option available to the decision maker would be to determine the optimal number of servers. The power at each decision option is shown in Table 1 for the example under consideration. The knee is found at 13 servers, which would be the optimal number of servers. When applied to call centers, empirical experimentation indicates that the knee approach favors limiting customer wait times more than minimizing server costs. However, determining the knee provides an effective means to approximate a starting point for call center system optimization.

Table 1: Power Function Example

Servers	Utilization	Ws (min)	Power
20	50%	15.01	0.4998
19	53%	15.01	0.5259
18	56%	15.03	0.5545
17	59%	15.07	0.5857
16	63%	15.14	0.6191
15	67%	15.31	0.6533
14	71%	15.65	0.6847
13	77%	16.43	0.7024
12	83%	18.37	0.6804
11	91%	25.23	0.5404

Because service times often consist of non-value-added activities, server utilizations can be lowered without increasing the number of servers. Given the hockey-stick functional relationship, reducing service times by a small amount can have disproportionate benefits for customers. For the call center example, the average service time was 15 minutes and 40 customers arrived 40 per hour. With 11 servers, server utilization is 90.9% and the average wait time for callers is 10.2 minutes. If the 15 minute average service time can be reduced by 45 seconds to 14.25 minutes, 11 servers can now serve 46.3 customers in one hour and server utilization is 86.4%. The average customer waiting time would decrease by about 5 minutes

(Figure 1). This result (a 5 minute waiting time reduction after a service time reduction of 45 seconds) may seem counter-intuitive, but it is a direct result of the hockey-stick function.

METHOD

Cooper (2015) defines meta-analysis as “the quantitative procedures that are used to statistically combine the results of studies.” It is advantageous over a single study because it systematically studies several articles that look at a similar problem. It also enables the analysis of some questions that are not answered by individual studies (Deeks et al., 2019). Another advantage is that it reveals conflicting results from similar studies and thus enables the investigation of reasons for different perspectives. Cooper (2015) lists 7 steps of meta-analysis as follows: (1) formulating the problem; (2) searching the literature; (3) gathering information from studies; (4) evaluating the quality of studies; (5) analyzing and integrating the outcomes of studies; (6) interpreting the evidence; and (7) presenting the results. These steps were followed in this research.

The use of meta-analysis in several disciplines has been growing. Although meta-analysis has been used heavily in healthcare, its use in service systems in general is relatively limited. Maleyeff (2006) is one of the earliest studies that performs meta-analysis of about 60 internal service systems to draw conclusions regarding the problems and opportunities that exist in the implementation of Lean principles to services. The authors believe that the research presented here is the first to perform meta-analysis of call center operations for the following purposes: (1) study the relationship between customer satisfaction and several performance metrics; (2) create a process map of a generic call center and identify wasteful activities; and (3) identify typical solutions that were reported in the literature that help increase call center productivity.

The framework shown in Figure 2 describes the procedure for a meta-analysis that integrates process mapping and analysis, Lean methods for waste removal, operational improvements, and their business implications (reduced labor cost and/or customer experience improvements). By instituting a Lean-based process analysis, activities that do not add value (i.e., wasteful activities) are identified. Lean (and other methods) can be used to reduce time spent on these activities, which will impact operational performance. Normally, call centers take this as an opportunity to lower costs by reducing call center personnel. As shown by the queueing analysis, this reduction in handling time can also positively impact the customer experience in a disproportionately large way. In many cases, a combination of these approaches (lowering costs and reducing waiting time) can be implemented.

RESULTS

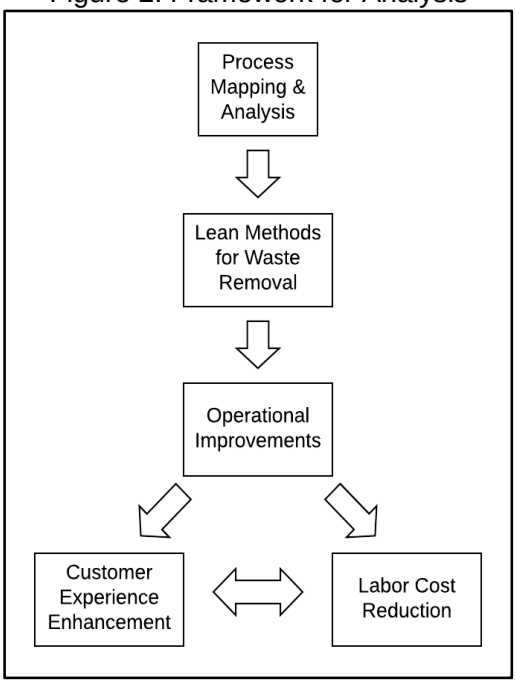
Table 2 provides a set of codes that is used in the following sections. These codes will note how each reference listed was incorporated into the meta-analysis.

Relationships Identified

This section shows the results of the meta-analysis indicating how the use of Lean methods will impact performance (Figure 3) and how those impacts have been shown to enhance business competitiveness. The focus is on the correlations that have been uncovered. It is noteworthy that these correlations show a direct link between visible quantifiable impacts (e.g., handling time reduction) and those that may be hidden from view because they cannot be easily

measured (e.g., customer retention). In summary, Lean methods have been shown to have a positive impact on the three main operational goals: better (e.g., higher quality), cheaper (e.g., lower labor costs), and faster (less customer waiting). In addition, the analysis shows that these impacts will improve revenue by improving customer retention rates. Figure 3 uses the assigned codes to identify the source of each uncovered correlation.

Figure 2: Framework for Analysis



Call Center Process Analysis

A process flow map that represents a generic, but representative, call center was developed based on the types of call center operations that were analyzed in the articles cited. It also reflects the work reported by researchers who applied Lean methods of analysis and improvement to various call centers. The main assumption is that the intent of the call center is to solve the customer's problem during a single call, although the potential exists for the customer to call back if the solution posed by the call center representative was ineffective.

The generic process map (Figure 4) shows the entire process from a customer contacting the call center by telephone to the solution being presented and evaluated by the customer. Within the process map, a set of 19 wasteful activities that have been cited by various authors are highlighted by number. For example, customers often wait in the queue (and many of them abandon the call at this point). Or, customers may choose an inappropriate option in the electronic sorting system before talking to a representative, which results in wasted time while they are reconnected to the proper staff member. Table 3 lists the source of each wasteful activity identified by number in the process map, including its category and description.

Table 2: Reference List Codes

Code	Resources
A&D	Al-Akwaa & Dunahay, 2015
BG&J	Bhat, Gijo, & Jnanesh, 2014
DW&F	De Ruyter, Wetzels, & Feinberg, 2001
FKH&K	Feinberg, Kim, Hokama, & Keen, 2000
GAM&G	Garcia, Archer, Moradi, & Ghiabi, 2012
GS&V	Gruber, Szmigin, & Voss, 2006
H&S-05	Homburg & Stock, 2005
H&S-94	Heskett & Schlesinger, 1994
KAC&W	Keiningham, Aksoy, Cooil, & Wahren, 2006
L&D	Laureani & Douglas, 2010
MDK&F	McAdam, Davies, Keogh, & Finnegan, 2009
P&R	Piercy & Rich, 2009
P&S	Pruyn & Smidts, 1998
PMMM&J	Peevers, McInnes, Morton, Matthews, & Jack, 2009
S	Saif, 2016
S&F	Sergeant & Frenkel, 2000
T&N	Tuten & Neidermeyer, 2004
Wetal	Weaver, Greeno, Goughler, Yarzebinski, Zimmerman, & Anderson, 2013
W	Whitt, 2006
WSO&A	Wojtys, Schley, Overgaard, & Agbabian, 2009)
YS&Y	Yoon, Seo, & Yoon, 2004

Other wasteful activities are more general because they cause waste in multiple process locations. They are listed below:

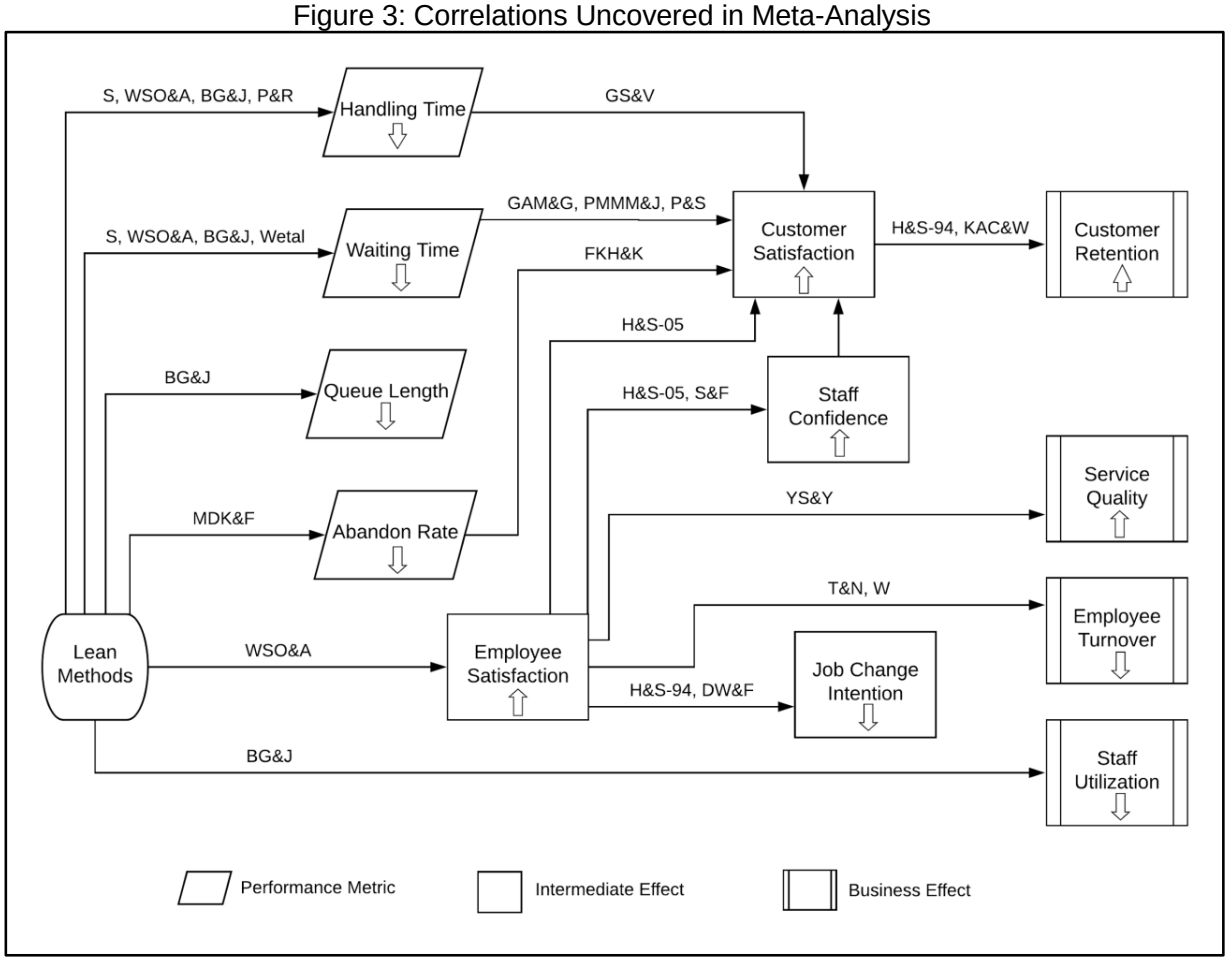
- Some customer problems require a particular long time to solve and the learning curve of new employees can be slow (Saif, 2016).
- The solution that call centers provide may not be accurate, resulting in a call back (Al-Akwaa & Dunahay, 2015).
- Operators may need to concentrate on non-core tasks, such as printing documents, photocopying, and faxing, which impacts their efficiency (Laureani & Dougals, 2010).

Effect of Reducing Handling Times

Figure 5 illustrates the disproportionate improvements in customer service that can occur when reductions in service times results in call center operators working with decreased utilizations. It is assumed that the call center operates with 20 servers. Customers contact the call center at a rate of 285 per hour and follow a Poisson process. Service times are assumed to follow a gamma distribution with a coefficient of variation of 75%. The server utilization is controlled by changes in the average service time. For example, when the average service time is 3.75 minutes per customer, the 20 servers can serve an average of 320 calls per hour and, therefore, their utilization would be 89% (the ratio of 285 incoming calls per hour to the service rate of 320

calls per hour). The graph also includes a fifth order equation that accurately predicts average waiting time (Wq, or y) based on the server utilization (ρ, or x). The usefulness of the fit is evident based on the R^2 value of 97.3%.

Figure 3: Correlations Uncovered in Meta-Analysis



For this call center, the knee is 86% and the optimal capacity buffer is 14%. Table 4 shows the benefits that would be realized by customers of the call center given a set of assumed reductions in average service times. For example, if call center servers were currently utilized at 98% (i.e., the average service time was 4.13 minutes), the average waiting time would be about 7.0 minutes. By reducing time spent on non-value-added activities, perhaps the service time can be reduced to 3.79 minutes (i.e., 20 seconds). This change can reduce server utilization to 90% (above the knee but a reasonable wait given the hockey-stick curve). In this case, the average waiting time would be reduced to about 1.4 minutes (as shown in Table 4), a decrease of 5.6 minutes. In summary, removing 20 seconds from the average service time will save the average customer almost 6 minutes of waiting. As shown in the articles detailed below, this magnitude of service time reduction is quite possible. In fact, service time reductions have been achieved at many call centers studied in this meta-analysis.

Figure 4: Generic Call Center Process Map

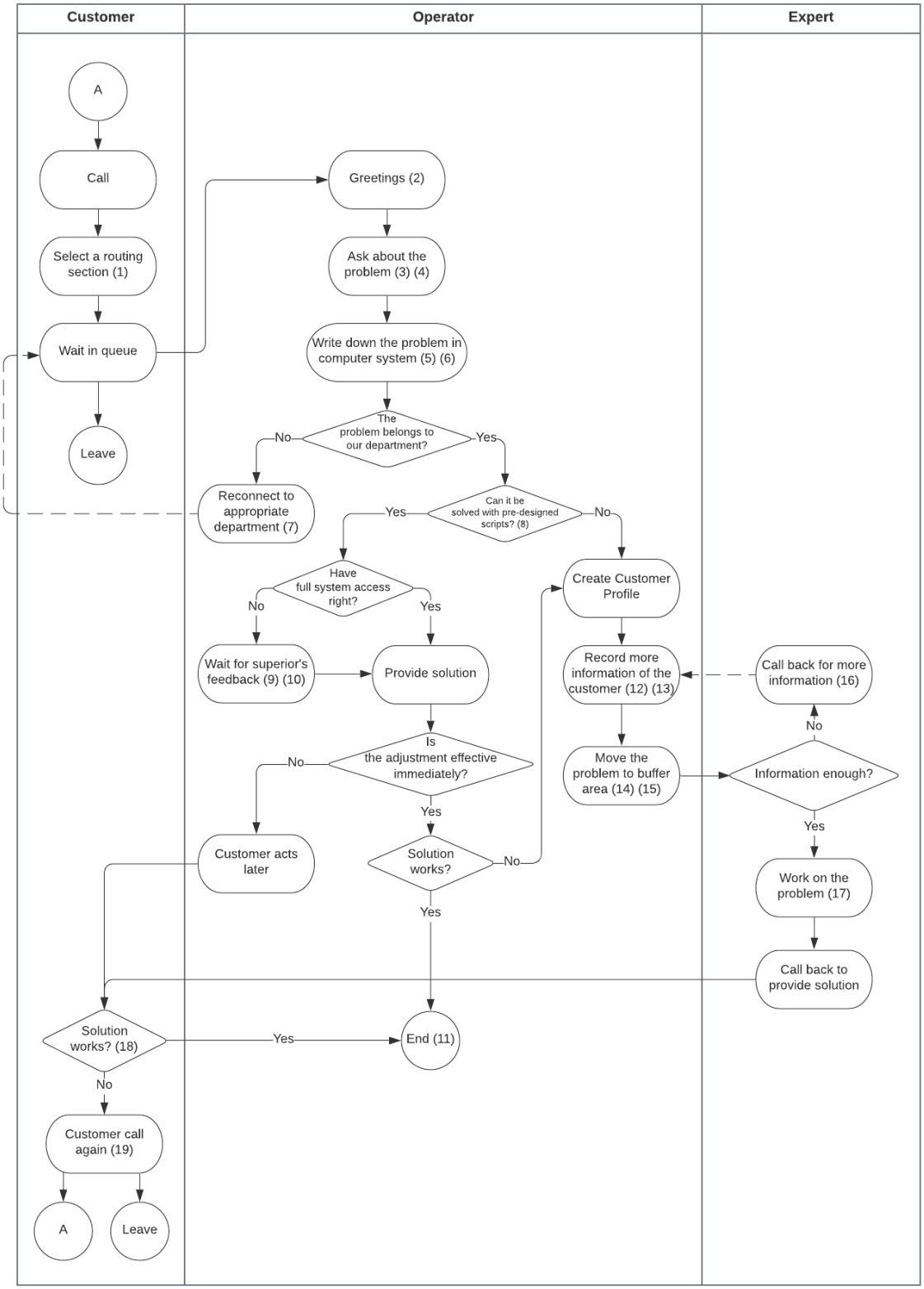


Table 3: Wasteful Activities Identified in Meta-Analysis

No.	Wasteful Activities		Code
	Type	Problem Description	
1	Mistake	Upon entering the process, customers selected an inappropriate option from the electronic routing system (either due to an oversight or due to their actual problem not being an option and selecting the closest approximation).	P&R
2	Delay	Long greetings and a long call hold time.	S
3	Mistake	Operative accepted the call and started serving a customer while still completing the notes from their previous customer; this led to frequent problems of incorrect or incomplete information being entered.	P&R
4	Duplication	Typing mistakes were created by the operator while entering the customer's information to the computer system.	BG&J
5	Processing inefficiency	Customers with different mother-tongues were getting service in the call center, thus the operator as well as customers had difficulty in communication.	BG&J
6	Processing inefficiency	Operators with a slow typing speed tended to have a high average handling time.	S
7	Movement	Routing a call to another agent, which increased the risk of delaying, incorrectly routing, or losing the call.	A&D
8	Processing inefficiency	Call operators were unable (due to the system design) or unwilling (due to the performance measures) to resolve customer issues; they placed increased amounts of work into the buffer area.	P&R
9	Delay	Lack of access to the required information for the call center agent, which produced longer wait times for customers.	A&D
10	Delay	Call center operators could not give first contact responses because they were waiting for feedback from either another department or staff in supervisory roles.	L&D
11	Processing inefficiency	After every call, operators must input the reason for the call into the customer relationship management system; but the process of capturing the insight gained from the call was unnecessarily lengthy, requiring the operators to work through a series of four-to-five steps.	S
12	Mistake	Errors in identifying the information provided by the customer due to incorrect spellings of customers name and address.	BG&J
13	Reviews	Unnecessarily long summing up when operators reconfirmed the customer's query and the solution provided before ending the call.	S
14	Processing inefficiency	Operators were deployed from the general staff group to reduce the amount of work in the buffer area, which reduced the number of available staff to answer incoming queries.	P&R
15	Duplication	Technical queries that could not be solved immediately on first contact and had to be processed offline by another analyst.	L&D
16	Reviews	Staff verified and clarified the information in a data entry form, but it needed to be checked to determine if all the information was correct (if not, the request would be returned to the customer).	BG&J
17	Duplication	When the call has already been over-possessed by different operators on the team, then assigned to a team lead; the issue has become more complex.	A&D
18	Mistake	Incorrect solutions were considered as incomplete solutions leading to repeated calls about the same issues.	L&D
19	Mistake	When the solution was not provided on the first call, the customer contacted the call center again for the same inquiry.	A&D

Figure 5: Example Hockey-Stick Function

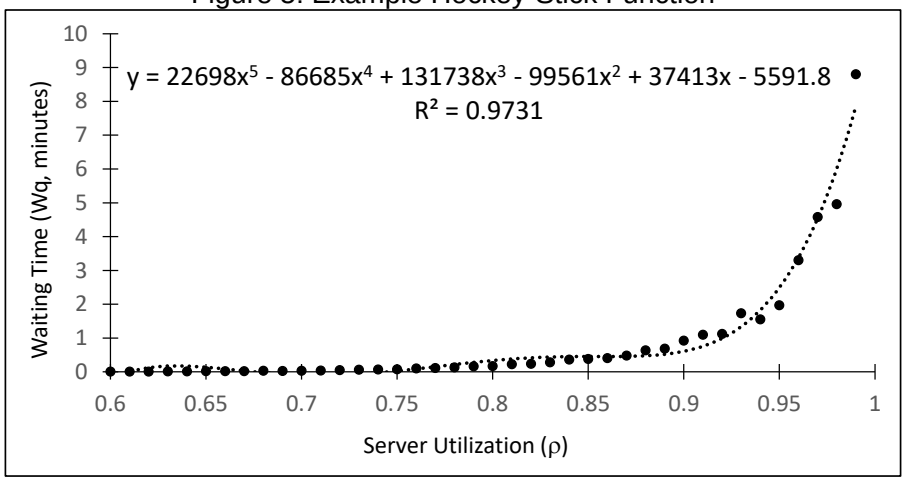


Table 4: Numerical Results

Server Util (ρ)	Avg Svc Time	Avg Total Time (Ws)	Avg Queue Time (Wq)	Power
86%	3.62	4.03	1.16	0.774
87%	3.66	4.16	1.19	0.767
88%	3.71	4.36	1.23	0.748
89%	3.75	4.42	1.29	0.755
90%	3.79	4.72	1.41	0.723
91%	3.83	4.94	1.58	0.706
92%	3.87	5.01	1.84	0.712
93%	3.92	5.66	2.21	0.644
94%	3.96	5.49	2.73	0.677
95%	4.00	5.98	3.42	0.636
96%	4.04	7.36	4.34	0.527
97%	4.08	8.66	5.52	0.458
98%	4.13	9.09	7.02	0.445
99%	4.17	12.97	8.89	0.318

As an alternative, a call center manager can also “split the difference” by simultaneously reducing server utilization and the number of servers. The resulting decrease in server utilization would not be as significant, but the improvement in the customer experience will still be noticeable to customers. For example, removing 20 seconds from service times can lead to a decrease in labor assignments along with a decrease in labor utilization from 98% to 94%. In this case, as shown in Table 4, the average customer waiting time will decrease from 7.0 minutes to 2.7 minutes (a decrease of 4.3 minutes per customer).

Lean-Based Improvement Potential

This section provides examples of how Lean methods were used within call centers to remove wasteful activities. Lean methods can make up for the shortcomings of the original process,

improve the efficiency of the call center, and play a role in preventing problems. The Lean approach would transition call center management from top-down control of workers with narrow job-roles and rigidly enforced efficiency measures to a cross-skilled workforce, rewarding operators for customer service through a supportive management system (Piercy & Rich, 2009). It would eliminate over-reliance on monitoring and using strictly maintained performance metrics.

The main Lean methods listed in Table 5 include: Jidoka (separating work best done by machines from work best done by humans), 5S (organizing the workplace to keep data and other information in close proximity and easy to locate), poka-yoke (mistake proofing the activity to eliminate or reduce the likelihood of a mistake), visual workplace (showing information useful for operators in a location that is easily seen by all personnel), and SMED (reducing the "setup" time for transitioning from customer to customer).

Table 5: Lean Solutions Uncovered in Meta-Analysis (Examples)

Lean Method	Description	Code
Jidoka	Incorporated greetings into an interactive voice response, which shortened the greetings by agents to customers.	S
5S	Relocated the call center to an area close to staff from both the claims processing and technical support areas - when complex problems were received they were resolved more quickly.	P&R
5S	Stopped providing sum-up dialogues that were not adding value for the customer.	S
Poka-Yoke	Only correct customer information forms can be accepted by the system.	BG&J
Poka-Yoke	Created a single pool of workers to help diagnose a type of customer requests rather than using an electronic call routing sort system.	P&R
5S	Gave specialized training to operators for understanding particular problems and specific situations.	BG&J
Standard Work	Arranged in-house training related to English grammar, quick typing skills, and other related activities such as the use of short-cut keys.	BG&J
Standard Work	Established a minimum required typing speed of 25 words per minute for contact center agents.	S
5S	Trained operators so they could resolve queries without constant referral to other departments or a buffer area; redesigned the performance measurement systems of each company to reduce dysfunctional behavior.	P&R
Visual Workspace	Designed a new performance system based on sampling calls and evaluating them on a new scale that measured how well the customer requirement was handled - these scores were plotted and used to identify new streams of customer demand and to enhance future training programs.	P&R
SMED	Simplified the insight-capturing process in the customer relationship management system from four steps to one step; the interface was changed to ensure that insights can be captured in one window with a single click.	S
5S	Empowered operators to act independently of set scripts based on their own knowledge of the operational requirements of each type of customer issue.	P&R

DISCUSSION

Call centers are necessarily focused on cost savings to achieve success due to the highly competitive nature of most businesses. This work highlighted how Lean methods have been successfully applied to identify wasteful activities and improve competitiveness by reducing time spent on these non-value-added activities. Due to the non-linear hockey-stick function that relates server utilization to customer waiting time, opportunities can be lost to improve the customer experience without incurring higher costs, because previous efforts have been devoted entirely to cost reductions.

This work was limited by its focus on publications that highlighted the use of Lean (in some cases in conjunction with Six Sigma) methods. The meta-analysis can be expanded to include similar efforts that have been undertaken using various other industry jargon, such as total quality management, continuous quality improvement, or “good old” common sense. Although this work focused on call centers with relatively stable demand, future research should also focus on how call center labor capacity can or should be rapidly redeployed during times of turmoil, such as during the 2020 COVID-19 pandemic.

ACKNOWLEDGEMENTS

The authors would like to thank Ningjian Huang and Tiange Shen for their contributions during the initial phases of this effort.

REFERENCES

Aiello, J., & Kolb, K. (1995). Electronic performance monitoring and social context: Impact on productivity and stress. *Journal of Applied Psychology*, *80*(3), 339-353.

Al-Akwaa, N., & Dunahay, P. (2015). *The application of Lean Six Sigma principles in the technical support call center: First contact resolution*, ProQuest Dissertations and Theses.

Amoako, G., Dzogbenuku, R., & Doe, J. (2016). How service experience leads to brand loyalty: perspective from the telecom sector in Ghana. *IUP Journal of Brand Management*, 13(2), 33-56.

Anton, J., Monger, J., & Perkins, D. S. (1997). *Call Center Management by the Numbers*. Purdue University Press.

Antony, J., Antony, F. J., Kumar, M., & Cho, B. R. (2007). Six Sigma in service organizations. *International Journal of Quality & Reliability Management*. *24*(3), 294-331.

Ball, D. R., & Maleyeff, J. (2003). Lean management of environmental consulting. *Journal of Management in Engineering*, *19*(1), 17-24.

Bhat, S., Gijo, E., & Jnanesh, N. (2014). Application of Lean Six Sigma methodology in the registration process of a hospital. *International Journal of Productivity and Performance Management*, *63*(5), 613-643.

Bielen, F., & Demoulin, N. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality: An International Journal*, *17*(2), 174-193.

Bowen, D., & Youngdahl, W. (1998). "Lean" service: In defense of a production-line approach. *International Journal of Service Industry Management*, *9*(3), 207-225.

Cleveland, B., & Mayben, J. (1997). *Call center management on fast forward: Succeeding in today's dynamic inbound environment*. ICMI Inc.

Cooper, H. (2015). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach* (Vol. 2). Sage publications.

Davis, M. M., & Vollmann, T. E. (1990). A framework for relating waiting time and customer satisfaction in a service operation. *The Journal of Services Marketing*, *4*(1), 61.

Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta- analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, 241-284.

De Ruyter, K., Wetzels, M., & Feinberg, R. (2001). Role stress in call centers: Its effects on employee performance and satisfaction. *Journal of Interactive Marketing*, *15*(2), 23-35.

Edvardsson, B., Johnson, M., Gustafsson, A., & Strandvik, T. (2000). The effects of satisfaction and loyalty on profits and growth: Products versus services. *Total Quality Management*, *11*(7), 917-927.

Emiliani, M. (2004). Improving business school courses by applying Lean principles and practices. *Quality Assurance in Education: An International Perspective*, *12*(4), 175-187.

Feinberg, R., Hokama, L., Kadam, R., & Kim, I. (2002). Operational determinants of caller satisfaction in the banking/financial services call center. *The International Journal of Bank Marketing*, *20*(4), 174-180.

Feinberg, R., Kim, I., Hokama, L., & Keen, C. (2000). Operational determinants of caller satisfaction in the call center. *International Journal of Service Industry Management*, *11*(2), 131-141.

Garcia, D., Archer, T., Moradi, S., & Ghiabi, B. (2012). Waiting in vain: Managing time and customer satisfaction at call centers. *Psychology*, *3*(2), 213-216.

Genesys Telecommunications Labs. (2007). *Genesys Global Consumer Survey*. Retrieved from: http://www.crmxchange.com/uploadedFiles/White_Papers/PDF/Genesys_Global_survey.pdf

Genesys Telecommunications Labs. (2019). *Genesys Global Consumer Survey*. Retrieved from: https://www.genesys.com/collateral/genesys-usa-consumer-survey-2019

George, M. L. (2003). *Lean Six Sigma for Service*. New York.

Gronholdt, L., Martensen, A., & Kristensen, K. (2000). The relationship between customer satisfaction and loyalty: Cross-industry differences. *Total Quality Management*, *11*(4-6), 509-514.

Gruber, T., Szmigin, I., & Voss, R. (2006). The desired qualities of customer contact employees in complaint handling encounters. *Journal of Marketing Management*, *22*(5-6), 619-642.

Haley, Flint, & Mcnally. (2012). The effects of employee perceptions of monitoring procedures on turnover. *Problems and Perspectives in Management*, *10*(3), 75-82.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *International Journal of Service Industry Management*, *7*(4), 27-42.

Heskett, J., & Schlesinger, L. (1994). Putting the service-profit chain to work. *Harvard Business Review*, *72*(2), 164-174.

Hillmer, S., Hillmer, B., & McRoberts, G. (2004). The real costs of turnover: Lessons from a call center. *Human Resource Planning*, *27*(3), 34-42.

Holman, D., Batt, R., & Holtgrewe, U. (2007). *The Global Call Center Report: International Perspectives on Management and Employment [Electronic version]*. Ithaca, NY: Authors.

Homburg, C., & Stock, R. (2005). Exploring the conditions under which salesperson work satisfaction can lead to customer satisfaction. *Psychology and Marketing*, *22*(5), 393-420.

IBISWorld (2020), *Telemarketing & call centers in the US market size 2002–2025, IBISWorld, New York*, https://www.ibisworld.com/industry-statistics/market-size/telemarketing-call-centers-united-states/, retrieved June 7, 2020.

Keiningham, T., Aksoy, L., Cooil, B., & Wahren, B. (2006). Call center satisfaction and customer retention in a co-branded service context. *Managing Service Quality*, *16*(3), 269-289.

Kleinrock, L. (2018). Internet congestion control using the power metric: Keep the pipe just full, but no fuller. *Ad hoc networks*, *80*, 142-157.

Kovach, J. V., & Borikar, S. (2018). Enhancing financial performance: An application of Lean Six Sigma to reduce insurance claim denials. *Quality Management in Healthcare*, *27*(3), 165-171.

Labach, E. (2010). Improving customer retention through service quality at call centers. *International Journal of Management and Information Systems*, *14*(3), 71-76.

Laureani, A., & Douglas, A. (2010). Lean Six Sigma in a call centre: A case study. *International Journal of Productivity and Performance Management*, *59*(8), 757-768.

Liat, C. B., Mansori, S., & Huei, C. T. (2014). The associations between service quality, corporate image, customer satisfaction, and loyalty: Evidence from the Malaysian hotel industry. *Journal of Hospitality Marketing & Management*, *23*(3), 314-326.

Maleyeff, J. (2006). Exploration of internal service systems using Lean principles. *Management Decision*, *44*(5), 674-689.

McAdam, R., Davies, J., Keogh, B., & Finnegan, A. (2009). Customer-orientated Six Sigma in call centre performance measurement. *The International Journal of Quality & Reliability Management*, *26*(6), 516-545.

Millhiser, W. P. & Veral, E. A. (2019). A decision support system for real-time scheduling of multiple patient classes in outpatient services. *Healthcare Management Science*, 22 (1), 180 -195.

Ohno, T. (1988). *Toyota Production System: Beyond Large-Scale Production*. CRC Press.

Peevers, G., McInnes, F., Morton, H., Matthews, A., & Jack, M. (2009). The mediating effects of brand music and waiting time updates on customers' satisfaction with a telephone service when put on-hold. *The International Journal of Bank Marketing*, *27*(3), 202-217.

Piercy, N., & Rich, N. (2009). Lean transformation in the pure service environment: The case of the call service centre. *International Journal of Operations & Production Management*, *29*(1), 54-76.

Piturro, M. (1989). Employee Performance Monitoring . . . Or Meddling? *Management Review*, *78*(5), 31.

Pruyn, A., & Smidts, A. (1998). Effects of waiting on the satisfaction with the service: Beyond objective time measures. *International Journal of Research in Marketing*, *15*(4), 321-334.

Rizkya, I., Syahputri, K., Sari, R. M., & Siregar, I. (2018). Lean Thinking in banking industry: Waste identification on credit disbursement process. *E3S Web of Conferences*, *73*, 07004.

Saif, S. (2016). Every Second Counts: How one call center applied Lean and Six Sigma to reduce its average call-handling time. *Quality Progress*, *49*(2), 24-28.

Sarel, D., & Marmorstein, H. (1999). Managing the delayed service encounter: The role of employee action and customer prior experience. *The International Journal of Bank Marketing*, *17*(6), 286-294.

Salvador, S., & Chan, P. (2004). Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence, 576-584.

Sergeant, A., & Frenkel, S. (2000). When do customer contact employees satisfy customers? *Journal of Service Research*, *3*(1), 18-34.

Sprigg, C.A., and Jackson, P.R. (2006), Call centers as Lean service environments: Job-related strain and the mediating role of work design, *Journal of Occupational Health Psychology*, 11(2), 197-202.

Tuten, T., & Neidermeyer, P. (2004). Performance, satisfaction and turnover in call centers: The effects of stress and optimism. *Journal of Business Research*, *57*(1), 26-34.

Weaver, A., Greeno, C., Goughler, G., Yarzebinski, D., Zimmerman, H., & Anderson, K. (2013). The impact of system level factors on treatment timeliness: Utilizing the Toyota production system to implement direct intake scheduling in a semi-rural community mental health clinic. *The Journal of Behavioral Health Services & Research*, *40*(3), 294-305.

Whitt, W. (2006). The impact of increased employee retention on performance in a customer contact center. *Manufacturing & Service Operations Management*, *8*(3), 235-252.

Wojtys, E., Schley, L., Overgaard, K., & Agbabian, J. (2009). Applying Lean techniques to improve the patient scheduling process. *Journal for Healthcare Quality*, *31*(3), 10-16.

Womack, J. P., & Jones, D. T. (1996). Beyond Toyota: How to root out waste and pursue perfection. *Harvard Business Review*, *74*(5), 140-172.

Yoon, M. H., Seo, J. H., & Yoon, T. S. (2004). Effects of contact employee supports on critical employee responses and customer service evaluation. *Journal of Services Marketing*. *18*(5), 395-412.

DECISION SCIENCES INSTITUTE

U.S. Ports of Entry's Decision Model with an Anti-Human Trafficking Focus

Priscila de A. Drummond
University of Texas Rio Grande Valley
Email: priscila.deazevedodrummond01@utrgv.edu

Hiram Moya
University of Texas Rio Grande Valley
Email: hiram.moya@utrgv.edu

ABSTRACT

Ports of the entry (POE) present the last chance to prevent people from being trafficked into the U.S. A decision-making tool that considers the human trafficking problem and its different costs at the POE is a necessity. To fulfill this goal, a new border security measures is presented using traveler's internet content analysis and a Bayesian decision model to improve identification and minimize costs. We showed, using a numerical example, how the security system would be used to improve the proposed security measures based on the expected monetary value from the possible outcomes.

KEYWORDS: Decision Analysis; Bayesian Decision Model; Decision Tree; Human Trafficking

INTRODUCTION

Human trafficking mobilizes worldwide attention. People trafficked are forced to work with almost nonexistent payment, most of the time in insalubrious conditions. Since the 2000's Palermo Protocols, international organizations work to improve their responses against this crime without hindering its rise (UN News, 2007; U.S. DOS, 2019). Surprisingly, most trafficked people may be entering the U.S. using legal routes (Bouche, 2018). Traffickers also use the border to employ different kinds of detectable criminal activities. They usually cross illegal profit to launder money in other countries facilitating their business practices (Shelley, 2003).

The most challenging aspect of this problem, however, is the diversity of forms this crime is committed. Finding an accurate model to analyze their operation is a complicated task. There are few commonalities between the practice of different types of exploitation (eg.: labor, sex, child exploitation). But, by definition, they all follow the same Modus Operandi (United Nations, 2003). First, they recruit people using false promises. Those promises go from jobs overseas to marriage. Then, they must transport the victims. This is a delicate part of the operation because of the flea risk. The last part is the actual exploitation. The recruitment and transportation parts are the ones most concerning the POE. Research from NGOs has shown that recruitment relies heavily on social media (Bouche, 2018). Hence, improving custom border procedures, with the help of social media analyses, could help deter it.

Anyone lawfully entering the United States should go through a Port of Entry (POE). The U.S. Customs and Border Protection (CBP) and the POE's have the responsibility to screen all visitors and products, as well as the American citizens crossing the border. In the U.S., in 2019, there were 328 land, air, and seaports divided into 20 operational sectors with more than 42,000 CBP officers and Border Patrol agents (U.S. CBP, 2019). The CBP already has as one of their function to identify potential trafficking victims as they seek to enter the U.S., but their focus is on the training of border patrol agents.

Effective management of border security requires considering different alternatives and constant improving measures to fill the gap in security and increase the flow at the border. Another concern is also not disrupting the free movement rights of people and goods between countries. Besides, improve systems of identification of transnational criminal activities, like human trafficking, can be the key its deterrence (Miller, Baumeister, 2013). The importance of building a framework combining Intelligence information and a good decision-making tool cannot be overstated.

CBP must find a way to effectively identify potential threats and victims in the border control apparatus but also consider the costs associated with each decision. People's freedom has to be taken into consideration, as well as, excessive delays at the border crossing process. Those don't only cause downsides to trade and tourism; it causes environmental costs and personal inconvenience too.

We present a screening framework capable of identifying people being trafficked during the screening process and a way to optimize the selection of travelers for a second screening. In practice, this tool can help POE officers make more knowledgeable decisions. This paper begins by demonstrating the problem of trafficking in person and POE current numbers. Then, we explain the theory behind the proposed screening process and the model used to analyze it. In the third section, a numerical example is given using public and assumed data, followed by the results and conclusion. We expect that this research helps in the development of a framework that can improve decision at the U.S. Ports-of-Entry, increasing the detection of human trafficking activities while considering the operation limited resources.

LITERATURE REVIEW

Verify the POE efficacy is a challenging task because a lot of information is unknown. Most of the published papers have the goal to prevent terrorists' attacks or increase economic efficiency of border management. To the best of our knowledge, no paper was found that dealt exclusively with optimizing POE operations to counter human trafficking activities.

Decision Models

There are different approaches when dealing with the improvement of decisions at the border management. Some authors (McLay, Adrian & Sheldon, 2010; McLay, Jacobson & Kobza, 2007) use statistical models and integer programming to optimize the probability of accurate assessment at the airport screening process. Moya & Rueda (2019) focused on a Deterministic Equivalent Problem to analyze the cost of outcomes depending on the assessment of potential threats by officers at a land POE between the U.S. and Mexico. Other focus on using a Game Theory model to give the POE agents a tool to help prevent the entrance of illicit cargo in the U.S. (Wein, Wilkins, Baveja, Flynn, 2006). The use of the Multi-Tree Committee (Romero et al., 2016) for the same purpose has proved to be a valid model. We have chosen to base our research in the paper that least rely on the use of physical technological tools (like x-ray) as an aid of the screening process (Majeske & Lauer, 2012). They use a Bayesian Decision Model to optimize the screening process at airports considering primary and secondary screening. They tested two models that consider the cost of security failure from the perspective of passengers and government. Their model was modified to assess our proposed measures in this paper. One thing that all papers above have in common is the screening of cargo, for commercial or non-commercial travelers, in at least one of the steps of the inspection. None of those works focus on a secondary inspection by human-specialists in trafficking in persons.

Human trafficking Identification at the U.S. POEs

For the automatic screening processes, we based our theoretical framework in some models that were developed over the last decades to help prosecute or predict the occurrence of human trafficking. Hultgren et al. (2018) used a knowledge management approach that analyses online advertisements to identify potential victims of sex trafficking. They improved identification using a machine learning model and text mining enhancements. Besides, Alvari, Shakarian, & Snyder (2016) presented a semi-supervised learning approach that discerns potential patterns of human trafficking. Andrews, Brewster, and Day (2018) improved the detection of organized crime with the use of a Natural Language Programming (NLP) model in a social media context. They separate content by geolocation and pre-determined concepts of organized crimes. They used a type of Social Network Analysis (SNA) model, the Formal Concept Analyses (FCA), to cluster the data media into Organized Crimes identifiers. This model will be the base to crawl and classify events from the user's database in the proposed framework. Besides, Rabbany, Bayani, & Dubrawski (2018) introduced a method that extracts connections using an algorithm called Active Search of Connections in the human trafficking network. They introduce user's feedback to guide inference about the relevance and connectedness of nodes. We are also using a feedback approach to improve efficacy through time.
We have considered that the U.S. government already uses the *US Automated Targeting System*, which already assigns a score of risk to travelers according to internal databases and profiling. This is used only as a tool to support the human decision, the border agent (Heyman, 2011). What we are proposing is a modification of the current operation to encompass the human trafficking focus.

METHOD AND MODEL

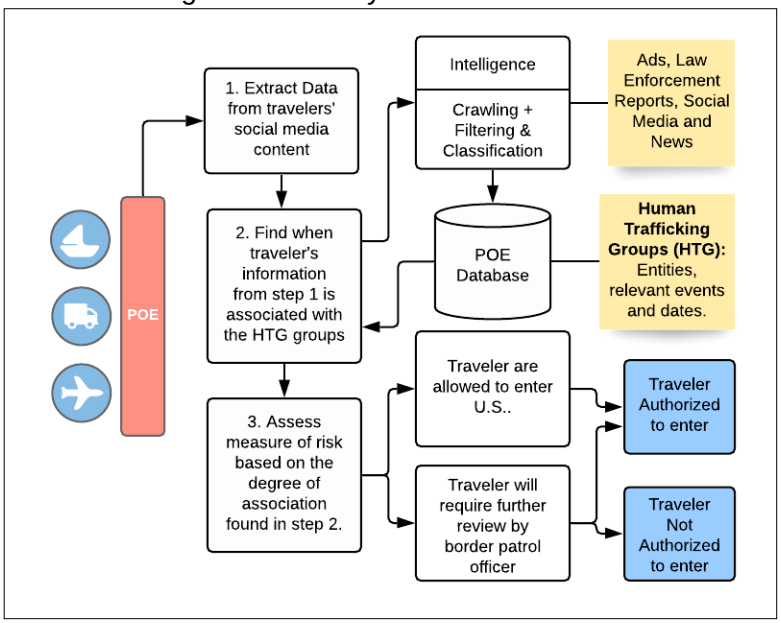
We propose a method to help the POE officers define when to call a secondary inspection at the screening process. The proposed model provides the number of risk factors (decision criteria) that someone would need to have that would minimize the cost for the U.S. government. We assume that, given the traveler's social media activities, we can expect a probability of risk that the traveler is being trafficked. The POE security measure for human trafficking detection is explained and optimized using a Bayesian decision model. We have used the expected monetary value for the possible outcomes and a Decision Tree to help us model the problem.

The POE Security Measure and the Proposed Model to Human Trafficking Detection

For national security reasons, how CBP assigns a measure of risk to individuals at the POE is classified information. But the Transportation Security Administration (TSA) and CBP provide some guidelines on how they proceed. It is known, for example, that the DHS started to analyze people's social media to look for any undesirable characteristics that can pose a potential threat to the U.S. (NY Times, 2019). At the air POE, TSA implemented in 2009 the program Secure Flight. It is a prescreening system, with names of people with known or assumed connections to the Terrorist Screening Database watchlist as soon as the person buys an airplane ticket and assign "risk score" to that person. This risk score is color-coded. The risk Assessments use intelligence-driven rule sets to identify the level of risk of the passengers (U.S. TSA, 2020).
At land POE, according to a GAO evaluation report (2019), the process is done in the moment of the border crossing. They run documents, names, license plates with known law enforcement and intelligence agencies databases. There is no indication that the POE uses any risk analysis model or decision-making tool. The orientation described in the same report is that if the individual or vehicle appears in any database, an officer will proceed to do further questionings in a second screening. (U.S. GAO, 2019). At sea POE, the process is similar. Additionally, they

require that the vessel should provide all documentation needed to proceed with a background check previous to the arrival.
According to the CBP, the current efforts to combat Human Trafficking at POE are restricted to training agents and that background check made by officials, according to CBP (U.S. CBP, 2019). The purpose of this paper is to address this threat, offering a decision-making tool that could help to utilize resources effectively based on the cost of possible outcomes. And since we have limited information regarding the risk assessment made by the DHS, we do not assume a possibility to compare with the current operation. The new security measure would make use of this apparatus to further improve the detection of human trafficking activities.
In Figure 1, the proposed model is illustrated. At first, the model uses an NLP model to extract all the information and proceed with an automatic generic cleanse and structure. The indicators used to crawl are pre-determined. At this stage, if the data exists, the model returns with people and organizations (names), events and dates, and geolocation from the traveler's media. The algorithm proposed to make it possible is the TwilCal (Andrews, Brewster & Day, 2018). The new security measure algorithm output is a structured file containing the data found in this first search. Then, using an FCA model with hand-crafted hierarchical rules based on known identifiers, the model classifies the data into human trafficking groups (HTG) of risk.

Figure 1: Security measure flowchart



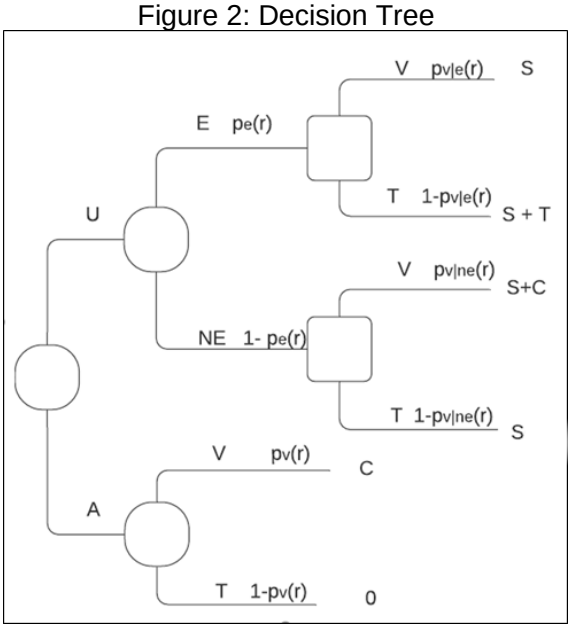
This number in the interval [0, G], in which G is the total number of HTG. This number is constantly changing since the model allows the new information from the traveler and Intelligence from other agencies to be added into the POE Database. The third stage will then compare HTGs found in the first stage to the POE mapped criminal network to identify, using an SNA model. It will look for all the possible connections between the traveler and the network. The output is going to be the number that represents how great it the connection between the trafficking network and the traveler. The final output will be a number between [0,∞], that we call Risk of traveler is being trafficked, ***r***. The measure ***r*** is, in practice, the quantifiable connections with events, dates, geolocation, activities, and people already known as human trafficking stakeholders by the United States agencies. Because we understand that not all people who have a high number of human trafficking connections necessarily are a victim or perpetrator,

this can be considered a possible measurement that will lead to further enquires at the POE screening process.
Utilizing tools that scrape and analyze the social media accounts in real-time, the system aims to assist border patrollers to make their final decision. In order to analyze the best value of *r*, which will lead to the best cost without great prejudice in the detection rate, we propose a Bayesian decision model.

Bayesian Decision Model

The performance of a classification scheme will depend on the distribution of the input. The model used is the one proposed by Majeske & Lauer (2012) adapted to fit the environment. The decision tree for this model can be seen in Figure 2 and is described as follows.

Figure 2: Decision Tree



Upon arrival at the POE, the traveler will go through the first screening. They can be a common traveler (T), or they could be a potential victim of human trafficking (V). In this part of the screening, the patrol officer will look for the traveler's documents, ask routine questions and run the documents, license plate, and biometrics through the database as explained in the section above. Depending on the number of undesired characteristics, r, the model will assign two possible categories for the traveler, authorized (A) or unauthorized (U). The authorized traveler will be granted entrance in the U.S., despite their real risk of being a victim, which is unknown. The unauthorized traveler, on the other hand, is going to be submitted for another screening. A specialized trained patrol office will conduct this screening. In this step, there are some extend of officer discretionary in the decision, as they perceive or not some suspicious behavior of the individual. The officer, then, is going to allow the entrance of the person to the U.S. depending on finding an extra piece of evidence or not (E or NE). This evidence can be hard evidence or subjective pieces of evidence. A few examples of possible indicators that officers can find are employment perspectives, employer analysis, psychological signs, distress signs, etc. If the officers do not find new evidence (NE), they will authorize the person to enter the country. Contrarily, if they find new evidence (E), the person will be forbidden entrance. In any of those scenarios, a wrong decision is possible. Thus, an authorized person (NE) can be a person at

risk of being trafficked. The opposite is also true, so the person can be forbidden entrance even though she/he is not at risk. For those with rejected entrance, the ideal policy would provide help and maybe resources so the person can corroborate to an ongoing investigation against this crime. But it is not under the POE's direct responsibilities, so we will not consider it.
For the scenarios presented in Figure 2 we can derive the probabilities as follows. The $\boldsymbol{p_v(r)}$ or $\boldsymbol{p(v|r)}$ is the probability of a traveler, given *r*, is going to be held in the U.S. in slavery conditions. Similarly, $\boldsymbol{p_t(r)}$ is the probability of, for a given *r*, a traveler is not be considered connected to human trafficking activities, which is the same as Equation 1.

$$p_t(r) = 1 - p_v(r) \tag{1}$$

We assume that as the screening process is looking for human trafficking indicators, so as *r* increases, $\boldsymbol{p_v}$ should follow. The errors can be computed as, on one hand, $\boldsymbol{p_{e|t}(r)}$, that represents the probability of finding new evidence in the second screening when the person is a common traveler (Type I error). On the other hand, $\boldsymbol{p_{ne|v}(r)}$ is the probability of not finding evidence in the second screening when the person is a victim (Type II error). The first one is easier to estimate by observation, but the second one is difficult since the number of people being trafficked is unknown. The probability of finding new evidence in the second screening, $\boldsymbol{p_e(r)}$**,** is given by the combination of the two errors as can be seen in Equation 2:

$$p_e(r) = p_{e|v}p_v(r) + p_{e|t}p_t(r) = \left[1 - p_{ne|v}\right]p_v(r) + p_{e|t}[1 - p_v(r)] \tag{2}$$

Assuming that the training for the specialized patrol officers is efficient and focus on finding new pieces of evidence, and the model is finding characteristics that are more common to the victims than to a common traveler, it can be inferred that as *r* increases so should $p_e(r)$. So, finding evidence or not is related to the risk factors (*r*) the person is assigned. When the probability of finding evidence is higher on victims ($\left[1 - p_{ne|t}\right] > p_{e|t}$), it corroborates with the assumptions of this model.
The posterior probabilities, $p_{v|e}(r)$ and the $p_{t|ne}(r)$, are the measure of the predictive validity of the model (Equations 3 and 4). The other two posterior probabilities are $p_{t|e}(r)$ and $p_{v|ne}(r)$, and they represent the errors in deciding to make the traveler go through a new screen (Equations 5 and 6). And since the $p_v(r)$ is a function of r, the all posterior probabilities will be also a function of r.

$$p_{v|e}(r) = \frac{p_{e|v}p_v(r)}{p_e(r)} = \frac{\left[1 - p_{ne|v}\right]p_v(r)}{\left[1 - p_{ne|v}\right]p_v(r) + p_{e|t}[1 - p_v(r)]} \tag{3}$$

$$p_{t|ne}(r) = \frac{p_{ne|t}p_v(r)}{p_{ne}(r)} = \frac{\left[1 - p_{e|t}\right][1 - p_v(r)]}{1 - \left[1 - p_{ne|v}\right]p_v(r) + p_{e|t}[1 - p_v(r)]} \tag{4}$$

$$p_{v|ne}(r) = 1 - p_{t|ne}\,(r) \tag{5}$$

$$p_{t|e}(r) = 1 - p_{v|e}(r) \tag{6}$$

To optimize the model, the evaluation of the cost of every outcome is going to be described in the next section.

Costs Associated with Each Outcome

Different from Majeske & Lauer (2012), we considered all the costs (from the State and traveler's perspective) together. The possible cost for each outcome is shown above in Figure 2 and are explained as follows. When a common traveler goes to a second screen and they are considered a possible victim, they are denied entrance (there could be other outcomes under the "denied" umbrella but they are not under POE officers responsibilities, thus they will be ignored). In this scenario, the expected cost is going to be the added security cost (S). S is an estimation of the cost which the government will have to implement the new security system, considering the possibility of increasing wait times, as well as the cost of hiring and training new employees. New facilities and accommodation are not going to be considered because it is assumed that a different room would not be necessary. Also, this decision would be an error, which lead to a cost (T) that can be computed adding the traveler's cost on non-reimbursed tickets, and the loss of legitimate job opportunities along with the U.S.'s government opportunity costs in tourism receipts. Tourism in the U.S. represented 2.8% of the Gross Domestic Product in 2018 (U.S. DOC, 2020) which would justify its addition.

In another outcome, the victim goes to a second screen and she/he is accepted, we are going to compute the added security costs (S) too. Although, we will have to add a higher cost since a victim are going to complete their cycle of exploitation. The cost of harm for the trafficked person and the Government's costs of law enforcement to prosecute this crime (C) are computed. The value is going to be an average of the cost associated with law enforcement and prosecution in cases of human trafficking in the U.S.

When a victim is accepted at the first step of the security system the only cost associated is going to be the cost of harm and prosecution (C). Then, if a traveler goes to a second screen and she/he is accepted. The only cost associated with this outcome is the security cost added (S). One could also associate it with an inconvenience cost. It would represent the cost for a common traveler that goes through extra procedures (e.g. late for a meeting). However, we considered the cost to be minor compared to the others then it was not included in the equation. The best outcomes regarding costs are when a real victim goes to a second screen and is denied access to the U.S., or when a common traveler is accepted on the first screen. The first one only incurs in added security costs (S) and the second, no cost. In future research, one could consider the value related to letting travelers enter the country (GPD) and the reduce in Wait-Time (Robert et al, 2014) at the POE as negative costs.

Optimal Risk Factor Value (Decision Criteria)

The optimal policy to the decision maker is the one that will minimize the Expected Monetary Value (EMV). To evaluate the framework, we have used costs for each outcome as described above. Since we want to optimize it, we need to minimize the EMV that is given by the cost and the probability of each outcome. $\boldsymbol{EMV_a}$ is the expected monetary value for a person authorized in the first screening (see Equation 7). The $\boldsymbol{EMV_u}$, EMV of the travelers send to a second screening, is going to be given in Equation 8.

$$EMV_a = p_v(r)\ C \tag{7}$$

$$EMV_u = p_e(r)\left[S + T\left(1 - p_{v|e}(r)\right)\right] + [1 - p_e(r)]\left[S + C\ p_{v|ne}(r)\right] = S + Tp_{e|t}[1 - p_v(r)] + Cp_v(r)p_{ne|v}(r) \tag{8}$$

In that case, to find the optimal policy for the new human trafficking detection system will require to find the minimum of the total expected cost with respect to some value of $\boldsymbol{r}$. The optimal value of *r* is the one that will equate the EMV costs and we are going to call $\boldsymbol{r_f}$**.** The total expected cost of the screening process regarding r is ES* (See Equation 9).

$$ES^*(r) = \begin{cases} p_v(r)C & if\ r < r_f \\ S + Tp_{e|t}[1 - p_v(r)] + Cp_v(r)p_{ne|v}(r) & if\ r \geq r_f \end{cases} \quad (9)$$

When the expected cost from authorizing the traveler at the first screening is equal to the expected value of having a second screening, it is indifferent to the expected cost if the officer patrol sends the person to further screen or not. That is the optimal state since we do not want to decrease the operational cost at the expense of greatly diminishing the rate of detention. In the ideal world, we would have infinite resources to implement additional security measures that would virtually capture all possible victims, but that is not feasible. So, the best policy would be to balance both costs, of increasing security measures and of letting a possible victim enter the country. In this case, when $\boldsymbol{EMV_u = EMV_a}$, so we can find $\boldsymbol{r_f}$ values (Equation 10):

$$p_v(r_f) = C + S + Tp_{e|t}[1 - p_v(r_f)] + Cp_v(r_f)p_{ne|v}(r_f) = \frac{S+Tp_{e|t}(r_f)}{C(1-p_{ne|v}(r_f))+T\,p_{e|t}(r_f)} \quad (10)$$

NUMERICAL EXAMPLE

A numerical example is provided to demonstrate the described model. All the costs and parameters are going to be estimated. We would like to reiterate that it is just a numerical example, and if tested with more accurate numbers, the result could change considerably. All the values assumed in this example would have to be estimated, in a real-world problem, using historical data for better accuracy.

For this example, first, we need to find an adequate distribution to $\boldsymbol{r}$, $\boldsymbol{f(r_f)}$. The exponential distribution (with parameter $\boldsymbol{\lambda_r}$) is a possible distribution to assume when considering the arrival of potential victims at the POEs having a specific measure of risk. It is expected that just a very low proportion of people are being trafficked in comparison to all the people trying to enter the U.S. The Department of State estimates that around 16,000 people are trafficked into the United States each year and the U.S. receive around 79.6 million of non-residents visitors annually (U.S. DOS, 2019). With that data we can make a roughly estimate of $\boldsymbol{p_v} = 0.0002$ and $\boldsymbol{p_t} = 0.9998$. In order to find the $\boldsymbol{p_{v|r}}$, we are going to use the same normal based Reflected Normal loss function as Majeske & Lauer (2012) but with the values adjusted to the problem presented in this paper. Since $\boldsymbol{r}$ has no intrinsic meaning one can set a value of **r** that matches the proportion of potential victims (0.0002). For this example, we will set $\boldsymbol{P(r > 100) = 0.0002}$ and solve for $\boldsymbol{\lambda_r = 0.08157}$. Using the Z distribution (Φ) we would have that the probability of $\boldsymbol{p_v}$, given by the Equation 11.

$$\boldsymbol{p_v = \phi\left(\frac{[r-\mu_v]}{\sigma_v}\right)} \quad (11)$$

We can see that the $\boldsymbol{\mu_v}$ represents the value for **r** thar results in $\boldsymbol{p_v(r) = 0.5}$. For the purpose of this example, we will assume that two people in five thousand travelers will have $\boldsymbol{p_v(r) \geq 0.5}$. This would lead to the results shown in Equation 12. The standard deviation of $\boldsymbol{\sigma_v}$ was assumed to be 7 to reflect the mismatch of the link between the risk factors found in the model and the potentiality of the person be victimized, producing a lower slope. A higher slope would represent

a more accentuated distinction between the number of factors of victim and non-victim traveler, which would probably not represent the reality. The figure 3 shows the $\boldsymbol{f}(\boldsymbol{p_v r})$.

$$\mu_v = -\ln(0.0004)/\lambda_r = 95.9182 \quad (12)$$

To roughly estimate the costs, we got the numbers from the official government and NGO's reports. The POE already carries a similar screening system with two levels, so the expected cost will only compute the increase in costs for the new detection system. For the additional security costs, we estimated based on current employment and average salary (U.S. CBP, 2019). For traveler's cost and U.S. losses on tourism, we used national data. We are computing the value of the person's held in slavery conditions as an estimate of the life insurance benefits, even though it is not the best measure, for the example should be enough approximation. Future research could use better methodologies for this estimation. The criminal justice costs range from 2,000 to 44,000 in violent cases (Hunt, Anderson & Sanders, 2017), due to the crime complexity, we estimated the higher cost and added $500 for police expenditures. Those estimated values can be seen in Table 1. The accuracy of the models proposed to the new security system will help us estimate a value for $\boldsymbol{p_{ne|v}}$, which in this numerical example is going to be 0.3. Let the $\boldsymbol{p_{e|t}}$, the rate of false positive be 0.01 (1%).

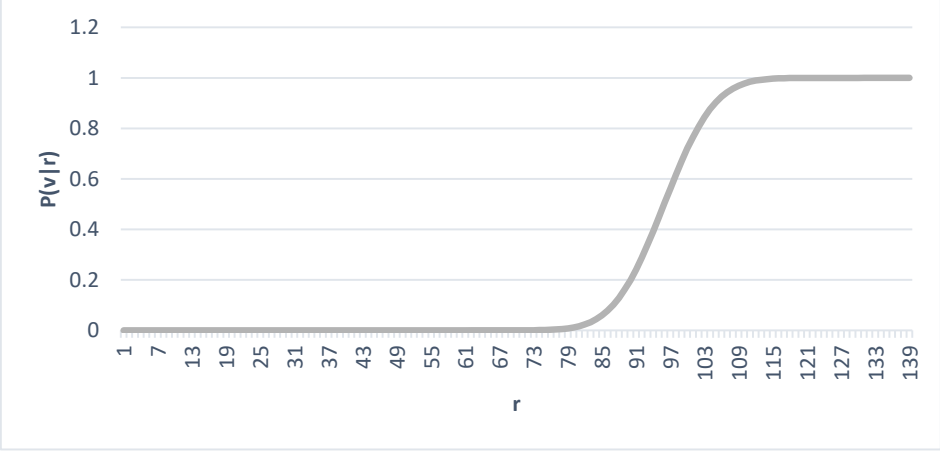
Table 1: Parameters used

Costs estimated	**US$**	**Other parameters estimated**	
Travel costs (T)	13,500	Pv	0.0002
Cost of type I error (C)	1,044,500	$Pv(r)$	$\Phi\left(\frac{r - 95.9182}{7}\right)$
Added Security Measures (S)	100	$p_{e\|v}$	0.01
		$p_{ne\|v}$	0.3

RESULTS

In Figure 3, notice that the conditional probability of the victims is dependent on r. Using the Equation 10 we find the $\boldsymbol{r_f}$ =72.0278, which means that the number to the risk factors found should be around 72 risk factors to a second screening to be worthy regarding costs of operation. The results of the outcomes and their given probabilities are presented in Table 2.

Figure 3: Probability of victims versus risk factors (*r*)



We can see that the proportion of passengers that are not authorized at the first screening is given by $\int_{r_f}^{\infty} f(r)\, dr$, which is 0.002808. This number represents about 223.5 thousand of non-residents per year being sent to a second screening. The probability of finding evidence in a second screening is given by the Equation 2, which is 0.010222 at r_f. This resulted is a little above 1% that the second screening will lead to new information. It is high, indicating that the cost of the second screening is not worth bellow the r_f. Around 1.02% of the people sent to be screened for the second time are real potential victims, but in a lower proportion, the victims are being authorized entrance. Results also indicate that almost 99.97% of the people that go to the second screening, and are assigned as possible victims, turn out to be common travelers. From the total that has been granted entrance to the U.S, only about 0.000097 are victims. The posterior probabilities are shaded in gray in Table 2 and are computed using Equations 3 to 6. They represent the performance of this policy, and how it will affect the daily operation and identification rate. The numbers show that the second screening leads to a better proportion of victims not authorized entrance at the second screening.

Table 2: Proportion of travelers screened by possible outcome and posterior probabilities

Outcomes	**All**	**Travelers**	**Victims**
Sent to second screen	0.002808	0.0000002	0.999677
Authorized entrance at first screen	0.997192	0.9897787	0.000323
Unauthorized entrance (New evidence found)	0.010222	0.999777	0.000223
Authorized entrance at second screen	0.989778	0.999903	0.000097

CONCLUSION

This paper uses a Bayesian Decision Model to the problem of identifying human trafficking activities at the border. It also offers an overview of a possible human trafficking detection framework. Using known models, this paper considers the expected cost associated with human trafficking and the POE operations. With the proposed framework, a model is offered to help assess decision criteria to improve human trafficking detection, while observing the costs associated with combating the illicit activity. Results offer a reduced number of second screenings while improving the identification probability at the secondary screening. Future research could adapt to other models, normally used in the criminal justice system, so they can be effectively used at the POE. It could also use this model to compare the identification rate of different NLP or ML model parameters when considering the POE costs. Moreover, they could try challenging some of the limitations of this paper, like accessing real data, find ground truth for the machine learning algorithms, and addressing issues related to security and data privacy from the travelers. The framework presented in this paper shows a possible way to implement different models to make a deeper search into the human trafficking activities at the POE improving its identification. We offered a tool, using this Bayesian Decision Model with the expected value cost of the possible outcomes, that helps improve decisions at the U.S. POEs.

REFERENCES

Alvari, H., Shakarian, P., & Snyder, J. K. (2016). *A non-parametric learning approach to identify online human trafficking*. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI). IEEE. 133-138. September.

Andrews, S., Brewster, B., & Day, T. (2018). *Organised crime and social media: a system for detecting, corroborating and visualising weak signals of organised crime online*. Security Informatics, 7(1), 3.

Borgatti, S. P. (2003). *Identifying sets of key players in a network. In IEMC'03 Proceedings. Managing Technologically Driven Organizations:* The Human Side of Innovation and Change IEEE. No. 03CH37502. October, 127-131.

Bouche, V (2018). Survivor insights: The role of technology in domestic minor sex trafficking". THORN. USA.

Heyman D (2011) *Testimony of David Heyman, Assistant Secretary, Office of Policy, before the House Committee on Homeland Security Subcommittee on Counterterrorism and Intelligence*. June 13.

Hultgren, M., Whitney, J., Jennex, M. E., & Elkins, A. (2018). *A knowledge management approach to identify victims of human sex trafficking.* Communications of the Association for Information Systems, 42(1), 23.

Hunt, P., Anderson, J., & Saunders, J. (2017). *The price of justice: new national and state-level estimates of the judicial and legal costs of crime to taxpayers.* American journal of criminal justice, 42(2), 231-254.

Majeske, K. D., & Lauer, T. W. (2012). *Optimizing airline passenger prescreening systems with Bayesian decision models*. Computers & operations research, 39(8), 1827-1836.

McLay, L. A., Jacobson, S. H., & Kobza, J. E. (2007). *Integer programming models and analysis for a multilevel passenger screening problem.* IIE Transactions, 39(1), 73-81.

McLay, L. A., Lee, A. J., & Jacobson, S. H. (2010). *Risk-based policies for airport security checkpoint screening*. Transportation science, 44(3), 333-349.

Miller, R., & Baumeister, S. (2013). *Managing Migration: Is border control fundamental to anti-trafficking and anti-smuggling intervention?* Anti-Trafficking Review, 2.

Moya, H., & Rueda, G. (2019). *A Deterministic Equivalent Problem to Study the Effects of Security Policies.* In New Global Perspectives on Industrial Engineering and Management. Springer, Cham, 117-126.

Polaris Project. (*2018) 2018 statistics from the national human trafficking hotline. Polaris: Freedom happens now*. USA. Retrieved from: https://polarisproject.org/.

Rabbany, R., Bayani, D., & Dubrawski, A. (2018). *Active search of connections for case building and combating human trafficking*. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2120-2129. July.

Roberts, B., Rose, A., Heatwole, N., Wei, D., Avetisyan, M., Chan, O., & Maya, I. (2014). *The impact on the US economy of changes in wait times at ports of entry.* Transport Policy, 35, 162-175.

Romero, P., Graneri, J., Viera, O., Moscatelli, S., & Tansini, L. (2016). *A Multi-Tree Committee to assist port-of-entry inspection decisions.* European Journal of Operational Research, 253(1), 170-177.

United Nations (2003). *Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the United Nations Convention against Transnational* Organized Crime. General Assembly resolution 55/25. Nov 15.

UN News (2007). *UN and partners launch initiative to end 'modern slavery' of human trafficking.* Retrieved from: https://news.un.org, March.

U.S. Customs and Border Protection (2016). *Performance and Accountability Report.* USA.

U.S. Department of Commerce (2019). *Travel and Tourism Sector Supports 7.8 Million Jobs and Accounts for 2.8% of U.S GDP" in 2018*. Retrieved from: https://www.commerce.gov/news/blog/2019/11/travel-and-tourism-sector-supports-78-million-jobs-and-accounts-28-us-gdp-2018, November 20.

U.S. Department of State (2019). Trafficking in Persons Report. June 10.

U.S. Transportation Security Administration. (2020). *TSA Secure Flight Program.* Retrieved from: https://www.tsa.gov/news/press/testimony/2014/09/18/tsa-secure-flight-program, January 18.

U.S. Government Accountability Office (2019). *Land Ports of Entry: CBP Should Update Policies and Enhance Analysis of Inspections.* Highlights of GAO-19-658, a report to the Committee on Homeland Security and Governmental Affairs, U.S. Senate. Retrieved from: https://www.gao.gov/assets/710/700758.pdf.

Wein, L., Wilkins, A., Baveja, M., & Flynn, S. (2006). *Preventing the importation of illicit nuclear materials in shipping containers.* Risk Analysis, 26 (5), 1377–1393.

DECISION SCIENCES INSTITUTE

Exploratory Analysis of Factors Affecting the Response Capability of Governments to COVID-19

Harmit Patel
University of New Haven
Email: hpate28@unh.newhaven.edu

Sagar Thakkar
University of New Haven
Email: sthak3@unh.newhaven.edu

Gokhan Egilmez
University of New Haven
Email: gegilmez@newhaven.edu

ABSTRACT

This paper proposes an exploratory analysis of policy making areas affecting countries' capability to responding and affectively coping with the COVID 19 outbreak. Data was collected from the government organizations and other official sources. To do so, Lasso, Ridge regression and neural network approaches were used, where the output was the fatality rate. The data consists of 208 countries, 20 number of input and output variables, the period between 01/21/2020 and 04/10/2020. The most important predictors were found to be international travel, hospital beds per capita, GDP per capita, cumulative total tests per capita, and international restriction decisions.

KEYWORDS: COVID-19, Policy making, Lasso, Ridge, Regression, Neural Networks

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the coronavirus 2 extreme acute respiratory syndrome (SARS-CoV-2). The disease was first detected in Wuhan, the capital of the Hubei province of China, in December 2019 and has since spread globally resulting in the ongoing coronavirus pandemic of 2019–20. More than 3.2 million cases have been registered across 186 countries and territories as of 30 April 2020 (*Coronavirus COVID-19 (2019-NCoV)*, 2019). Without vaccination, countries across the world are introducing various strategies including "social distancing" to slow down and eventually prevent the spread of the highly infectious virus. This social distance takes many forms, but its core aim is to keep people apart by confining them to their homes to reduce contact rates. Experts and many government regulatory agencies suggest 6ft social distance (Greenstone & Nigam, 2020).

Due to the unavailability of a vaccine, most countries responded with various forms of non-pharmaceutical interventions (NPI), including lockdown & curfew (home quarantine), social distancing, closure of schools / universities and closing of non-essential works (Gössling et al., 2020). Therefore, the virus has affected virtually various segments of the value chains across all industries and countries. Service industries majorly felt the impact due to cancelled public events (i.e. concerts, sports events, movie theaters, dining places, etc.), closed accommodations and shut-down attraction sites and public parks. Restaurants had to close as well, though a switch to take-away / delivery services allowed some to continue operations in some countries. In all these

developments, governments across the world, along with the support of citizens took actions to various degrees towards slowing down the spread of virus.

Indeed, while taking the actions such as lock downs, curfew restrictions; these actions resulted in significant number of economic loos and causes unemployment across the industries. Therefore, the disease not only tested the immune-system-wise resilience of people, but also test the governments' sustainability and resilience in terms of economy, healthcare system, and other resources. While taking response actions to stop the spread of the pandemic, economic and social impacts were severely felt, which also affected the infection trends and death tolls as a ripple effect. The virus has affected virtually various segments of the global and local supply chains.

In this context, several articles and preprints have been written in recent months to better understand the disease and its impacts. All of the above mentioned developments made is necessary to visualize and analytically study the pandemic data and associated data that could be used in conjunction with so as to understand the degree of relationship between the response management actions and the pandemic-related fatality rates across the countries. To do so, in this paper, we propose a data-driven exploratory analysis framework to understand the relationships of the government actions and availability of resources with the fatality rate because of the disease.

LITERATURE REVIEW

Several works have recently been conducted in various countries to identify the potential relationship of social, economic, and political factors with the spread of COVID19 and similar past pandemics. Various policy making interventions and their impacts on dealing with pandemics have been investigated. During a pandemic outbreak, it is important to control population mobility especially after a country lost its chance to stop the spread in the very beginning phase of the outbreak. In this context, population mobility, thus lock downs and restrictions, become necessary and have been employed as important policy initiatives by governments. Population mobility responds strongly to shifts in perceived prevalence of illness trajectory and stay-at-home announcements by the government. A rise in local infection rates from 0 percent to 0.003 percent decreases mobility by 2.31 percent as a ripple effect, and a government restriction order decreases mobility by 7.87 percent. It was also found as critical minimize the infection rate by closing non-essential places like cinema, restaurant, clothing stores. (Engle et al., 2020). Based on the epidemic of COVID-19 in mainland China, it was found that there is strong association between the social distancing and lockdown and rate of disease's spread. They also added that the stepwise implementation of social distancing & lockdown first in epicenter city later on state and entire country was found of importance (Lau et al., 2020). In Chicago most of the cases in February and March noted due to public gathering and parties(Ghinai et al., 2020). Findings indicate that the importance of adhering to current recommendations on social distance, including guidance on avoiding any crowded meetings and following state or local residence orders. As per the one investigation in the U.S. out of 95 confirmed cases, 35 cases reported due to visiting a religious place. Also, after 14 days preceding A lot of children with COVID-19 have more asymptomatic infections or They have milder symptoms and lower hospitalization rates than adults.

Travel is critically linked with the population mobility. A recent study conducted for the Hubei, China stated that controlling domestic air traffic helped slowing down the spreading rate of virus. While the described lockdown efforts in the Hubei province achieved measurable effects, it is also important to redirect the policy attention to the new epicenters outside of China to slow COVID-

19 spread. The rate of infections showed increasing trends within the states that were outside of lockdown (Lau, Khosrawipour, Kocbach, Mikolajczyk, Schubert, et al., 2020). A recent study also indicates that air travel is majorly important policy making area which could critically contribute to the spreading virus. In China after implementation of flight restrictions, they noticed significant decrease on number of cases. Therefore, controlling or banning domestic flights as well international flight created positive impact on the infection rates (Lau, Khosrawipour, Kocbach, Mikolajczyk, Ichii, et al., 2020).

In addition to government restrictions, there are areas that can hardly be controlled in the context of policy making but may still affect the spread of disease. One of the areas studied in the literature is the weather. Based on the latest numbers for the 2019-nCoV distribution, it would be important to consider the potential relationship between the number of cases and weather factors such as humidity and temperature. Bukhari found in some countries strikes the number of cases with increasing of the temperature. From the analysis we can show that the number of cases developed more in temperature range between 4 to 17 degree Celsius and humidity between 3 to 9 g/m^3. The higher number of cases also depending on number of testing per day, population of country, government policies and many more. (Bukhari & Jameel, 2020). In Jakarta spreading rate of COVID-19 virus was scientifically correlated with average of temperature(Tosepu et al., 2020).It was found that the virus lifetime is more and stable at cold places and also the airborne remain longer in dry air. As indicated in Wang's study conducted for China, the study suggested that the lockdown, public gathering limitations and increasing the health care system capacity in the short run could be of importance in reducing the transmissibility of the disease (Wang et al., 2020).

From the study of Raina, the country with a high age population and high income having a high death rate due to the virus. In Australia 16% of population age is more than 65 years where in China it is only 9%(MacIntyre, 2020). Peto Julian cites the urgent need for increased capacity to serologically test frontline health-care workers to check their immunity to the covid-19 virus (Peto, 2020). Frederik Verelst found that, In Italy work pressure was 8 times high on Health care system on 25th March compare to 11th March while at that time they have 12.5 critical care beds per 100,000 population(Verelst et al., 2020) which was second highest among the other world countries(Best Healthcare In The World 2020, 2020).Vishan Nigam thinking it seems fair to believe that social distancing would enhance the standard of treatment for non-COVID-19 medical conditions by reducing the burden on medical services , facilities and supplies (Greenstone & Nigam, 2020). Their lethality will decline slowly as 2019nCOV number of transmissions increases among the human population. During the early stage during the Covid-19 pandemic, the PIBA model correctly estimated 1.6% fatality for symptomatic patients during China. The model can be extended to predict case fatality for any infection, to predict the risk of serious diseases and to predict death rates for patients who contract serious diseases (L. Wang et al., 2020). Using a forecast study by Tanujit, he guided that we can reduce COVID-19 cases by keeping social distancing, isolated peoples above the age of 65 years, apply complete look down and increasing medical facilities (Chakraborty & Ghosh, 2020). By the travelling ban and closing restaurants China's daily growth rate dropped more than 7% in February 2020(Perc et al., 2020). From the SIRD model of Duccio Fanelli conclude that increasing ventilators helps to reduce fatality rate. In Italy and China, they are able to reduce 1% and 3% (Fanelli & Piazza, 2020). Using the PIBA method Lishi estimates the real-time mortality risk of a new infectious disease and forecasts potential deaths. The model can be extended to predict case fatality (including asymptoma) for any infection, to predict the risk of serious diseases and to predict death rates for patients who contract serious diseases (Chakraborty & Ghosh, 2020). Modified SEIR and AI prediction virus will reduce transmission and help to mitigate the impact of the potential peak to build new hospitals and quarantine centers designed to improve medical care and reduce exposure risks. The Wuhan

government increased the medical facility and government restriction that caused the control of spreading viruses in the city (Yang et al., 2020). The linear model forecasting of utilization of ICU beds in Italy helps the government regarding the Additional services, such as relocation of critically ill patients to other areas, emergency funding, staffing and ICU equipment .The objective is to ensure that every patient who requires an ICU bed is available (Grasselli et al., 2020). Forecasting using R for Hubei and outside Hubie in China, recovery and death data of COVID-19, reported publicly, are generally collected in hospitals where access to such information is warranted (Song et al., 2020). The Bai in their study in Wuhan Institute of Virology, inferred that when the index case start date was set, the approximate cCFR (estimated the risk of death from confirmed cases) value was 5.3%. The figures fitted with other model parameters on 8 December 2019 and 8.4% as explosive epidemic growth occurred(Bai et al., 2020). The epidemiological pattern of COVID-19 prevalence in Italy, Spain, and France, among the most affected European countries was predicted by the models Auto-Regressive Integrated Moving Average (ARIMA). ARIMA models are sufficient in future to analysis the trends of the outbreak and provide an insight into their epidemiological status such that it enables politicians and health agencies to prepare and provide services soon, including personnel, beds, and intensive care facilities(Ceylan, 2020).

METHODS

As COVID-19 disease is rapidly increasing across the world, it is important to analyze the impact of this disease across the world. To analyze the impact, we reviewed the fatality rate for total 208 different countries such that to know how each country cop up with this disease during pandemic situation. This paper focused on how country's fatality rate behaved with country's economic wellbeing, Average 1 quarter weather, government regulations regarding to prevent the spread of covid-19, healthcare system & testing capacity and tourism. As above mentioned, we considered all the necessary respective variables that might affected the fatality rate. To analyzed which variable, affect the fatality rate most or least, we conducted lasso regression analysis, Ridge regression analysis and deep neural network approach using machine learning.

The standard linear regression or the ordinary least squares method performs poorly with a large multivariate data set containing variables higher than the sample size(7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression - Statistics By Jim, 2018). The penalized regression allows a linear regression model to be generated which is penalized for too many variables by applying a constraint in the equation to the model for unnecessary variables. This is also known as methods of shrinking or regularization (Penalized Regression Essentials: Ridge, Lasso & Elastic Net - Articles - STHDA, 2019)(*Machine Learning Essentials: Practical Guide in R - Datanovia*, 2019). The consequence of imposing this penalty is that the coefficient values are reduced or decreased to zero. Hence, the variables which contributes less to output have coefficient is close to zero or equal to zero. The shrinkage requires the selection of a tuning parameter (lambda) that determines the amount of shrinkage. In this approach, first we conducted lasso and ridge regression analysis between the dependent and independent variables. Then, we split the entire dataset into train data set and test data set with 3:1 ratio. Lastly, we analyzed the model performance of test dataset.

Lasso Regression

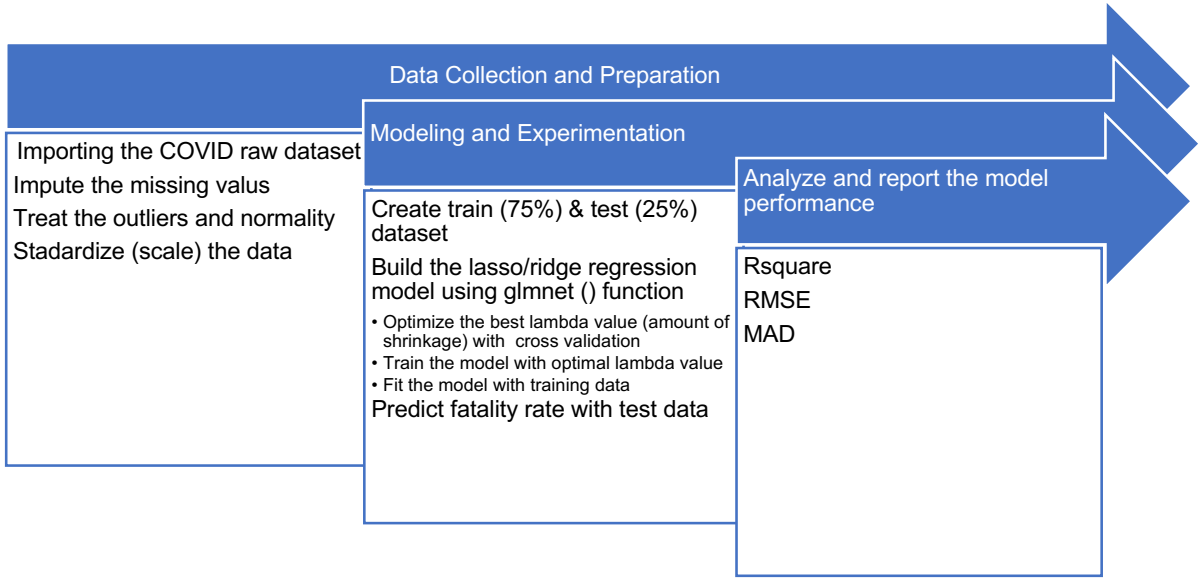
Lasso is defined as 'Least Absolute Shrinkage and Selection Operator'. Which means it reduces the regression coefficient towards zero by penalizing the regression model with lasso regression penalty. Lasso regression penalty is the sum of the absolute coefficient. To specify the amount of

coefficient shrinkage, we defined range of lambda value (amount of shrinkage). Then, with help of cross validation we found the best lambada value such that it produced the least amount of error with least bias & variance (Linear, Ridge and Lasso Regression Comprehensive Guide for Beginners, 2019) (*Machine Learning Essentials: Practical Guide in R - Datanovia*, 2019). From these different lambda value, we choose the optimal lambda value by replicating the lasso model ten times such that it minimizes cross validation prediction error rate. Optimal lambda value is calculated by taking the average of best lambda value which we obtain with each time replicating lasso model as given in APPENDIX B (sheet 9).
We conducted lasso regression approach between dependent variable and independent variables as described in APPENDIX B (Sheet 3) using RStudio. Flowchart for lasso regression modeling is provided in Fig. 1. To perform lasso regression, we assume elastic net mixing parameter α is equal to one & also we define range of lambda value (amount of shrinkage) from 10^{-2} to 10^{2}.From these range of lambda value, we choose the best lambda value by cross validation such that it minimizes cross validation prediction error rate.

Ridge Regression

Ridge regression shrinks the regression coefficient such that the variable with less contribution have coefficient closer to zero. The shrinkage of coefficient is achieved by penalizing the regression model with ridge regression penalty term which is the sum of the squared coefficients. To specify the amount of coefficient shrinkage, we define range of lambda value (amount of shrinkage). As lambda value increases the model's prediction sensitivity to the input variable decreases. Hence, ridge regression distinguishes higher and lower contributing input variables according to impact on output variable. Then, with help of cross validation we found the best lambada value such that it produced the least amount of error with least bias & variance. We got different best lambda value as we replicate this model. From these different lambda value, we choose the optimal lambda value by replicating the ridge model ten times such that it minimizes cross validation prediction error rate. Optimal lambda value is calculated by taking the average of best lambda value which we obtain with each time replicating ridge model as given in APPENDIX B (sheet 9). We built ridge regression models as described in APPENDIX B (Sheet 3). The flowchart for ridge regression modeling is given in Fig. 1.

Fig. 1. Lasso & Ridge Regression Procedure



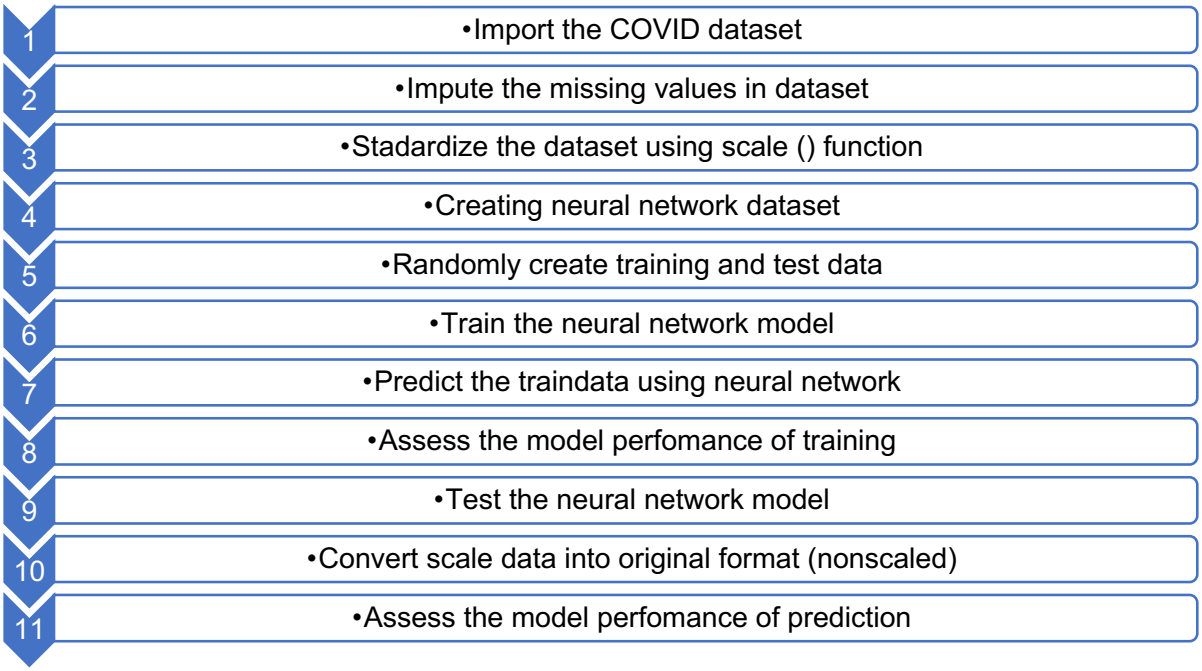
To perform Ridge regression, we assume elastic net mixing parameter α is equal to zero & also we define range of lambda value (amount of shrinkage) from 10^{-2} to 10^{2} same as lasso regression.

Neural Network

Artificial neural network (ANN) is consists of simple elements known as neuron, which is useful to make simple mathematical decisions and deliver this decision to other neurons which are properly organized in interconnected layers. It is flexible approach that helps to improve prediction accuracy of given dataset by combining neurons such that it helps to analyze impact of independent variables on the fatality rate. We create deep neural network with an input layer of independents variables, output layer of dependent variable (fatality rate) and three numbers of hidden layers. The numbers of neurons in each hidden layer is decided with rule of thumb method such as following:

The number of hidden layers is between the number of independent and dependent variables. Also, the number of hidden neurons is approximately 2/3 the size of input layer plus the size of output layer. By this rule of thumb methods we create deep neural network model with an input layer of 20 independent variables as described in APPENDIX B (sheet 3), with 16 number of hidden neurons in first hidden layer, 12 numbers of hidden neurons in second hidden layer, 9 numbers of hidden neurons in third hidden layer & 7 numbers of hidden neurons in forth hidden layer and with one output layer of dependent variable (fatality rate) as described in APPENDIX B (Sheet 3)(*The Number of Hidden Layers | Heaton Research*,2017). Following Figure 2 is the flowchart for the deep neural network model with above mentioned layer to analyze the model performance of test dataset.

Fig. 2. Neural network procedure



Model Performance

Model performance of test data set include R^2 value, Root mean square error and Mean absolute error of test data set. R^2 value indicates proportion of variation in dependent variable is explained by independent variables. The higher R^2 value indicates that model is better and well fit to data set. R^2 value of any dataset X is calculated by,

$$R^2 = 1 - \frac{\sum(Xi - \hat{X})^2}{\sum(Xi - \bar{X})^2} \qquad \text{(Eq. 1)}$$

where, $\hat{X} = Predicted\ value\ of\ X$, $\bar{X} = Mean\ value\ of\ X$.

Mean absolute error is the average absolute difference between observed and predicted values. The value of MAE indicates how close the observed data points are to the model's predicted values. Lower value of MAE indicates that observed value are very close to the predicted value. It is less sensitive to the outliers in the dataset compare to RMSE.
MSE of any dataset X is calculated by,

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{X}\right)^2 \quad \text{(Eq. 2)}$$

where, $\hat{X} = Predicted\ value\ of\ X\ and\ \bar{X} = Actual\ value\ of\ X$, and N is the total number of observations.
Root mean square error is the square root of the mean squared error (MSE), which is the average squared difference between the observed actual outcome values and the values predicted by the model. The value of RMSE indicates the absolute fit of the model to the data. Lower values of RMSE indicate better fit of model with data. RMSE is a good measure of how accurately the model predicts the output. RMSE of any dataset X is calculated by,

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{X}\right)^2}$$ (Eq. 3)

Where, $\hat{X} = Predicted\ value\ of\ X$, $\bar{X} = Actual\ value\ of\ X$, and N= Total number of observations.

We conduct regression approach and deep neural network model between dependent(output) and independent(input) variables to analyze which variable affecting fatality rate most or least as attached in Appendix

Data Collection and Preparation

To visualize the impact of covid-19 disease across worldwide, we collected data for total 208 countries to understand how the epidemic spreads, how the pandemic impacts people around the world and whether the preventive steps taken by respective government of countries are effective or not. The collected dataset refers to effect of covid-19 diseases for 208 different countries with respect to some other factors such as countries' economic wellbeing, Average annual weather, government regulations regarding to prevent the spread of covid-19, healthcare system & testing capacity and tourism.
As COVID-19 is current ongoing pandemic situation, we collected data up to 10th april,2020. The data was collected by using the following sources: Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) from various sources including the World Health Organization (WHO), DXY.cn. Pneumonia. 2020, BNO News, National Health Commission of the People's Republic of China (NHC), China CDC (CCDC), Hong Kong Department of Health, Macau Government, Taiwan CDC, US CDC, Government of Canada, Australia Government Department of Health, European Centre for Disease Prevention and Control (ECDC), Ministry of Health Singapore (MOH). These research paper accounts for a total of 27 different variables which provides the information about country and impact of COVID-19 in respected country with country's economic wellbeing, Average annual weather, government regulations regarding to prevent the spread of covid-19, healthcare system & testing capacity and tourism.
Table 1 gives the information about a total of 27 different variable which are necessary to analyze pandemic impact across the different countries.

Table 1. Input and Output Variables

Division	Variable Name	Abbreviation	Unit of Measurement	Source
Country	Country/Region Name	Country	Name of Country	(Roser et al., 2020)
Pandemic Impact	Fatality Rate of Country	Fatality. Rate	Percentage Rate of Fatality	(Roser et al., 2020)
Economic wellbeing	GDP per capita in $	GDP.per.capita.int.…	US Dollar	(Worldometer, 2020)
Healthcare system	Doctors (per 1,000 people)	Doctors. per. 1000.people	Number of Doctor/1000 People	(Roser et al., 2020)
Healthcare system	Hospital beds per 1000	Hospital.beds.per.1000	Number of Hospital Beds/1000 People	(Roser et al., 2020)
Weather	Weather (in F)	Weather.in. F.	Fahrenheit °F	(Weather Averages 2020)
Testing capacity	Cumulative total tests per thousand	Cumulative.total.test.per.thousand	Number of cumulative test/1000 People	(Roser et al., 2020)
Government Regulation	Curfew Restriction in Days	Curfew_DAYS	Total number of days from when country applied curfew restriction up to the day, when we collect our data.	(Roser et al., 2020)

Government Regulation	Curfew Restriction Binary	Curfew_BINARY	1 = Curfew restriction in country 0 = No curfew restriction in country	
Government Regulation	Partial Lockdown in Days	PL_DAYS	Total number of days from when country applied partial lockdown restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	Partial lockdown Binary	PL_BINARY	1 = Partial Lockdown in country 0 = No Partial Lockdown in country	
Government Regulation	Full Lockdown in Days	FL_DAYS	Total number of days from when country applied full lockdown restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	Full lockdown Binary	FL_BINARY	1 = Full Lockdown in country 0 = No Full Lockdown in country	
Government Regulation	International Restriction in Days	IR_DAYS	Total number of days from when country applied international restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	International Restriction Binary	IR_BINARY	1 = International Restriction in country 0 = No International Restriction in country	
Government Regulation	Domestic Restriction in Days	DR_DAYS	Total number of days from when country applied domestic restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	Domestic Restriction Binary	DR_BINARY	1 = Domestic Restriction in country 0 = No Domestic Restriction in country	
Government Regulation	Limit Public Gathering in Days	LPG_DAYS	Total number of days from when country applied limit public gathering restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	Limit Public Gathering Binary	LPG_BINARY	1 = Limit Public gathering Restriction in country 0 = No Limit Public gathering Restriction in country	
Government Regulation	Limit Import and Export in days	LPI.E_DAYS	Total number of days from when country applied limit import/export restriction up to the day, when we collect our data.	(Roser et al., 2020)
Government Regulation	Limit Import/Export Binary	LPI.E._BINARY	1 = Limit import/export Restriction in country 0 = No Limit import/export Restriction in country	
Tourism	International Travelers	International Travel	Number of international Travelers	(Roser et al., 2020)

Missing data

We initially conducted missing value analysis to understand the degree of missing data in each variable. The total missing datapoints and its percentages are shown in Table 2. The missing data amounts and % shares are completely attributed to the data availability at the time of data collection, but they are important to consider prior to modeling and experimentation.

Table:2 Number of Missing value percentage in each variable

Sr. No	Variables	No of Missing Value	% of Missing Value
1	Cumulative total tests per thousand	127	61.06
2	Limit Import and Export in days (LPI/E DAYS)	110	52.88
3	International Restriction in Days (IR_DAYS)	95	45.67
4	Domestic Restriction in Days (DR_DAYS)	90	43.27
5	Limit import/export Binary (LPI/E BINARY)	44	21.15
6	Limit public gathering in Days (LPG_DAYS)	42	20.19
7	Curfew restriction in Days (Curfew_DAYS)	27	12.98
8	Full Lockdown in Days (FL_DAYS)	24	11.54
9	International Travelers	21	10.09
10	Hospital beds per 1000	19	9.13
11	Doctors (per 1,000 people)	11	5.29
12	Partial Lockdown in Days (PL_DAYS)	11	5.29
13	Limit public gathering Binary (LPG_BINARY)	4	1.92
14	Domestic Restriction Binary (DR_BINARY)	3	1.44
15	Curfew restriction Binary (Curfew_BINARY)	2	0.96
16	Full lockdown Binary (FL_BANARY)	2	0.96
17	Partial lockdown Binary (PL_BINARY)	2	0.96
18	International restriction Binary (IR_BINARY)	1	0.48
19	Fatality Rate	0	0
20	Weather (in F)	0	0
21	GDP per capita (int.-$)	0	0

As shown in Table 2, the highest amount of missing values were found in variables such as Cumulative total tests per thousand (61.06%), limit import/export in days (LPI/E_ DAYS: 52.88%), international restriction in days (IR_DAYS: 45.67%) and domestic restriction in days (DR_DAYS: 43.27%).These variables may potentially affect the results of experimentation. On the other hand, there are no missing values found in Fatality Rate, Weather (in F), GDP per capita (int.-$) variables.

The missing values were handled by following the following rules:

1.If any continuous variable has more skewed data points than missing value of that variable is replaced with median of that variable.

2.If any continuous variable has less skewed data points than missing value of that variable is replaced with mean of that variable.

3.For categorical variable, we replace the missing value with mode of that variable (Data Mining Survivor: Impute - Mean/Median/Mode, 2010).

As given in the Appendix B (Sheet 7), The variable with skewness between –1 to 1 is considered as less skewed variable. Here, doctors per thousand people and limit public gathering in days (LPG_DAYS) have skewness between -1 and 1. Missing value of these variables are replaced with mean of that variable. The variable with skewness below –1 or above 1 is considered as highly skewed variable. Missing value of this variable is replaced with median of that variable. As given in Appendix B (Sheet 7) shows the mode value of each categorical variable. Missing value of this categorical variable is replaced with the mode.

Outliers

In table 3, the number of outliers and it's percent in the corresponding column are shown. Also, we conducted the outlier analysis to check the percentage of outliers in each variable. Results indicate that the extent of outliers ranged between 0 to 17% across the variables. To cope with the deteriorating impacts of outliers, we normalized and scaled all the variables between 0 to 1 as a standardization procedure prior to modeling and experimentation.

Table:3 Outlier percentage of continuous variable

Sr. No	Variables	Number of Outliners	% of Outliers
1	International travelers	22	10.58
2	Fatality Rate	13	6.25
3	Limit Import and Export in days (LPI/E DAYS)	13	6.25
4	Full Lockdown in Days (FL_DAYS)	12	5.77
5	International Restriction in Days (IR_DAYS)	10	4.81
6	Weather (in F)	8	3.85
7	Hospital beds per 1000	8	3.85
8	GDP per capita (int.-$)	7	3.37
9	Cumulative total tests per thousand	4	1.92
10	Domestic Restriction in Days (DR_DAYS)	4	1.92
11	Limit public gathering in Days (LPG_DAYS)	3	1.44
12	Doctors (per 1,000 people)	2	0.96
13	Partial Lockdown in Days (PL_DAYS)	2	0.96
14	Curfew restriction in Days (Curfew_DAYS)	0	0

Normality

To understand the extent of normality in the data, we took the skewness and kurtosis measures. Table 4 provides the results of skewness and kurtosis check. According to table 4, the highest positive skewness was found to be in full lockdown in days (FL_DAYS: 4.64) and highest negative skewness for international restriction binary (IR_BINARY: -8.12). Moreover, less skewed variables such as limit public gathering in days (LPG_ DAYS: -0.23), limit import/export binary (LPI/E _BINARY: 0.07) and curfew restriction binary (Curfew_BINARY) were evident. Additionally, the kurtosis results indicate that variables having kurtosis value is greater than seven is an indication of non-normality. Hence, extreme kurtosis issue was found in international restriction binary (IR_BINARY: 67.01), full Lockdown in days (FL_DAYS: 25.83), Cumulative total tests per thousand (26.56) and international travelers (19.42). Also, variable having the kurtosis value greater than seven are potential concern for normality. Also, we visualized the skewness & Kurtosis of variables and provided in APEENDIX B (Sheet 4).

Table 4: Skewness & Kurtosis of each variable

Sr. No	Variables	Skewness	Kurtosis
1	International restriction Binary (IR_BINARY)	-8.12	67.01
2	Full Lockdown in Days (FL_DAYS)	4.64	25.83
3	Cumulative total tests per thousand	4.35	26.56
4	International Travelers	3.85	19.42
5	Limit Import and Export in days (LPI/E DAYS)	3.01	11.33
6	Hospital beds per 1000	2.32	10.39
7	International Restriction in Days (IR_DAYS)	2.17	8.91
8	Limit public gathering Binary (LPG_BINARY)	-1.99	4.97
9	Fatality Rate	1.9	7.24
10	Partial Lockdown in Days (PL_DAYS)	1.84	7.84
11	Full lockdown Binary (FL_BANARY)	1.8	4.26
12	GDP per capita (int.-$)	1.79	7.07
13	Domestic Restriction in Days (DR_DAYS)	1.79	7.99
14	Weather (in F)	-1.67	6.00
15	Domestic Restriction Binary (DR_BINARY)	-1.39	2.93
16	Curfew restriction in Days (Curfew_DAYS)	1.08	2.50
17	Doctors (per 1,000 people)	0.95	3.71
18	Partial lockdown Binary (PL_BINARY)	0.5	1.25
19	Curfew restriction Binary (Curfew_BINARY)	0.4	1.16
20	Limit public gathering in Days (LPG_DAYS)	-0.23	4.08
21	Limit import/export Binary (LPI/E BINARY)	0.07	1.01

Correlation Analysis

In our dataset fatality rate was the dependent variable and the rest was independent variables. The highest correlation rate noticed for domestic restriction in days (DR_DAYS). Other than these variables, fatality rate is highest positively corelated with international travelers & highest negatively corelated with domestic restriction in days (DR_DAYS). Variables having an absolute correlated value of 0.01 is least correlated variables with fatality rate. Here, international restriction in days (IR_DAYS), curfew restriction binary (Curfew_BINARY), limit public gathering in days (LPG_DAYS), & limit import/export binary (LPI/E BINARY) variables are least correlated variables with fatality rate.

Table 5: Correlation analysis between fatality rate & other independent variables

Sr. No	Variable Name	Fatality Rate
1	Fatality Rate	1.00
2	Domestic Restriction in Days (DR_DAYS)	-0.15
3	GDP per capita (int.-$)	-0.13
4	Hospital beds per 1000	-0.13
5	Domestic Restriction Binary (DR_BINARY)	-0.13
6	Cumulative total tests per thousand	-0.11
7	International Travelers	0.10
8	Doctors (per 1,000 people)	-0.08
9	International restriction Binary (IR_BINARY)	0.07
10	Limit public gathering Binary (LPG_BINARY)	0.04
11	Weather (in F)	-0.03
12	Full Lockdown in Days (FL_DAYS)	0.03
13	Partial lockdown Binary (PL_BINARY)	-0.03
14	Limit Import and Export in days (LPI/E DAYS)	0.03
15	Curfew restriction in Days (Curfew_DAYS)	0.02
16	Full lockdown Binary (FL_BANARY)	0.02
17	Partial Lockdown in Days (PL_DAYS)	0.02
18	International Restriction in Days (IR_DAYS)	-0.01
19	Curfew restriction Binary (Curfew_BINARY)	-0.01
20	Limit public gathering in Days (LPG_DAYS)	-0.01
21	Limit import/export Binary (LPI/E BINARY)	0.01

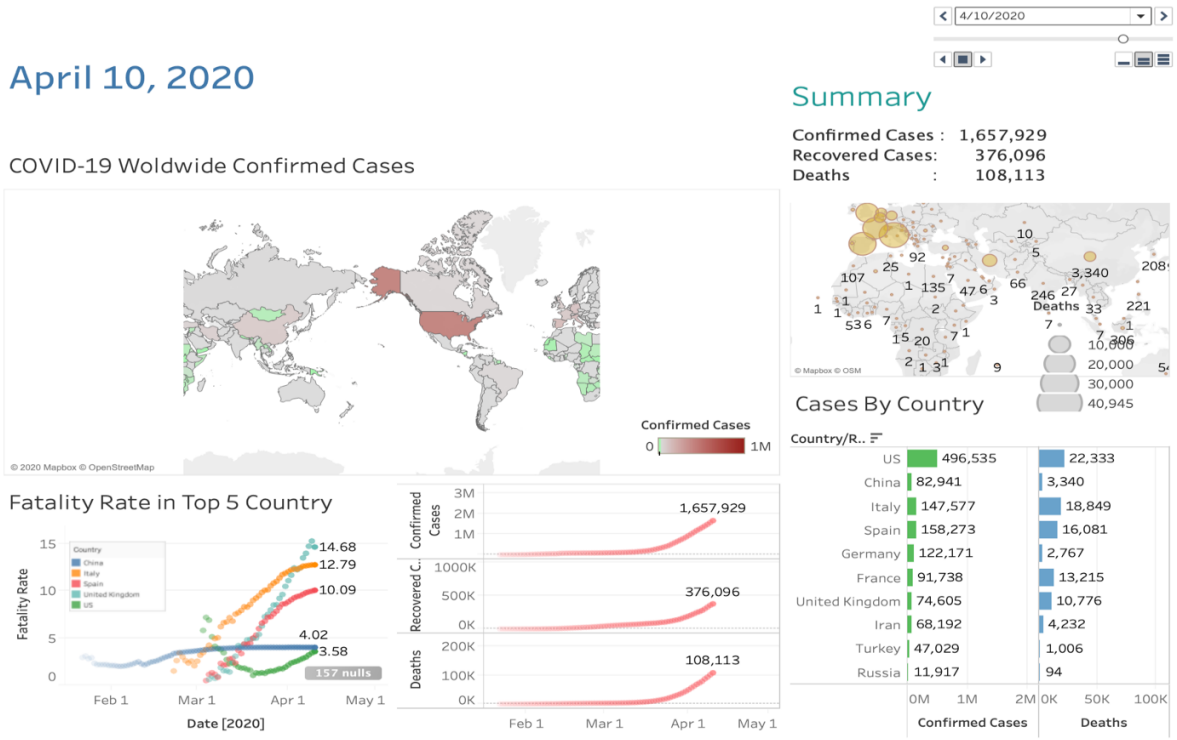
RESULTS

Data Visualization

The focus of visualization analysis (See Fig. 3) is the average daily statistics of the four main variables: reported incidents, deaths, recoveries, and fatality rate. The data apply to total daily cases and cover the period from January 21, 2020 to April 10, 2020. Data visualization part is carried out in Tableau and a link is also provided in Appendix-A (link 1).

The right-side top graph shows the number of death cases in worldwide. The size of circle indicates the number of cases. In above image the biggest circle with Italy with 27,682 which indicates the highest number of death rate. There are many countries death number is 1 which indicates with dot.

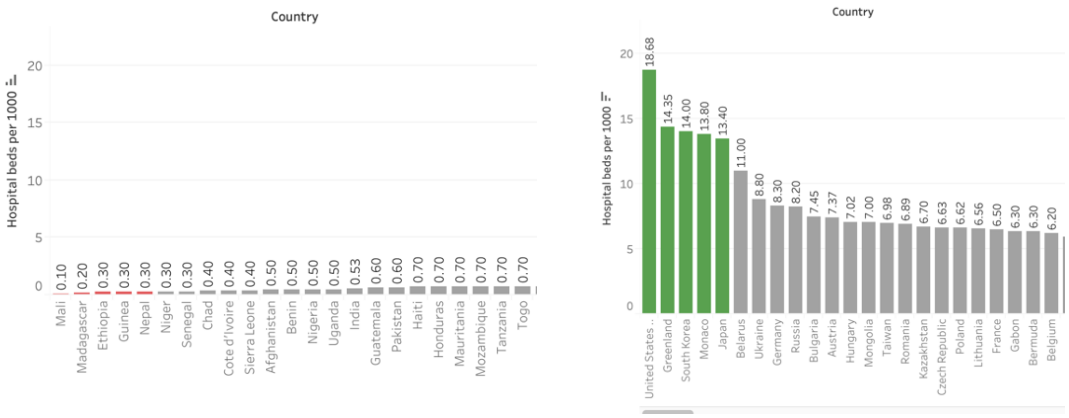
Fig.3. Data Visualization using Tableau



The world's top 10 country which facing highest number of cases and deaths due to COVID-19 shown in above horizontal bar graph in right side bottom. From this graph we can clearly see that USA is leading with 496,535 confirmed cases and 22,333 deaths. In below fatality graph include fatality rate of top 5 country include China, Italy, Spain, United Kingdom, USA. From the graph we can observe that United Kingdom (indicated as Light Blue color) has highest number of Fatality rate which is 14.68% that indicate that amount of people died due to coronavirus and least rate at USA which is almost same as China at 3.58%.

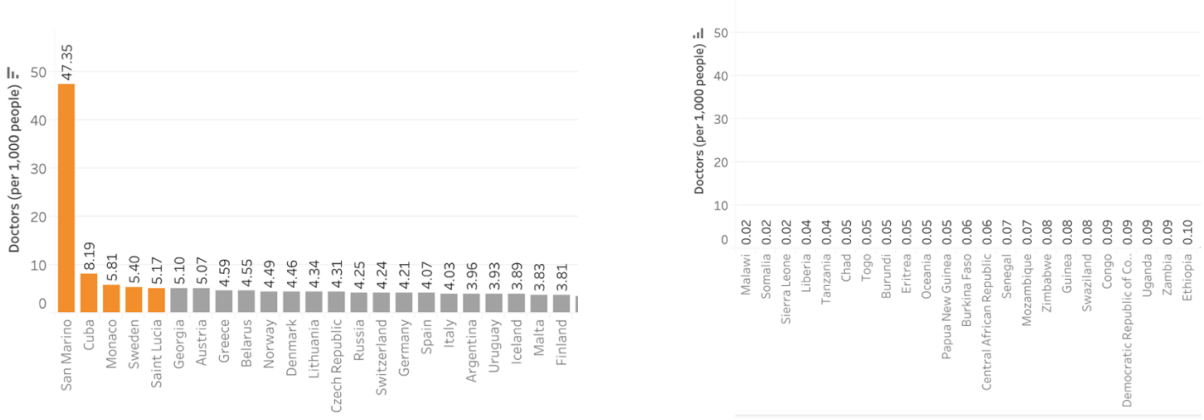
The dashboard includes graph of COVID-19 world confirm cases, Fatality rate in top 5 countries, Cumulative graph of (Death, Recovered and confirmed cases), Cases by top 5 country and world death graph in circle graph. This top left side chart shows the higher number of cases that shaded in red (Europe, United Kingdom) and which country has cases number less than 5000 that in green color. This bottom middle chart represents culminative line graph of confirmed cases, recovered cases and death. From the graph we can say that the after March 20th drastically change noticed in all the three categories. On the date of April 10^{th,} the total confirmed cases are 1,657,929, Recovered cases 376,096 and 108,113 total death happened worldwide.

Fig. 4 Graph of Hospital beds per 1000 people



This graph given in Fig. 4 (Appendix-A, Link-2) gives information of the number of Hospital beds available per 1000 people in 208 countries. The red color indicates least number of beds available (left side graph) and green color for highest number of beds available (right side graph) for different country. From this graph we can conclude that the most least number of hospital beds available were in Mali, Madagascar, Ethiopia, Guinea and Nepal and their rates are respectively 0.10, 0.20, 0.30 for both the country Nepal and Guinea. For the most beds available in United States, Greenland, South Korea, Monaco, and Japan which rates are 18.68, 14.35, 14.00, 13.80 and 13.40.

Fig.5. Graph of Doctors per 1000 People



Above both graphs give (Appendix-A, Link-2) information of available of number of Doctors per 1000 people. The orange color represents highest number of doctors available. The highest number of doctors available at San Marino which is 47.35 doctors per 1000 peoples. Later, that Cuba, Monaco at 8.19 and least number at 0.02.

Experimentation Results

Ridge Regression:

We conducted ridge regression analysis between dependent variable and independent variables using glmnet () function in R studio. Then, the optimal lambda value was obtained with cv.glmnet () function and then the model was fit with optimal lambda value such that input variables with minor contribution to outcome will have negligible coefficient values. The following equation was obtained from the ridge regression model:

Fatality rate = 0.15 + 0.033*(International Traveler) + 0.023*(international restriction binary) + 0.009*(limit import and export in days) + 0.008 *(full lockdown in days) +0.007*(partial lockdown in days) + 0.006*(limit public gathering Binary) + 0.002*(full lockdown binary) + 0.002*(curfew restriction in days) - 0.041*(cumulative total tests per thousand)-0.039*(domestic restriction in days) - 0.034*(hospital.beds.per.1000) - 0.026*(GDP.per.capita.int....) - 0.010*(domestic restriction binary) - 0.010*(doctors. Per. 1000.people.) - 0.006*(weather in F.) - 0.002*(partial lockdown Binary) - 0.002*(curfew restriction binary) - 0.002*(International restriction in days) - 0.001*(limit import/export in Binary) -0.001*(limit public gathering in days)

Furthermore, it was found that if a country has sustainable and enough hospital beds and doctors per capita than the fatality rate of that country were found to be low. Besides, government restriction such as number of days under domestic restriction, partial lockdown decision (Yes/No), limiting the import/export decision (Yes/No) and number of days under international restriction were found to have significant decreasing impact on the fatality rate. Whereas, the level of economic stability (GDP) of a country leads to decrease the fatality rate because one unit increase in GDP per capita leads to decrease fatality rate by 0.026.

Prior the fitting the model, we split 75% of the entire dataset into training dataset and remaining 25% of dataset into test dataset. For ridge regression elastic net mixing parameter value was taken as 0 while as for lasso it is 1. We trained the model, predicted the outcomes with the test data, and calculated the model performance as shown in Table 9. The R^2 value was found to be 0.25, which indicates that 25% variation in fatality rate was explained by the input variables considered. The RMSE was found to be 0.17 while the MAE was found to be 0.12. The error rates of the model were found to be acceptable while the explanatory power of the model was found to be acceptable but low.

Table 9: Model performance of test data using Ridge regression

Model performance of test data set for ridge regression approach	
R^2 Value	0.25
RMSE	0.17
MAE	0.12

Lasso Regression

The Lasso Regression Equation was found as follows:

Fatality rate = 0.18 + 0.096*(International traveler) + 0.012*(international restriction binary) - 0.125*(domestic restriction in days) - 0.078*(Hospital beds per 1000 people) - 0.066*(GDP.per.capita.int....) -0.050*(cumulative total tests per thousand) - 0.001*(domestic restriction binary)

It was found that out of the variables considered, 7 input variables such as international travel in terms of international restriction decision (IR_ Binary), domestic restriction time (DR_DAYS) , health care system capacity in terms of hospital beds per capita, economic strength in terms of GDP per capita, cumulative total tests per capita, domestic restriction decision (DR_BINARY) have critical predictive importance. All other variables have negligible importance as their coefficient value is zero. Out of 7 variables, international travel had the highest positive impact on fatality rate because one-unit increase in the variable lead to 0.096-unit increase in fatality rate. While as domestic restriction in days has the negative impact on fatality rate which means if domestic restriction in days increases, fatality rate of that country will decrease. Furthermore, increment in healthcare system leads to decrease the fatality rate because hospital beds per 1000 and cumulative total test have negative coefficient value in regression equation. Lastly, Table 10 depicts the average model performance of test dataset for the ten replications of the lasso model (see the full data of 10 models' results in APEENDIX B sheet 9). Similar results were also wound with Lasso. The model's explanatory power is 28% while acceptable RMSE and MAE results were found.

Table 10: Model performance of test data using Lasso regression

Model performance Parameter	Value
R^2 Value	0.28
RMSE	0.16
MAE	0.11

From the above output of model performance (Table 10), we can conclude as follows:

Neural Network Approach:

The model performance of test data set using neural network approach is given in Table 11. It was found that neural network gave significantly lower R^2, 0.06. Which indicates that only 6% variation in fatality rate is explained by all other input independent variables. Thus, we did not move forward with the sensitivity analysis of the input variables in the neural network model.

Table 11: Model performance of test data using Neural network

Model performance parameter	Value
R^2 Value	0.06
RMSE	0.26
MAE	0.18

DISCUSSION AND CONCLUSIONS

In this paper, we collected a COVID-19 data set for worldwide countries for different 27 variable include Weather, Government restriction, Air traffic control, testing capacity and many more. While different datasets (i.e. World Health Organization, CDC, Hopkins University, Ourwold indata.org) are used for different countries, we get clear estimates. Based on the epidemic of COVID-19 in mainland China and later that all over the world. To understand what type of policy making mechanisms and regulatory or non-regulatory actions would have critical impact on the fatality rate, we collected data to investigate the impact of a set of input variables on fatality rate.

The data was prepared by checking missing values, outliers, and normality, and split into 75% training and 25% test data. We developed ridge, lasso, and neural network models to conduct the exploratory analysis. Both Lasso and Ridge models provided acceptable error rates with explanatory power between 25% and 30% of the total variance in fatality rates based on ten replications were completed on each model. The neural network model did not yield a satisfactory performance in terms of explanatory power, which was around 6%. This is attributed to the neural network architecture and further experiments would be needed to optimize the number of hidden layers and activation functions.

In terms of important variables, based on the results of Lasso and Ridge models, the most important five predictors were found to be as International travel, hospital beds per capita, GDP per capita, cumulative total tests per capita, and international restriction decision. Furthermore, cases of COVID-19 disease are much higher per passenger number than they are on an international scale. In the foreign district of COVID-19 instances, air travel is a significant facilitator, while domestic COVID-19 spread is also associated with the other means of transportation, including trains and buses. Also, the health care system capacity is other important intervention measure and policy making area. The future extensions of the current work include further experimentation with other machine learning approaches including decision trees, support vector machines, and random forest. The data could be extended to have more variables that will add to the explanatory power of the models as well.

APPENDIX

APPENDIX A: Visualization

1.Below link provides the information regarding Tubule Visualization of total 208 countries data with its confirmed cases, recovered cases, and death numbers due to COVID-19
https://public.tableau.com/profile/harmit2381#!/vizhome/COVID-19_15875053039360/Dashboard?publish=yes

2. Following link gives visualization about the number of hospital beds and available doctors per 1000 peoples in respected country.
https://public.tableau.com/profile/harmit2381#!/vizhome/COVID-1910AprilByCountry/Dashboard1?publish=yes

APPENDIX B: Data and Experimental Results

Data file (See the link below) consists of the following content:
Sheet1: Raw Data: - Raw Dataset collected from different official sources.
Sheet2: COVID Dataset: - Data set which used to analyze the model.
Sheet3: Input and Output variable Information: - Detail information of input and output variable.
Sheet4: Histogram of Each Variable: - Plotting a histogram of variables to check the normality using RStudio
Sheet5: Ridge Coefficient: Impact of output variables on fatality rate using Ridge coefficient in RStudio
Sheet6: Lasso Coefficient: Impact of output variables on fatality rate using Lasso coefficient in RStudio
Sheet7: Missing value Imputation: Treatment of missing value in each variable
Sheet8: Mode of categorical Variable: Finding a Mode of binary variable for missing value imputation of categorical variable
Sheet9: Lambda value (Amount of shrinkage) & Model performance metrics for all ten replications of ridge and lasso regression
https://unhnewhaven-my.sharepoint.com/:x:/g/personal/hpate28_unh_newhaven_edu/EeI64h0w5BhJlh5JOYE74TMBKE_GWrph9o5PuQnUfdv_Eg?e=RGAvIt

REFERENCES

7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression - Statistics By Jim. (n.d.). Retrieved June 13, 2020, from https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/

Bai, Z., Gong, Y., Tian, X., Cao, Y., Liu, W., & Li, J. (2020). The Rapid Assessment and Early Warning Models for COVID-19. *Virologica Sinica*, *12250*. https://doi.org/10.1007/s12250-020-00219-0

Bukhari, Q., & Jameel, Y. (2020). Will Coronavirus Pandemic Diminish by Summer? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3556998

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of the Total Environment*, *729*, 138817. https://doi.org/10.1016/j.scitotenv.2020.138817

Coronavirus COVID-19 (2019-nCoV). (n.d.). Retrieved June 10, 2020, from https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

Engle, S., Stromme, J., & Zhou, A. (2020). Staying at Home: Mobility Effects of COVID-19. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3565703

Ghinai, I., Woods, S., Ritger, K. A., McPherson, T. D., Black, S. R., Sparrow, L., Fricchione, M. J., Kerins, J. L., Pacilli, M., Ruestow, P. S., Arwady, M. A., Beavers, S. F., Payne, D. C., Kirking, H. L., & Layden, J. E. (2020). Community Transmission of SARS-CoV-2 at Two Family Gatherings - Chicago, Illinois, February-March 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(15), 446–450. https://doi.org/10.15585/mmwr.mm6915e1

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*, *0*(0), 1–20. https://doi.org/10.1080/09669582.2020.1758708

Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy. *JAMA*, *323*(16), 1545. https://doi.org/10.1001/jama.2020.4031

Greenstone, M., & Nigam, V. (2020). Does Social Distancing Matter? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3561244

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., Bania, J., & Khosrawipour, T. (2020). The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. *Journal of Microbiology, Immunology and Infection*, *53*(3), 467–472. https://doi.org/10.1016/j.jmii.2020.03.026

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of Travel Medicine*, *March*, 1–7. https://doi.org/10.1093/jtm/taaa037

Linear, Ridge and Lasso Regression comprehensive guide for beginners. (n.d.). Retrieved June 12, 2020, from https://www.analyticsvidhya.com/blog/2017/06/a-comprehensive-guide-for-linear-ridge-and-lasso-regression/

Machine Learning Essentials: Practical Guide in R - Datanovia. (n.d.). Retrieved June 30, 2020, from https://www.datanovia.com/en/product/machine-learning-essentials-practical-guide-in-r/?url=/5-bookadvisor/54-machine-learning-essentials/

Penalized Regression Essentials: Ridge, Lasso & Elastic Net - Articles - STHDA. (n.d.). Retrieved June 12, 2020, from http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/153-penalized-regression-essentials-ridge-lasso-elastic-net/

Song, P. X., Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., & Eisenberg, M. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. In *medRxiv*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/2020.02.29.20029421

The Number of Hidden Layers | Heaton Research. (n.d.). Retrieved June 28, 2020, from https://www.heatonresearch.com/2017/06/01/hidden-layers.html

Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the Total Environment*, *725*. https://doi.org/10.1016/j.scitotenv.2020.138436

Wang, J., Tang, K., Feng, K., & Lv, W. (2020). High Temperature and High Humidity Reduce the Transmission of COVID-19. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3551767

DECISION SCIENCES INSTITUTE

The Implementation of Decision Tree Classification Methods in Engineering and Aviation Sciences to Examine the Design and Safety Performances using Project Based Learning

Willie Brown, Dinesh K. Sharma, Ibibia Dabipi, Lei Zhang, & Weiwei Zhu
University of Maryland Eastern Shore
Emails: {wlbrown, dksharma, ikdabipi, lzhang, wzhu}@umes.edu

ABSTRACT

The study explores methods in decision tree classification using project-based learning to examine set parameters and performance indicators associated with course learning objectives and outcomes for engineering and aviation education. This approach establishes techniques to support an assessment standard to map key measures with regard to performance outcomes in project-based learning. The ability to apply the model creates output measures in the environment of engineering and aviation education that identifies ways to implement task requirements and learning measures. This study will examine the decision approach using project based learning to determine the overall effectiveness of engineering and safety design models.

KEYWORDS: Engineering, Aviation Science, Safety, Performance, Project Based Learning, Data Analytics, Machine Learning, Decision Tree Classification Methods.

INTRODUCTION

The understanding of deep learning techniques in related practices can be overwhelming by the various approaches and challenges (Gosh et. al, 2019). These challenges are due to the unconstrained environment in addressing the specific types of performance decision in engineering design and development for image segmentation applications (Gosh et. al, 2019). Image segmentation is an application with considering the uniqueness of a technical perspective as it defines boundaries between separate semantic entities of an image (Gosh et. al, 2019). In understanding a modern conceptual approach, the aspect of operation, science and technology using real life application of fields involving engineering, transportation, and medical allows image-processing of other related areas. These problems can be defined using image segmentation of a specific (identified) target and clusters based on a digital imaging (Jaiswal, Sharma & Varma, 2019). As such, the application rationale can be applied to various practices and used for detection, recognition and processing according to the methods given in the applied application (Jaiswal, Sharma & Varma, 2019). The methods of image segmentation and the steps to object detection shows significant results on the datasets as the study will explore and examine the evaluation process for modeling purposes (Jaiswal, Sharma & Varma, 2019). The study has adopted R (2020) software as the language and environment for statistical computing and graphics.

In considering the classification and decision-tree model evaluation approach, the objectives aligned with areas of the module (2019) learning outcomes. This includes in summary the understanding of classification and methods for interpretation in decision tree methods with model evaluation and implementation strategy. The implementation strategy in this study of image segmentation considers properties obtain from the dataset (e.g., a group with regard to region, color, texture and intensity) (Jaiswal, Sharma & Varma, 2019). This technique using supervised

learning for local boundary detection could be introduced the necessary boundaries in each segment pertaining to the characteristics and optimal solution to initiate methods of statistical fit (Jaiswal, Sharma & Varma, 2019). The researchers in these fields mentioned the work to study image segmentation and identified problems to build on the proposed methods and functions. This paper will further explore the practices for advancement by providing a course case study using the data instances to best determine the specific applications of interpretation in this investigation (Jaiswal, Sharma & Varma, 2019).

PROJECT DESIGN TO DEMONSTRATE MODEL DEVELOPMENT FOR COURSE OUTCOMES

This section on analysis and demonstration of model development is presented to understand exploratory analysis and the data pre-processing approach applied. To examine the dataset, the study also explored grouping variables into factors and classes. This theoretical concept is described using qualitatively and quantitatively through exploratory analysis with the following outcomes:

1. Ability to evaluate the engineering design and aviation safety performance accuracy for application development;
2. Ability to apply and implement data-driven decision appropriate for techniques in the engineering design process; and
3. Ability to create an application that examines performance patterns and methods in handling decision for precise measuring and model development.

The modification of system environments provides the user different trajectories as defined in key applications as situational awareness. The evaluation of specified situations to understand the represented course objectives and learning outcomes to track relationships for clarification purposes. The project was able to simulate the most rational or efficient way to model results for interpretation based on uniqueness and the implications of course learning objectives and performance requirements. This development of requirements provided a significant effect on the applied each model and application design. The strategic approach also eliminated the concerns about introducing an unreasonable decision criteria; hereby, the optimization and development of feasible requirements were supported using a performance and validation accuracy matrix in the decision tree model. This is vital to the performance and creates a benchmark in the results for investigation purposes in the engineer design process.

This characterization of course performance in project-based learning has a significant impact on the properties to trace behavior and attributes of instances in the engineering design process (i.e., descriptive statistical analysis). The use of descriptive statistical analysis introduce an integrative approach for data collection and could be widely applied with regard to environment conditions and observation in the engineering design. This provides distinctive patterns in the decision tree model through association and prediction of classes using the parameters related to the statistical techniques. In exploring these relationships among items (variables), the rules and terms of the dataset could be proposed to forecast course performance according to project requirement and learning objectives. The analysis and model demonstration section explores the data used and contained to perform with the statistical software R. This approach using R language simplifies and enables access to data and manages statistical information from the point of view with regard to data management and acquisition that could aid in machine learning methods.

In the model development approach, the theoretical perspective of supervised learning implemented allow the various types and characteristics of the observations to reflect the performance and requirements relevant to the variables and classes. The modification of techniques will require a defined design approach that outlines the specifications and conditions to address the principles of performance challenges. Hereby, the data set and instances are adopted from the Center for Machine Learning and Intelligent Systems, which was randomly from a database of seven (7) outdoor images (UCI, 2019). According to UCI (2019), these images were hand-segmented by creating a classification for every pixel that includes an instance of 3x3 region. There are 19 areas in the attribute information of high-level numeric-valued with class and data set characteristics that introduce a multivariate totaling of 2310 instances (UCI, 2019).

The proposed method and framework described validates decision techniques analysis in engineering education. This examination was applicable to the course learning outcomes and contained attributes under certain conditions for project based learning. As these conditions were applied, the discovery of new opportunities with existing methods introduce solutions for project innovation. This insight into the application design to which the analysis and model is demonstrated derived on relationships among item sets from the course learning objectives. The approach may have several potential applications in engineering education and provides a degree of understanding to determine the various indicators of design requirements for project based learning. The goal of this assessment will be to study the targeted groups and examining the environment of course learning outcomes being deployed and delivered into the various projects.

METHODS: EXAMINING DATA TYPES IN THE ENGINEER DESIGN PROCESS USING DECISIONS AND APPLICATION STRATEGIES

In analyzing pattern of events, the objects in supervised learning and attributes are critical to both quantitative and qualitative methods. The quantitative approach created in this study was based on a processing method to examine applied environmental contexts and the hierarchy influences of the dependent outcomes (Leibovici, Bastin, & Jackson, 2011). Such variable selection methods in an exploratory analysis, this could also allow the creation of structures and scales for classification and regression trees through data in order to describe quantitatively (Leibovici, Bastin, & Jackson, 2011). Whereas, the qualitative exploratory analysis applied in this study examines the performance of a decision tree and the thought processes in which the effects of underlying structures are represented (Haworth, Bostani, & Sedig, 2010). This investigation supports the use of appropriate and interactive strategy in a qualitative exploratory study (Haworth, Bostani, & Sedig, 2010). By this determination, the require problem solving steps could reflect a systematic approach with analytical thinking and supervised methods in machine learning (Haworth, Bostani, & Sedig, 2010). This systematic approach includes the following: a) data preprocessing, b) decision tree model building, and c) methods to predict and examine related attribute of a decision tree. The study was designed to understand the decision tree and assigned rules according to the systematic approach. This is determined and measured to perform and classify each specific values with regard to the targeted variables (Ching-Chin, Yu-Jen & Bo, 2019). The display of the dataset structure is provided in Figure 1 below.

According to the findings of Cheng and Chen (2019), the study has proposed four (4) contributions in the data pre-processing steps. The first approach introduces proposed algorithms with unique identifiers to determine if the affected algorithms can be classified to support the results (Cheng & Chen, 2019). Hereafter, this study did not discover any proposed variables that could reduce

the accuracy of the results according to the first step. Since the data did not have any limits, the second approach in removing the unique identifiers and the continuation step was skipped in the second contribution. The examination of the second contributions in the data-processing steps were predicated on the first approach (Cheng & Chen, 2019). The third contributions in the data pre-processing steps featured the application of values and the validation of variables and components (Cheng & Chen, 2019). From the initial dataset structure findings, there were no variables in the class with numeric values. There are numeric values classifiers, but not as a class factor. Moreover, the correct validation with regard to the components are proposed in this study using the identified attributes and types. Lastly, the fourth contribution addresses the examination of parameters to perform the activities according to the desired methods as explained in the pre-processing steps. This approach ensures that all parameters are executed to support the requirements in decision tree methods. Figure two (2) explores the four contributions in the data-processing steps as explained in the section. The factors represented in figure two (2) indicates the class summary and the statistical function for evaluation.

```
> str(segment) #preview the structure
'data.frame':   2310 obs. of  20 variables:
$ region_centroid_col : int  218 113 202 32 61 149 197 29 1 69 ...
$ region_centroid_row : int  178 130 41 173 197 185 229 111 81 85 ...
$ region_pixel_count  : int  9 9 9 9 9 9 9 9 9 9 ...
$ short_line_density_5: num  0.111 0 0 0 0 ...
$ short_line_density_2: num  0 0 0 0 0 0 0 0 0 0 ...
$ vedge_mean          : num  0.833 0.278 0.944 1.722 1.444 ...
$ vegde_sd            : num  0.548 0.251 0.772 1.782 1.515 ...
$ hedge_mean          : num  1.111 0.333 1.111 9 2.611 ...
$ hedge_sd            : num  0.544 0.365 1.026 6.749 1.925 ...
$ intensity_mean      : num  59.63 0.889 123.037 43.593 49.593 ...
$ rawred_mean         : num  52.4 0 111.9 39.6 44.2 ...
$ rawblue_mean        : num  75.22 2.56 139.78 52.89 61.56 ...
$ rawgreen_mean       : num  51.222 0.111 117.444 38.333 43 ...
$ exred_mean          : num  -21.56 -2.67 -33.44 -12.11 -16.11 ...
$ exblue_mean         : num  46.8 5 50.2 27.9 35.9 ...
$ exgreen_mean        : num  -25.22 -2.33 -16.78 -15.78 -19.78 ...
$ value_mean          : num  75.22 2.56 139.78 52.89 61.56 ...
$ saturation_mean     : num  0.319 1 0.199 0.267 0.303 ...
$ hue_mean            : num  -2.04 -2.12 -2.3 -2 -2.02 ...
$ class               : Factor w/ 7 levels "brickface","cement",..: 5 3 6 5 5 5 4 1 3 1 ...
```

Figure 1. Image Segmentation's dataset structure

```
> summary(segment$class)
brickface    cement   foliage     grass      path       sky    window
      330       330       330       330       330       330       330
```

Figure 2. Class of image segmentation and statistical function of factors

RESULT INTERPRETATION AND MODEL EVALUATION

The experimental results is shown in this section using the proposed approach of image segmentation regarding the decision tree methods. For this reason, the technique utilized each

activities and focused on the analysis to implement classification features. The algorithms were used to evaluate segmentation and the measures based on image processing that are capable for methods of statistical fit in supervised learning development and classification. Therefore, the classification methods using the decision tree algorithm allowed in the accuracy of train modeling to obtain a threshold capable for testing. In addition, a sectional breakdown is provided in figure four (4) from the output results for review with regard to the tree structure. The overall output includes the conditional inference tree with 24 terminal nodes. The print response output had nineteen (19) class result inputs with a total of 1650 observations. As a result, the image processing technique accomplished in the study provides an approach to properly evaluate the decision tree classifiers and algorithms as it pertains to detection and recognition. The visualization of the tree was plotted in figure five (5) and figure six (6) below using the output print of nodes in the applied practices.

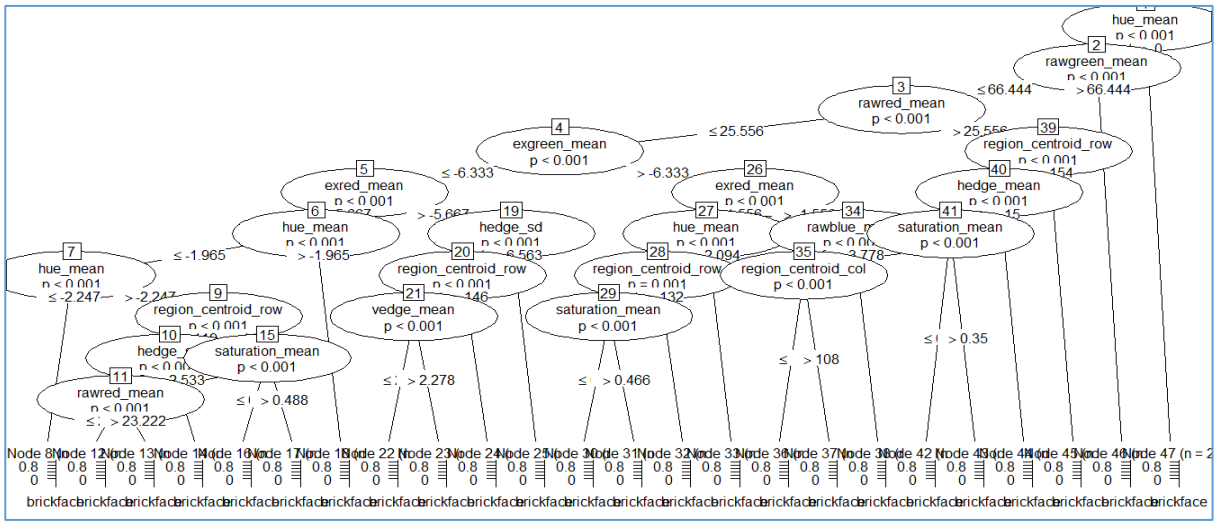


Figure 3. Sub-tree Starting from Node two (2)

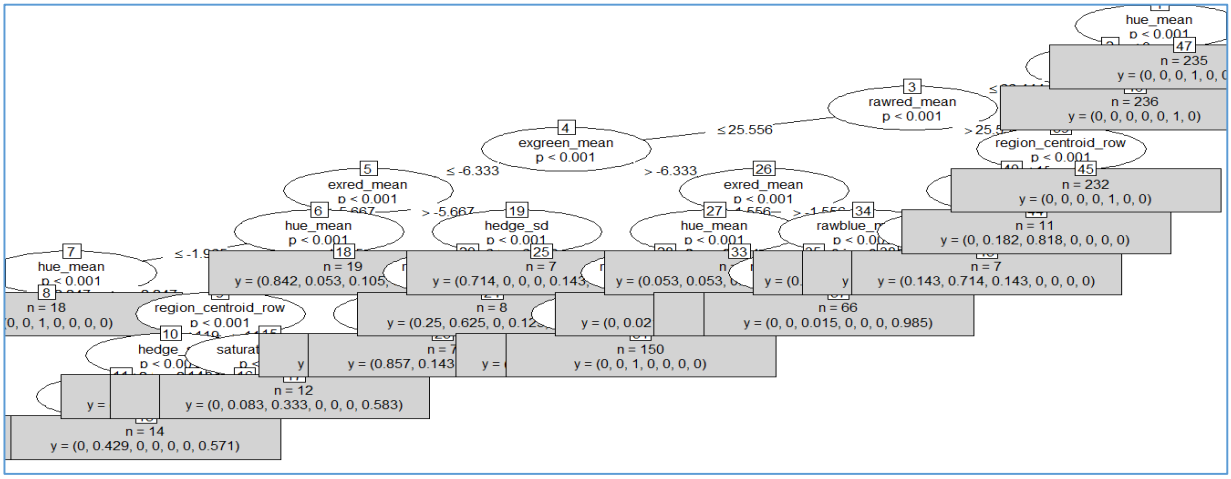


Figure 4. Tree Plot

As of the decision tree findings, node one (1) as illustrated in the results were plotted for evaluation with interest to understand the root node. From this variable name, the hue_mean was the splitting variable that determines the tree values. In node two (2), the rawgreen_mean was a direct value and attribute with rawred_mean as a direct child (node 3). Due to the complexity and learning process of the nodes found, the results could be considered as a demonstrative study. The result shows evidence of an outfitting model, which reinforces the approach of applying new data analysis techniques with regard to the decision tree learning as a useful tool for prediction (Choi, Song, & Lee, 2018). As the dominant value becomes the terminal node, the output and the results for the specific node shows that approximately 90% of instances have the class equal to 2. Subsequently, the proposed algorithm can improve the generalization of modeling and evaluation process of performance to implement conditions applied to the conventional tree methods (Choi, Song, & Lee, 2018). According to Choi, Song, and Lee (2018), this introduction and the improving of generalization in a new tree induction algorithm classifies and form a target functions in machine learning. These images are based on the detection strategy and supervised learning to support the use of nodes and the underlying mapping in the input/output variables. Successively, the understanding of the confusion matrix creates a visualization of the performance in support in addressing the proposed algorithm for machine learning.

The confusion matrix is the assessment indicator to cross-validate the various classification methods by achieving best accuracy rate of Support Vector Machine (SVM) (Choi, Song, & Lee, 2018). The number of correctly classified instances is the sum of (229, 221, 224, 235, 232, 236, and 199), which the total is equal to 1576. In comparison, the number of misclassified instances is the sum of (4, 1, 2, 1, 1, 1, 25, 5, 6, 3, 4, 9, 10, and 2), which the total is equal to 74. The number of instances in the training set was 1650. From the table output, the classification accuracy is 0.955 (95.5%). The number of machine learning techniques with the confusion matrix showed the relationships among attributes of each value in the intersection as a class instance and the corresponding probability. As stated in the course literature module and decision tree classification example (2019), “the sum of the numbers on the diagonal from upper left to lower right is the classification accuracy. The sum of all matrix entries is 1.” The study has introduce the outputs; in addition, this include the visualization as it relates to the sum of the numbers proposed in the confusion matrix for training set and the entries discovered using the prop.table() function.

```
> #confusion matrix
> table(predict(segment_ctree), train.data$class)

           brickface cement foliage grass path sky window
  brickface       229      4       2     1    1   0      6
  cement            3    221       1     1    0   0      5
  foliage           0      4     224     0    0   0     25
  grass             0      0       0   235    0   0      0
  path              0      0       0     0  232   0      0
  sky               0      0       0     0    0 236      0
  window            2     10       9     0    0   0    199
```

Figure 5. Detail Breakdown of Confusion Matrix for Training Set

```
> prop.table(table(predict(segment_ctree), train.data$class))

              brickface       cement      foliage        grass         path          sky       window
brickface 0.1387878788 0.0024242424 0.0012121212 0.0006060606 0.0006060606 0.0000000000 0.0036363636
cement    0.0018181818 0.1339393939 0.0006060606 0.0006060606 0.0000000000 0.0000000000 0.0030303030
foliage   0.0000000000 0.0024242424 0.1357575758 0.0000000000 0.0000000000 0.0000000000 0.0151515152
grass     0.0000000000 0.0000000000 0.0000000000 0.1424242424 0.0000000000 0.0000000000 0.0000000000
path      0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1406060606 0.0000000000 0.0000000000
sky       0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.1430303030 0.0000000000
window    0.0012121212 0.0060606061 0.0054545455 0.0000000000 0.0000000000 0.0000000000 0.1206060606
```

Figure 6. Prop Table

The evaluation of the model on a test data is described to build the confusion matrix as presented in the section. The ability to test data and map the accuracy of assessment based on the confusion matrix captures the variation in order to address classification error. The comparison of techniques within the classified imagery indicate in the results the classification accuracy and provides any uncertainty involving technique applied to image processing. The performance metrics to further verify the effectiveness of the proposed approach is critical to identify the defects using machine learning concepts and conditions. This draws a class distribution within the training and test dataset using the application associated with image segmentation. The number of correctly classified instances is the sum of (92+84+81+92+94+94+85), which the total is equal to 622. In comparison, the number of misclassified instances is the sum of (2+1+3+3+7+2+1+2+3+1+1+7+3+2), which the total is equal to 38. The number of instances in the training set was 660. From the table output, the classification accuracy is 0.942 (94.2%).

testPred	brickface	cement	foliage	grass	path	sky	window
brickface	92	0	3	0	0	0	1
cement	2	84	2	0	3	0	2
foliage	0	3	81	1	0	0	7
grass	0	0	0	92	0	0	0
path	0	1	0	0	94	0	0
sky	0	0	1	0	0	94	0
window	2	3	7	0	0	0	85

Figure 7. Detail Breakdown of the Confusion Matrix for Test Data

FINDINGS

The research findings project-based learning provides students with various chances to comprehend abilities that will be required. In our findings, the students would figure out how to work together and include inputs by building on reasoning and critical thinking abilities. This enables them to learn and apply learning outcomes to reach a resolution (see figures below).

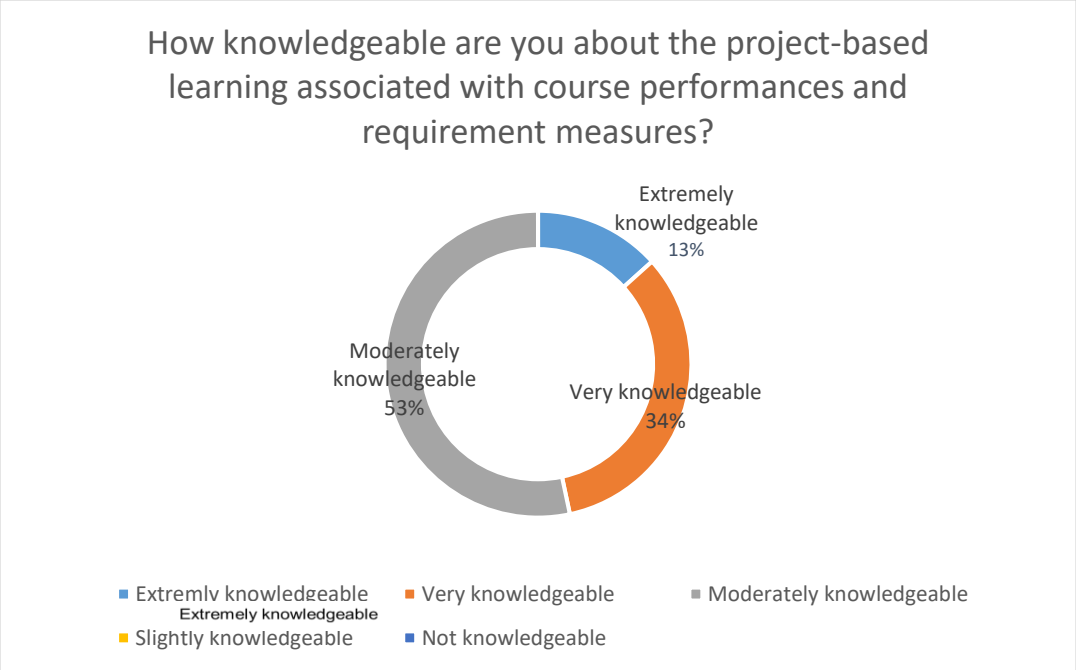


Figure 8. Students' knowledge about the project-based learning associated with course performance and industry partnership.

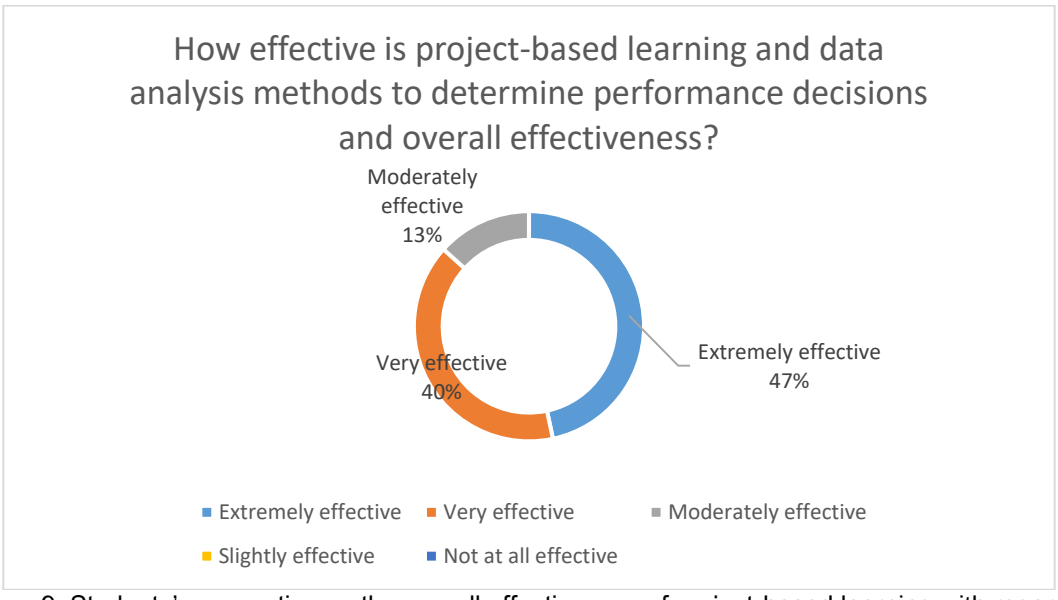


Figure 9. Students' perspective on the overall effectiveness of project-based learning with regard to course performance decisions.

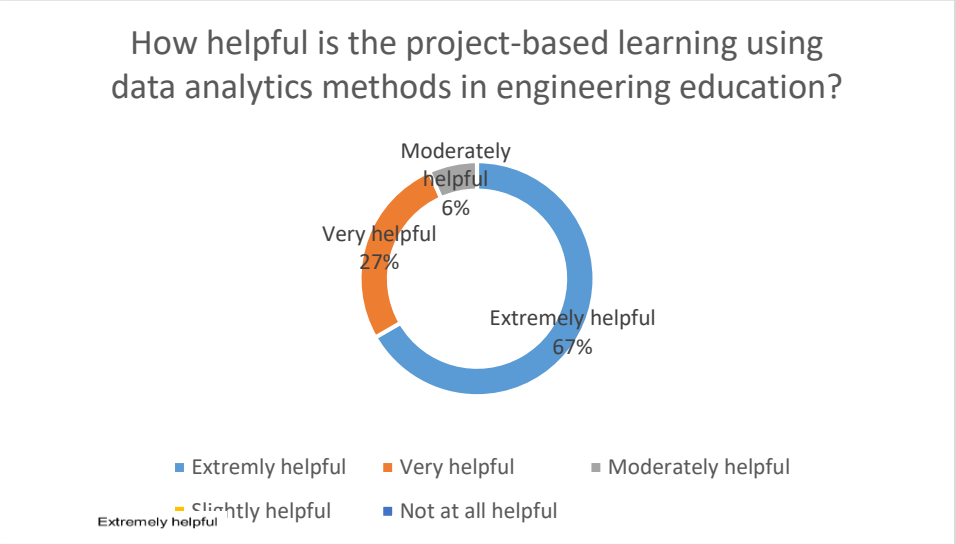


Figure 10. Students' perspective on how helpful is project-based learning with decision analytics to advance course performance decisions.

The purpose of these research questions was to provide a better understand the involvement of students with project-based learning. This model constructs learning significant to address real world issues; it establishes connections between the content by interacting course performances and decision methods. These connections formed allow students to understand there are opportunities for them to utilize the information they learned in the real world, particularly in the realm of engineering education. Project-based learning has advanced students by applying research technique abilities to expand curriculum and the course performance. Embracing this approach allows the researcher to explore the connections and better understanding of how project-based learning is affecting students. This research reports on a study that investigates the outcomes of project-based learning in course performance.

CONCLUSION AND FUTURE IMPLICATIONS

In conclusion, the applied practices of classification can be a complex problem and the assigned instances that best represents the entire dataset such as image segmentation presents a challenge for interpretation (Weinberg & Last, 2019). According to Weinberg and Last (2019), the decision tree models could create a challenge in the implementation approach that includes three (3) key factors – syntactic, semantic, and linear combination. In this study, the experiment to perform a test to classify and develop a decision tree model for evaluation of image segmentation highlights – the understanding of algorithms and challenge to ensure that the representative models selected must be compatible to the approach (Weinberg & Last, 2019).

In these findings, to understand the approach mentioned are best examined in describing the classification methods and proposed algorithms for addressing the problem. As such, this study adopted the supervised machine learning method and explored the target function of training data. Whereby, the methods of statistical fit used to estimate how well the approach was applied created limitation in parameters due to overfitting. For future implication, the selected dataset and model should be pruned and remove some of the details to perform an evaluation based on the sampling of data. In considering this approach, the study could determine the performance measures for validation by adopting the resampling methods and approach with regard to

classification and decision tree modelling evaluation. These implications are relevant in the design method to learn new discoveries in machine learning for modeling. The limitations and suggestions of this study could give a baseline to measure the applied application and aviation safety performances efforts. The performance efforts and factors considered as suggested should understand methods of statistical fit with regard to supervised learning based on the classification and decision tree model evaluation of image segmentation and processing.

Project based learning in engineering education is challenged depending on the dimensionality of the data due to the underlying course performance measures. In most cases, the performance indicators and outcomes are different factors (i.e., attributes, data sets inconsistencies and data types). Due to these reasoning, the first steps are critical in understanding the data performance measures and to become familiar with the variables. Hence, we have presented in this study a framework to ask if the first steps are being considered. This will aid in the engineering design practices to solve a problem and address requirements that are measurable in a pre-defined process. Overall, this is important using the various evaluation methods in order to become familiar with proposed engineering strategies used in the decision making process.

REFERENCES

Barbiero, A. (2014). Data transformation for confidence interval improvement: An application to the estimation of stress-strength model reliability. Advances in Decision Sciences.

Cheng, H., Feng, D., Shi, X., Chen, C. (2018). Data quality analysis and cleaning strategy for wireless sensor networks. Journal on Wireless Communication and Networking.

Ching-Chin, C., Yu-Jen, C., & Bo, H. (2019). Decision tree-based classifier in providing telehealth service. BMC Medical Informatics and Decision Making, 19(1), 1-15

Choi, J., Song, E., & Lee, S. (2018). L-tree: A local-area-learning-based tree induction algorithm for image classification. Sensors, 18(1)

Dua, D. and Graff, C. (2019). UCI machine learning repository. Irvine, CA: University of California, School of Information and Computer Science.

Ghosh, S., Das, N., Das, I., & Maulik, U. (2019). Understanding deep learning techniques for image segmentation, ACM Computing, 52(4).

Idri, A., Benhar, H., Dernandez-Aleman, J. L., Kadi, I. (2018). A systematic map of medical data preprocessing in knowledge discovery. Computer Methods & Programs in Biomedicine, 162, 69-85.

Jaiswal, V., Sharma, V., & Varma, S. (2019). An implementation of novel genetic based clustering algorithm of color image segmentation. Telkmnika, 7(3), 1461-1467

Leibovici, D., Bastin, L., & Jackson, M. (2011). Higher-order co-occurrences for exploratory point pattern analysis and decision tree clustering on spatial data. Computers & Geosciences, 37(3), 382-389

Luo, Q., Guo, C., Zhang, Y. J., Cai, Y., & Liu, G. (2018). Algorithms designed for compressed-gene-data transformation among gene banks with different references. BMC Bioinformatics, 19(1).

Haworth, R., Bostani, S., & Sedig, K. (2010). Visualizing decision trees in games to support children's analytic reasoning: Any negative effects on gameplay? International Journal of Computer Games Technology, 1-11

R-Project. (2020). http://www.r-project.org

Tran, D., T., Avard, A., & Jorm, L. (2017). Data cleaning and management protocols for linked perinatal research data: A good practice example for the Smoking Maternal Use of Medication and Safety (MUMS) study. BMC Medical Research Methodology, 17, 1-15.

Weinberg, A., I., & Last, M. (2019). Selecting a representative decision tree from an ensemble of decision-tree models for fast big data classification. Journal of Big Data, 6(1), 1-17

Piracy or Privacy in Video Streaming? A Game-Theoretic Model

Yiwen Gao
Temple University
yiwen.gao@temple.edu

Subodha Kumar
Temple University
subodha@temple.edu

Emre Demirezen
University of Florida
emre.demirezen@warrington.ufl.edu

ABSTRACT

In spite of huge revenue loss from piracy and account sharing, firms show contradictory attitudes, which make account sharing behavior's impact on profit unclear. Although it is critical for firms and users, it has not been rigorously examined. Hence, by considering privacy and blocking concerns, we explore how technology investment and sharing level shapes streaming service firm's performance. We find that technology investment has different impact in monopoly and duopoly markets. Furthermore, entrant's sharing level's impacts on sharing market size and entrant's profit depend on incumbent's sharing level and technology investment. Our results provide practical insights to streaming service firms.

KEYWORDS: Privacy, Piracy, Video Streaming, Technology Investment

INTRODUCTION

The growth of video streaming services has been tremendous in the last decade. The revenue of video streaming services in US reached $11 billion in 2019 and is expected to reach around $13 billion in 2024 (Statista 2019). Furthermore, the global revenue of video streaming services is expected to exceed $184.3 billion by 2027. However, with the gradual growth of streaming services, the account sharing behavior is also increasing. According to a report, 27% of US broadband households engage in some form of piracy or account sharing, and 20% use a piracy app or jailbroken device (ParksAssociates 2019). They estimate that in 2024, the expected global lost revenue due to piracy and account sharing will reach $12.5 billion (Lewis 2019).

In spite of the huge amount of revenue loss from piracy and account sharing, the firms show totally different strategies. For example, Netflix provides different plans to enable account sharing among friends and families, while Hulu limit to one simultaneous screen per account. However, Hulu recently relaxed the limit to two screens and Netflix started to crack down account sharing behavior (Despotakis 2019). The contradictory attitudes of streaming service firms and the changes of their attitudes make the impact of account sharing behavior on the profit unclear. Although it is an important issue for firms and users, it has not been rigorously examined. Hence, in this research, we attempt to explore how the account sharing behavior shapes the price and profit of the streaming service firms. We also study how the sharing behavior influences the other firm's revenue as well as how the strategies shape the other firm's revenue.

Motivation

Users pay a monthly or annually fee to the streaming services to get access to the content on the platforms. Part of the users pay the fee alone, while others cannot afford the fee and share the account with other users and split the bill with all that account users (Heinzman 2019a). However, these account sharing behavior raise two problems for the users: privacy concern and blocking concern. While sharing account to the acquaintance, the users could worry about their sensitive information will be exposed to their acquaintance. For example, Amazon Prime Video is connected with Amazon account. If the users share their account to others, their shopping preference and billing address may be shown to others (Heinzman 2019b). Furthermore, sharer also raises the risk of their other accounts getting hacked (Cybersecurity 2018). A recent survey shows that one-third Americans shared password with the coworkers, and 22% people reuse the same password for multiple accounts (Williams 2019). Furthermore, the sharing account behavior can raise the concerns that the user account could be blocked if the streaming service firms have invested in technology to detect whether the account has been shared and to what extent it has been shared (Vincent 2019; Welch 2019).

From the perspective of streaming service firms, although Amazon Prime Video users may naturally deter from password sharing (because of the privacy concerns), Hulu needs to enforce a strict policy on the number of simultaneous screens. With similar price (e.g., Netflix $7.99, Hulu $7.99, Amazon prime video $8.99), the number of users not paying on Amazon Prime Video and Hulu are both 5 million, however, Netflix (without natural deterrence or a strict policy) has 24 million users not paying (Swaminathan 2019). Besides, the existing platforms like Hulu and Disney invested in the technology to detect sharing behavior either by themselves or through third party services (e.g., Synamedia) (Vincent 2019). However, newly launched streaming services (e.g., Quibi) usually do not invest in technology to detect sharing.

Research Question and Contributions

Our research contributes to the information goods sharing literature (Tunca and Wu 2013) and streaming services literature (Aguiar et al. 2018; Lu et al. 2019) by answering how account sharing behavior shapes the market and firm performance.

Does technology investment always pay off? This research question raises a long term debate and two streams of literature holds two contradict standpoints (Brynjolfsson and Hitt 1996; Zahavi and Lavie 2013). The productivity paradox of technology investment is suggested to be better understood by considering the strategic role of information technology (IT), which is dominantly determined by the sharing level in this research (Dehning et al. 2003). Thus, this research raises the research question: Does the technology investment in detecting sharing increase the monopoly profit? Disney earned more profit within studio industry after adopting Synamedia technology to detect account sharing behavior (Szalai 2020). This suggests that high technology investment leads to high profit. However, our research finds that more technology investment can boost profit only when the sharing level is either high or low, but it hurts the profit when the sharing behavior is at a moderate level. Thus, managers should acknowledge for the sharing level when making decisions of technology investment.

Another debate is on the effect of sharing level on the firm performance and market share (Galbreth et al. 2012; Novos and Waldman 1984; Tunca and Wu 2013). We contribute to this debate and raise two research questions in the context of an incumbent form and an entrant

firm: 1) Does the higher sharing level at the entrant firm increase the total sharing market size? 2) Does the higher sharing level at the entrant firm reduce its profit?

One may think that the high sharing level at the entrant will lead to higher sharing market size as previous research claims that the network effect generated from piracy will increase market size (Chellappa and Shivendu 2005; Cheng et al. 2011). However, interestingly and counter-intuitively, our results show that when the incumbent sharing level is low, the increase in entrant's sharing level actually reduces the whole sharing market size. This result indicates that the managers need to be cautious in choosing strategies to determine the sharing and non-sharing market size.

The fact that Netflix lost $2.3 billion annually due to account sharing behavior (Swaminathan 2019) may lead us to believe that higher sharing level results in lower profit. However, surprisingly, our results show that the higher sharing level at the entrant actually improves its profit when the technology level of incumbent is high. Hence, the managers should bear in mind the mechanism with which the profit changes as the sharing level and technology investment change.

LITERATURE REVIEW

In this section, we briefly discuss two streams of literature: information goods sharing literature, and the streaming service literature. Here, we also highlight our contributions with respect to these studies.

Sharing Information Goods

Information goods sharing is commonly discussed in the privacy context. Two forms of information goods sharing exist, specifically the content sharing, in which the content (e.g. video, audio, etc) is sent directly to the receiver, and the account sharing, in which the password is shared to the receiver and they get access to the content as they want. Previous research mainly focused on the content sharing behavior. One stream of literature studied the effects of content sharing behavior on content provider's performance. For example, Tunca and Wu (2013) showed that information goods producers would be better off to provide copyright protection policy for individual piracy behavior rather than commercial piracy behavior. Another stream of literature focused on the piracy's effect on the social influence.

The above studies mainly focus on content sharing. However, in contrast, we focus on account sharing. There are mainly two features that distinguish the content sharing behavior and the account sharing behavior. For the direct content sharing, the sharer do not have privacy concerns of leaking their sensitive information to people who receive the sharing content (Cybersecurity 2018). However, users sharing their accounts could expose their sensitive information, for instance, billing address, preferences, etc, to others. Furthermore, account sharing behavior can be largely detected by technology and the providers can keep the situation under control (Vincent 2019). However, the sharing of the content cannot be fully captured by the content provider as the spread on the Internet is hard to trace and it takes time and effort for the firm to figure out the video or content are spread out from which account and further enforce punishment (Tunca and Wu 2013).

Some of the past studies have also examined the account sharing behavior. For example, Galbreth et al. (2012) study the social sharing of information goods among acquaintance, which also includes the subscription account sharing. However, they do not distinguish between

content sharing and account sharing behaviors. Therefore, they do not capture the unique features of privacy concerns and technology detection in account sharing behavior. We will further distinguish our paper with theirs in the subsequent literature review.

Streaming Services

Previous research studying the streaming service mainly focus on the piracy perspective. For instance, Aguiar et al. (2018) study the effect of shutdown of kino.to, a major platform for unlicensed video streaming, on the unlicensed as well as licensed consumption and find that the shutdown does not leads to less unlicensed consumption as the users turn to the alternative websites. In addition, Lu et al. (2020) study the mediation effect of electronic word of mouth on how the piracy influence the demand of sales and find that positive word of mouth (WOM) volume affects post release piracy but WOM valence has no significant mediation effect on the piracy impact. Different from previous research, which focus on the negative influence from piracy issue, this research attempts to explore both the benefits and damages generated from account sharing behavior.

Most Relevant Literature

One of the most relevant work is by Galbreth et al. (2012), which studies the impact of social sharing of information goods on the pricing and profits of the producer of the good and finds that if sharing is already widespread, additional shares will increase firm's profit. Our paper is different from theirs from four perspectives. First, they study the impact of social sharing and emphasize on the structure of user network and the monopoly setting. However, our research analyzes how the account sharing shapes the supplier market, specifically how competitor's strategies to detect sharing and their users' sharing behavior influence the focal firm's price and profit. We consider four different mixed markets in our paper.

Second, they discuss in the context of all the information goods, but our research focuses on the streaming services, which have a severe profit damages due to the account sharing behavior (Statista 2019). Also, different from other subscriptions, streaming services have unclear attitude towards the account sharing behavior (Houseman 2020). For example, Netflix encourages the users to share their account to their families and friends, but Hulu enables video watching on at most two screens per account. Thus, the profit and the optimal price for the streaming services become unclear.

Third, given the limited user information obtained by video streaming firms, it becomes hard for them to detect the structure of sharing network as centralized or decentralized, which is mainly discussed in Galbreth et al. (2012) paper. In contrast, our research does not distinguish the network but only considers the level of sharing, which can be easily measured by firms through the usage of different devices in different areas. Fourth, the unique features of account sharing behavior: privacy concern and technology investment, are not considered in Galbreth et al. (2012) paper. In our research, privacy concern is taken into account in the user decision, while technology investment is taken into account in the supplier side.

MODEL AND ASSUMPTIONS

In this section, we first discuss from the user problem, and then the platform problem.

Users

A recent survey shows that 12% of Netflix users share account to enjoy the service, and the attitudes towards account sharing behavior varies across different firms (Housman 2020). The ambiguous attitudes makes part of the users share accounts with others, while others still pay the service themselves (Heinzman 2019a). Thus, two different types of users are discussed in this research, specifically users who share account, and those do not share their accounts.
If the users share their account to their family or friends, the price each individual pay is divided, and therefore the cost of each user is $\frac{p}{s+1}$, where s indicates the sharing level. Here, we have $(s+1)$ in the denominator to ensure that the denominator is always non-zero. Furthermore, as discussed earlier, sharing account causes privacy concerns to the user. For example, Amazon prime video is linked with Amazon shopping cart and order history. The account sharing also exposes all the information (e.g., billing address, shopping preference) of the account holder to their acquaintance (Heinzman 2019b). Therefore, the cost of privacy concerns is considered to be $d * s$, which grows with the sharing level s. Here, d indicates the level of privacy concern subjects to $0 < d < 1$. We allow d to be different across users. Consistent to past studies (Aggarwal 2008), we consider d to be uniformly distributed on the standard Hoteling line between 0 and 1, where a user's index $d \in [0,1]$ indicates her cost of privacy concern per unit service level.

In addition, when the users share their account, they face the uncertainty of getting blocked by the firms (and therefore they may not be able to watch videos whenever they want). For example, Hulu only allows to watch different content on at most two screens. This could threaten the users that they will have risk of seeing nothing when they want to watch the videos. Thus, we use $a * k$ to capture the uncertainty cost of sharing, where a is the uncertainty that the account could be blocked by firms (or they could not watch the video as they want) and k indicates the investment of technology used to detect account sharing behavior. Put together, the ultimate utility function of users who share account would be

$$v^s = u - \frac{p}{s+1} - d * s - a * k \quad (1)$$

where u is the utility the users get from watching videos, and p is the membership fee.

For users who do not share account, they do not need to be concerned about the privacy as well as the blocking issues. Hence, the utility of users who do not share their account would be

$$v^{ns} = u - p \quad (2)$$

There will be no punishment or any uncertain cost to decrease their utility.

Streaming Service Firms

Although streaming service firms lost about $9.1 billion due to account sharing (Weprin 2020), it does not mean that the firms do not gain any revenue from account sharing users. However, they gain less revenue per user from those who share than those who do not share. Thus, the revenue of the firms would result from two parts, users who share account and those who do not share, specifically it would be $\frac{p}{s+1} * d_s + p * (1 - d_s)$.

In addition, firms have the option to leverage the technology to detect the password sharers (Weprin 2020). Hence, there will be a cost of the investment of technology used to detect account sharing. We denote the level (or cost) of technology investment by k. Furthermore, as the users use the same password for multiple accounts, the password sharing may lead to identity theft for other accounts beyond streaming services (Cybersecurity 2018). This may cause additional cost to the firm in terms of lawsuits, bad publicity, etc. (Cybersecurity 2018; Davis 2019). This cost would increase with s and decrease with k. Hence, we model it as $\frac{s}{k}$. In summary, the profit of the firms would be

$$\pi = \frac{p}{s+1} * d_s + p * (1 - d_s) - \frac{s}{k} - k \quad (3)$$

Firms and User Decisions

Following Li and Kumar (2018), we propose a three-stage sequential game. In the first stage, the incumbent chooses their price ($\boldsymbol{p_i}$), and in the second stage, the entrant decides whether to enter or not. If they decide to enter, they will decide their price ($\boldsymbol{p_e}$). Finally, in the third stage, the users will choose to use incumbent or entrant's service as well as decide whether to share account or not.

ANALYSIS AND MANAGERIAL INSIGHTS

In this section, we first discuss the monopoly scenario with both sharing and non-sharing market. Next, we discuss the duopoly scenario where the incumbent invests in technology while the entrant does not invest technology.

Monopoly setting

In this subsection, we consider the scenario with only one streaming service firm providing services for both sharing as well as non-sharing users. If the utility of sharing is higher than the utility of not sharing, namely when $v^s = u - \frac{p}{s+1} - d * s - a * k > v^{ns} = u - p$, the user chooses to share; otherwise, she does not share. We represent the privacy cost of indifferent user as d_{ms}. Hence, d_{ms} users would choose to share and $(1 - d_{ms})$ users would not share.
The incumbent's profit maximization problem is as below:

$$\max \pi_m = \frac{p}{s+1} * d_{ms} + p * (1 - d_{ms}) - \frac{s}{k} - k \quad (4)$$
$$s.t. 0 < d_{ms} < 1$$

We summarize the optimal decisions of the streaming services in Lemma 1. Proofs of all the lemmas are in the Appendix. Proofs of all the propositions are omitted. However, they are available upon request.

Lemma 1. Under the monopoly scenario,

a) *The optimal price is* $p_m^* = \frac{(1+s)(1+ak+s)}{2s}$
b) *The optimal profit is* $\pi_m^* = \frac{1}{4}(2 + 2(a-2)k + \frac{(1+ak)^2}{s} + \frac{(k-4)s}{k})$

From Lemma 1, we can infer that higher technology investment will increase the optimal price, but how sharing level influence the price is unclear. Furthermore, the impact of technology investment and sharing level on the profit is also unclear.

Whether technology investment pay off raises a long debate since last century. Zahavi and Lavie (2013) claim that the investment in unique technologies can limit the firm's gains from economies of scope, while Brynjolfsson and Hitt (1996) leverage firm-level data and find that the IS spending has made a significant contribution to firm output. The productivity paradox of technology investment is suggested to be better understood by under which condition the IT investment will pay off (Dehning et al. 2003).

Proposition 1. *Increase in technology investment actually improves the monopoly profit when the sharing level is low (i.e.,* $0 \le s \le \underline{s_0}$*), but, interestingly, it hurts the monopoly profit when the sharing level is in a higher range (i.e.,* $\underline{s_0} < s \le \bar{s_0}$*,). Here,* $\underline{s_0} = \frac{1}{4}(2k^2 - ak^2 + k\sqrt{-8a - 8a^2k + 4k^2 - 4ak^2 + a^2k^2})$ *and* $\bar{s_0} = \frac{1}{4}(2k^2 - ak^2 - k\sqrt{-8a - 8a^2k + 4k^2 - 4ak^2 + a^2k^2})$*.*

Report shows that Disney, which invests in the technology to detect the password sharing and block accounts with excessive sharing, has the highest profit among studio industry in 2019 (Szalai 2020). Hence, one may think that the high technology investment will lead to higher profit. However, the above proposition shows that this is not the case. In fact, we find that the technology investment helps more when the sharing level is below $\underline{s_0}$ rather than when $\underline{s_0} < s \le \bar{s_0}$. For completion, we would also like to add that the technology level will help when $s > \bar{s_0}$. This result can be explained as follows.

When the sharing level is high, the increase of technology investment will increase the potential cost of sharing account, users would like to choose not share their account and the sharing market size will decrease ($d_{ms}^* = \frac{1-ak}{2s} + \frac{1}{2}$), at the same time the non-sharing market size ($1 - d_{ms}^* = \frac{1}{2} - \frac{1-ak}{2s}$) will increase. The streaming service firm will gain more revenue from the non-sharing market. With low sharing level, more technology investment will increase the price ($p_m^* = \frac{(1+s)(1+ak+s)}{2s}$) so that the firm could counter the cost of technology investment. Thus the firms could get increasing revenue from the market with increasing price. However, when the sharing level is at the moderate level, the increasement of technology can neither increase the market size or the price, thus the market size as well as the price will as well in a moderate level and cannot achieve sharing market's economies of scale or monopoly price. Hence, with moderate sharing level, the technology investment will decrease the profit.

Our findings provide managerial insights for the practitioners to help them in optimizing the technology investment. It also gives the message the one-size-fit policy does not work. They need to carefully account for the sharing level. Either with high or low sharing level, the technology investment will pay off, however, with the modest sharing level, the streaming services should think carefully to invest in technology.

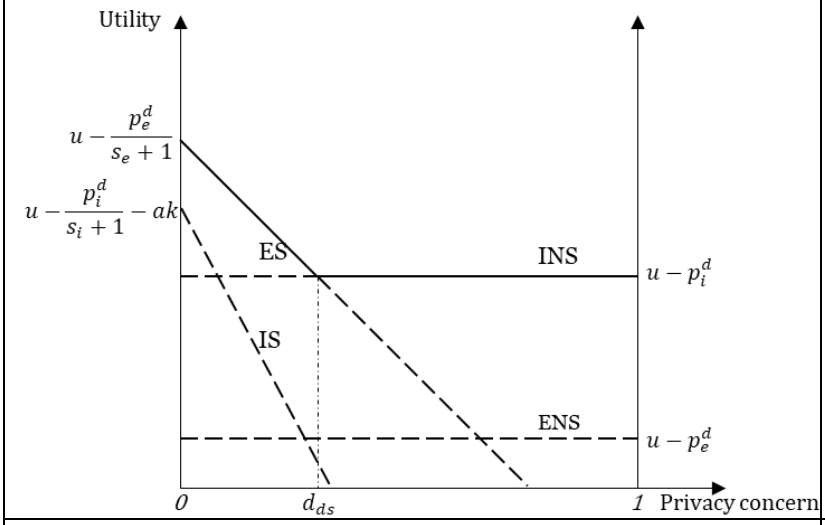
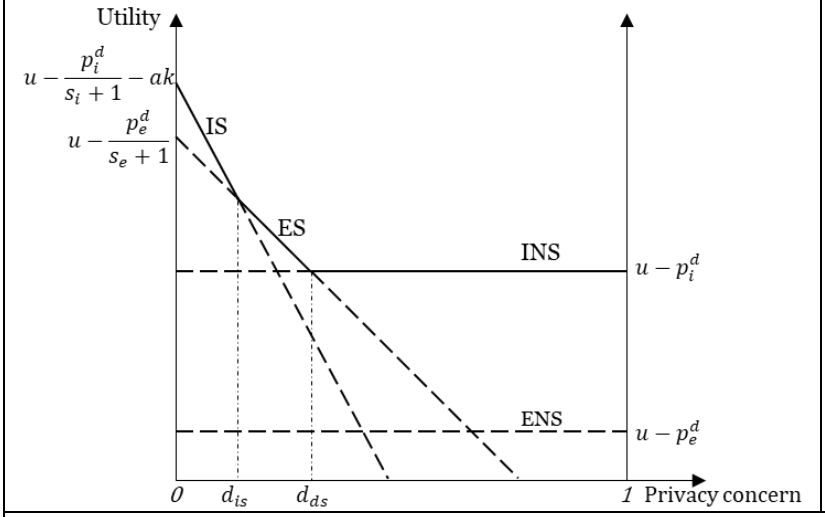
Duopoly setting

We assume that the sharing level in the incumbent will be larger than the entrant ($s_i > s_e$) given the observation that the existing behavior is most severe in the incumbent Netflix (Swaminathan, 2019), which launched in 1998, while Amazon prime video and Hulu launching in 2005 and 2008 respectively, suffer a less account sharing problem. In addition, as the

monthly fee for Netflix is $8.99, while Amazon Prime Video is $12.99, we assume that the incumbent price is less than entrant price ($p_i < p_e$).

We further discuss the markets with different relationship between the intercept of v_i^s and v_e^s. The plot of two markets and constraints are shown in Table 1. Lines IS and ES indicate customer utility of sharing account on the incumbent and entrant platform, respectively, and lines INS and ENS indicate customer utility of not sharing account on the incumbent and entrant platform, respectively. In the first market, the entrant takes the sharing market and the incumbent takes the non-sharing market, and in the second market, both the incumbent and entrant take the sharing market and only the incumbent takes the non-sharing market. We use backward induction to solve the problem.

Table 1. Summary of the hoteling models in duopoly market

Figures of user utility in the duopoly market	Constraints
Market 1 	$p_i < p_e$ $s_i > s_e$ $u - \frac{p_i^d}{s_i+1} - ak < u - \frac{p_e^d}{s_e+1}$
Market 2 	$p_i < p_e$ $s_i > s_e$ $u - \frac{p_i^d}{s_i+1} - ak > u - \frac{p_e^d}{s_e+1}$
Notes: Line IS indicates utility of user sharing account on incumbent platform Line ES indicates utility of user sharing account on entrant platform	

Line INS indicates utility of user not sharing account on incumbent platform
Line ENS indicates utility of user not sharing account on entrant platform

Does Higher Technology Investment Helps the Firm More?

In the first duopoly market, we assume that the intercept of v_i^s is smaller than the one of v_e^s ($u - \frac{p_i}{s_i+1} - a * k < u - \frac{p_e}{s_e+1}$).

When the incumbent's price is lower than the entrant's price ($p_i < p_e$), the users who do not share account always choose incumbent as $v_i^{ns} = u - p_i > v_e^{ns} = u - p_e$. As the incumbent's sharing level is higher than the entrant's sharing level and the intercept of v_i^s is smaller than the one of v_e^s, the users who share account always choose entrant as $v_i^s = u - \frac{p_i}{s_i+1} - a * k - s_i * d < v_e^s = u - \frac{p_e}{s_e+1} - s_e * d$. When the user's utility satisfies $v_e^s > v_i^{ns}$, she would share on the entrant, otherwise, she will not share and use incumbent's service. We represent the indifferent customers between sharing and not sharing as d_{ds}. Thus, d_{ds} users will share on the entrant, while $(1 - d_{ds})$ users will use incumbent service and not share.

Thus, the maximization problem of the incumbent and entrant

$$\max \pi_i = p_i * (1 - d_{ds}) - \frac{s_i}{k} - k \quad (5)$$

$$\max \pi_e = \frac{p_e}{s_e+1} * d_{ds} \quad (6)$$

$$\text{s.t. } 0 < d_{is} < d_{ds} < 1;\ s_i > s_e;\ p_1 < p_2$$

Lemma 2. *Under the duopoly scenario that the incumbent takes the non-sharing market and entrant takes the sharing market, the optimal price and profit is as below*

a) *The optimal price is $p_i^{d^*} = s_e$ for the incumbent and $p_e^{d^*} = \frac{1}{2}s_e(1 + s_e)$ for the entrant*
b) *The optimal profit is $\pi_i^{d^*} = \frac{s_e}{2} - \frac{s_i}{k} - k$ for the incumbent, and $\pi_e^{d^*} = \frac{s_e}{4}$ for the entrant.*

Proposition 2. Contrary to the monopoly market, increase in technology investment actually improves the incumbent's profit when the sharing level is high (i.e., $s_i \geq \underline{s_d}$), but it hurts the incumbent's profit when the sharing level is in the low range (i.e., $0 < s_i \leq \underline{s_d}$). Here, $\underline{s_d} = k^2$.

Contrary to the monopoly market, the increase in technology investment hurts the incumbent's profit when the sharing level is low in duopoly market.

The sharing behavior has been detected and forbidden by incumbent. The incumbent benefits only when the incumbent's sharing level is high in the duopoly market, in which the bad publicity ($\frac{s_1}{k}$) is much more severe than in the one with low incumbent's sharing level. Hence, the decrease in the bad publicity pays off the technology investment cost when the incumbent's sharing level is high.

This proposition implies that technology investment is a tricky decision. Managers need to carefully figure out their market position and further decide the technology investment. Furthermore, it suggests the managers to adjust the technology investment quickly when the market changes.

In the second duopoly market, we assume that the intercept of v_i^s is larger than the one of v_e^s ($u - \frac{p_i}{s_i+1} - a * k > u - \frac{p_e}{s_e+1}$).

The users who do not share account always choose incumbent as $v_i^{ns} = u - p_i > v_e^{ns} = u - p_e$. When the user's utility satisfies $\min\{v_i^s, v_e^s\} > v_i^{ns}$, she would either share on the entrant or on the incumbent, otherwise, she will not share. We represent the indifferent customers between sharing and not sharing as d_{ds}. Hence, d_{ds} users will share either on incumbent or entrant, and $(1 - d_{ds})$ will not share. In addition, when $v_i^s = u - \frac{p_i}{s_i+1} - d * s_i - a * k > v_e^s = u - \frac{p_e}{s_e+1} - d * s_e > v_i^{ns}$, the user chooses to share on incumbent, otherwise ($v_e^s > v_i^s > v_i^{ns}$), she will share on entrant. We use d_{is} to represent indifferent users sharing between incubment and entrant. Thus, among d_{ds} sharing user, d_{is} users will share on the incubment, while $(d_{ds} - d_{is})$ will share on the entrant, and for the other $(1 - d_{ds})$ users, they will use incumbent service and not share. Thus, the maximization problem of the incumbent and entrant is as below

$$\max \pi_i = \frac{p_i}{s_i+1} * d_{is} + p_i * (1 - d_{ds}) - \frac{s_i}{k} - k \quad (5)$$

$$\max \pi_e = \frac{p_e}{s_e+1} * (d_{ds} - d_{is}) \quad (6)$$

$$\text{s.t. } 0 < d_{is} < d_{ds} < 1;\ s_i > s_e;\ p_i < p_e.$$

Lemma 3. *Under the duopoly scenario that the incumbent takes both the sharing market and non-sharing market, and entrant takes only the sharing market, the optimal price and profit is as below*

a) *The optimal price is* $p_i^d = \frac{(1+s_i)(ak(s_i-s_2-1)+2(1+s_i)(s_i-s_e))s_2}{2s_i(1+s_i(2+s_i)-s_e(2+s_e))}$ *for the incumbent and* $p_e^d = \frac{s_e(1+s_e)}{4s_i}(2ak + \frac{(ak(s_i-s_e-1)+2(1+s_i)(s_i-s_e))(1+s_i-s_e)}{1+s_i(2+s_i)-s_e(2+s_e)})$ *d for the entrant*

b) *The optimal profit is* $\pi_i^d = \frac{4k(1+s_i)^2(s_i-s_e)^2s_e+a^2k^3s_e(1-s_i+s_e)^2-4k^2(s_i-s_e)(2s_i(1+s_i)^2+as_e-s_i(4+as_i)s_e+(a+(-2+a)s_i)s_e^2)-8s_i^2(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))}{8ks_i(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))}$ *for the incumbent, and* $\pi_e^d = \frac{s_e(2(1+s_i)(s_i-s_e)(1+s_i-s_e)+ak(1+3s_i^2-2s_i(s_e-2)-s_e(4+s_e)))^2}{16s_i(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))^2}$ *for the entrant.*

Does Higher Entrant's Sharing Level Enlarge the Whole Sharing Market Size?

The debate on the effect of sharing level on the performance of market and firms lasts for years. Galbreth et al. (2012) claim that for centralized sharing network, high level of sharing increase firm profit. While Novos and Waldman (1984) find that piracy causes significant profit losses for legal producers and Tunca and Wu (2013) find that the individual sharing can hurt the profit of firms. We contribute to this debate, and discuss under which scenarios, the sharing level will benefit or hurt the performance of market and firm.

Proposition 3. *Increase in entrant's sharing level enlarge the whole sharing market when the incumbent's sharing level is high (i.e.,* $s_i > \overline{s_i}$ *), but, interestingly, it will hurt the whole sharing market when the incumbent's sharing level is in a low range (i.e.,* $\underline{s_i} \le s_i \le \overline{s_i}$*). Here,* $\underline{s_i} = s_e - \sqrt{1 + 2s_e + 2ak(1 + s_e)}$ *and* $\overline{s_i} = s_e + \sqrt{1 + 2s_e + 2ak(1 + s_e)}$.

A stream of literature claims that the network effect generated from piracy will increase market size (Chellappa and Shivendu 2005; Cheng et al. 2011). Hence, one may think that the high sharing level will lead to higher sharing market size. However, the above result shows that the increase in entrant's sharing level indeed enlarge the whole sharing market when the incumbent's sharing level is higher than $\overline{s_i}$. While, surprisingly, increasing sharing level of entrant will shrink the sharing market size when $\underline{s_i} \leq s_i \leq \overline{s_i}$. This result can be explained as follows.

As the increase in incumbent sharing level will result in increasement of both incumbent (p_i^d) and entrant's price (p_i^d), the high incumbent sharing level will lead to high price of both incumbent as well as entrant. High price stimulates users to share their account as the decrease in the price ($\frac{p^d}{s+1}$) will compliment the privacy concerns ($d_s * s$) of users locating around the boundary between sharing and non-sharing market. At the same time, the increase in entrant sharing level will also lead to increasement of incumbent price. Hence, with high incumbent and entrant price, further increasement in entrant will lead to more sharing in the market. However, when the incumbent sharing level (s_i) is at the low range, the price (p^d) is originated at a relatively low range, although the raise in the entrant sharing level will result in increase in the incumbent's price (p_i^d), but the sharing price increasement ($\frac{p^d}{s+1}$) is relatively small due to the low price (p^d) and increasing entrant sharing level (s_e), thus it cannot make up the loss entrant sharing level causes to privacy concerns ($d_s * s$).

This proposition provides meaningful implications to both incumbent and entrant to optimize the whole sharing market size. Incumbent should be aware that their sharing level shift from a high level to a low one could result in the whole sharing market size shrink if the entrant further increase their sharing level, and evaluate how such shrink impacts the profit.

Does Higher Entrant's Sharing Level Increase its Profit?

Proposition 4. *Increase in entrant's sharing level will lead to lower entrant's profit when the incumbent's technology level is low (i.e., $k \leq k_0$), but, interestingly, it helps the entrant's profit when the technology level is in a higher range (i.e., $k > k_0$). Here, $k_0 = \frac{-((2(1+s_i)(s_i^2+3s_i^3+3s_i^4+s_i^5-3s_is_e-10s_i^2s_e-13s_i^3s_e-6s_i^4s_e+2s_e^2+11s_is_e^2+23s_i^2s_e^2+12s_i^3s_e^2-4s_e^3-19s_is_e^3-10s_i^2s_e^3+6s_e^4+3s_is_e^4))}{a(s_i+6s_i^2+12s_i^3+10s_i^4+3s_i^5-10s_is_e-22s_i^2s_e-18s_i^3s_e-6s_i^4s_e+4s_e^2+10s_is_e^2+10s_i^2s_e^2+8s_i^3s_e^2-6s_is_e^3-10s_i^2s_e^3+4s_e^4+5s_is_e^4))}$*

According to a report, the Netflix lost $2.3 billion annually due to account sharing behavior (Swaminathan 2019). Hence, one may think that the high sharing level will lead to lower profit. However, proposition 4 shows that the entrant's sharing level hurts its profit when the technology level of incumbent is higher than k_0, which is consistent with the report. However, surprisingly, the sharing level of entrant will boost entrant's profit when $k \leq k_0$. This result can be explained as follows.

The market size of entrant is $d_{ds} - d_{is} = \frac{ak}{4s_i-4s_e} + \frac{(1+s_i)(1+ak+s_i-s_e)}{2+2s_i(2+s_i)-2s_e(2+s_e)}$. When the incumbent's technology investment is high, the market size of entrant will be high because when the technology investment is high, the market size of incumbent will decrease, and users sharing password will transfer into entrant's sharing market. At the same time, the price of entrant (p_e^d) will as well be high with high incumbent technology investment. The increase in entrant's sharing level will attract more users from incumbent market who are scared by the technology. Thus, with high technology level, the increase in sharing level will increase the entrant's profit.

However, if the technology level is low, the market size of incumbent will be larger as the incumbent's sharing level is higher than the entrant. The benefit resulting from market size due to high technology investment diminishes, thus, further increasing sharing level will result in lower profit for entrant.

This proposition suggests that managers need to acknowledge the change in the profit as the sharing level and technology investment changes. Specifically, entrant would be better to evaluate their competitor's technology investment and make decisions to adjust their profit strategies.

Conclusion

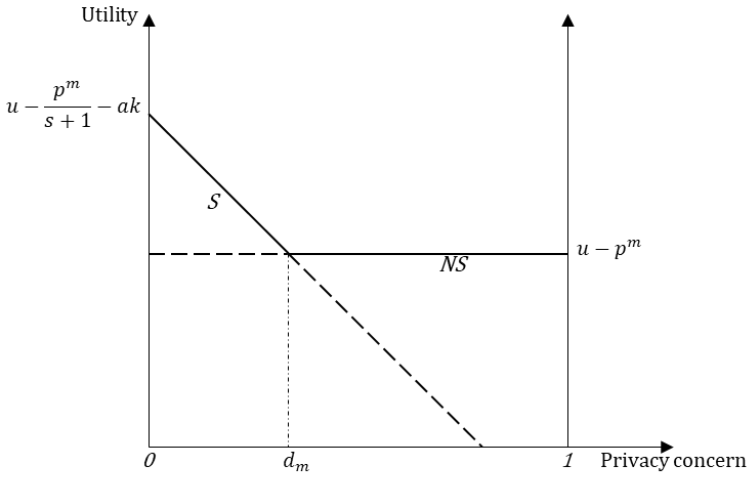
This research has examined how the sharing level shape the market size and firm profit in the duopoly scenarios, as well as how the technology investment affects the monopoly profit. We find some interesting results that the impact of technology investment on the monopoly profit will be enlarged with either high or low sharing level, however, the technology investment hurts the incumbent profit with medium sharing level. Furthermore, the sharing level of entrant exerts heterogenous impact on the sharing market size as well as entrant profit regarding incumbent's sharing level and technology investment. Our results provide insightful implications to practitioners. We plan to present more results regarding duopoly scenarios during the presentation.

APPENDIX

Proof of Lemma 1

When there is only one firm in the market, the users choose between sharing account or not. In Figure A1, lines S and NS indicate user utility of sharing account and not sharing account. We use $d_{ms} = \frac{p^m}{1+s} - \frac{ak}{s}$ to indicate the indifferent users choosing between sharing and not sharing account. Then, we substitute this value into the monopoly profit function, which is $\frac{p}{s+1} * d_m + p * (1 - d_m) - \frac{s}{k} - k$.

Figure A1. User utility in the monopoly market



Line S: Utility of user sharing account
Line NS: Utility of user not sharing account

By applying the first-order condition $\frac{\partial \pi^m}{\partial p^m} = 0$, the optimal price can be obtained as

$$p^m = \frac{(1+s)(1+ak+s)}{2s} \quad \text{(A1)}$$

Given $\frac{\partial^2 \pi^m}{\partial p^{m2}} = -\frac{2s}{(1+s)^2} < 0$, second-order condition is satisfied. We acquire the optional profit by substituting the optimal price in profit function:

$$\pi^m = \frac{1}{4}(2 + 2(a-2)k + \frac{(1+ak)^2}{s} + \frac{(k-4)s}{k}) \quad \text{(A2)}$$

Proof of Proposition 1

From Equation (A2), we have

$$\frac{\partial \pi_m}{\partial k} = \frac{1}{2}(a + \frac{a(1+ak)}{s} + \frac{2s}{k^2} - 2) \quad \text{(A3)}$$

Solving $\frac{\partial \pi_m}{\partial k} = 0$ generates two solutions as below.

$$\underline{s_0} = \frac{1}{4}(2k^2 - ak^2 - k\sqrt{-8a - 8a^2k + 4k^2 - 4ak^2 + a^2k^2}) \quad \text{(A4)}$$

$$\bar{s_0} = \frac{1}{4}\left(2k^2 - ak^2 + k\sqrt{-8a - 8a^2k + 4k^2 - 4ak^2 + a^2k^2}\right) \quad \text{(A5)}$$

The two solutions are both positive when $p^m > 0$ and $\pi^m > 0$.
When $\underline{s_0} < s_1 < \bar{s_0}$, $\frac{\partial \pi_m}{\partial k} > 0$ and when $0 < s_1 < \underline{s_0}$ or $s_1 > \bar{s_0}$, $\frac{\partial \pi_m}{\partial k} < 0$.

Proof of Lemma 2

In the Market 1, we use $d_{ds} = \frac{p_i - p_e + p_i s_e}{s_e + s_e^2}$ to indicate the indifferent users choosing between sharing and not sharing account. We substitute this value into the duopoly profit function for incumbent and entrant platform, which is $p_i * (1 - d_{ds}) - \frac{s_i}{k} - k$ and $\frac{p_e}{s_e+1} * d_{ds}$, respectively.

Using backwards deduction, we apply the first-order condition for entrant $\frac{\partial \pi_e^d}{\partial p_e^d} = 0$, and obtain the optimal entrant price as

$$p_e^d = \frac{1}{2}p_i(1 + s_e) \quad \text{(A6)}$$

Substituting this value into profit function of incumbent and taking the first-order condition of incumbent $\frac{\partial \pi_i^d}{\partial p_i^d} = 0$, we obtain the optimal price of incumbent as

$$p_i^d = s_e \quad \text{(A7)}$$

Substituting this value into p_e^d, we can acquire the optimal price of entrant:

$$p_e^d = \frac{1}{2}s_e(1 + s_e) \quad \text{(A8)}$$

Given $s_i > s_e$, the second order condition $\frac{\partial^2 \pi_i^d}{\partial {p_i^d}^2} = -\frac{1}{s_e} < 0$ and $\frac{\partial^2 \pi_e^d}{\partial {p_e^d}^2} = -\frac{2}{s_e(1+s_e)^2} < 0$ are satisfied.

By substituting the optimal price of both incumbent and entrant into the profit function, we can acquire the profit function of incumbent and entrant are shown as below

$$\pi_i^d = \frac{s_e}{2} - \frac{s_i}{k} - k \quad \text{(A9)}$$

$$\pi_e^d = \frac{s_e}{4} \quad \text{(A10)}$$

We check the condition for Market 2 $s_i > s_e > 0, p_e^d > p_i^d > 0, 0 < d_{ds} < 1$ and $u - \frac{p_i^d}{s_i+1} - ak < u - \frac{p_e^d}{s_e+1}$.

Proof of Proposition 2

From Equation (A10), we have

$$\frac{\partial \pi_e}{\partial k} = \frac{s_i}{k^2} - 1 \quad \text{(A11)}$$

When $s_i < k^2$, $\frac{s_i}{k^2} < 1$. Hence, $\frac{\partial \pi_1}{\partial k} < 1 - 1 = 0$. While $s_i > k^2$, $\frac{s_i}{k^2} > 1$. Hence, $\frac{\partial \pi_1}{\partial k} > 1 - 1 = 0$

Proof of Lemma 3

In the Market 2, we use $d_{ds} = \frac{p_i - p_e + p_i s_e}{s_e + s_e^2}$ to indicate the indifferent users choosing between sharing and not sharing account and $d_{is} = \frac{ak + \frac{p_i}{1+s_i} - \frac{p_e}{1+s_e}}{s_e - s_i}$ to indicate the indifferent users choosing between sharing on the incumbent and entrant platform. We substitute these two values into the duopoly profit function for incumbent and entrant platform, which is $\frac{p_i}{s_i+1} * d_{is} + p_i * (1 - d_{ds}) - \frac{s_i}{k} - k$ and $\frac{p_e}{s_e+1} * (d_{ds} - d_{is})$, respectively. Using backward induction, we apply the first-order condition for entrant $\frac{d\pi_e^d}{dp_e^d} = 0$, and obtain the optimal entrant price as

$$p_e^d = \frac{(1+s_e)(p_i s_i(1+s_i - s_e) + ak(1+s_i)s_e)}{2s_i(1+s_i)} \quad \text{(A12)}$$

Substituting this value into profit function of incumbent and taking the first-order condition of incumbent $\frac{\partial \pi_i^d}{\partial p_i^d} = 0$, we obtain the optimal price of incumbent as

$$p_i^d = \frac{(1+s_i)\big(ak(s_i - s_e - 1) + 2(1+s_i)(s_i - s_e)\big)s_e}{2s_i(1+s_i(2+s_i) - s2(2+s_e))} \quad \text{(A13)}$$

Substituting this value into p_e^d, we can acquire the optimal price of entrant

$$p_e^d = \frac{s_e(1+s_e)}{4s_i}\left(2ak + \frac{(ak(-1+s_i - s_e) + 2(1+s_i)(s_i - s_e))(1+s_i - s_e)}{1+s_i(2+s_i) - s_e(2+s_e)}\right) \quad \text{(A14)}$$

We check the condition $s_i > s_e, p_i^d > p_e^d > 0, 0 < d_{is} < d_{ds} < 1$ and $u - \frac{p_i^d}{s_i+1} - ak > u - \frac{p_e^d}{s_e+1}$.

Given $s_i > s_e$, the second order condition $\frac{\partial^2 \pi_i^d}{\partial {p_i^d}^2} = -\frac{1}{s_e} + \frac{\frac{1}{s_e - s_i} - s_i}{(1+s_i)^2} < 0$ and $\frac{\partial^2 \pi_e^d}{\partial {p_e^d}^2} = \frac{2s_i}{s_e(1+s_e)^2(s_e - s_i)} < 0$ are satisfied.

By substituting the optimal price of both incumbent and entrant into the profit function, we can acquire the profit function of incumbent and entrant as below:

$$\pi_i^d = \frac{4k(1+s_i)^2(s_i-s_e)^2 s_e + a^2k^3 s_e(1-s_i+s_e)^2 - 4k^2(s_i-s_e)(2s_i(1+s_i)^2 + as_e - s_i(4+as_i)s_e + (a+(-2+a)s_i)s_e^2) - 8s_i^2(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))}{8ks_i(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))} \quad \text{(A15)}$$

$$\pi_e^d = \frac{s_e(2(1+s_i)(s_i-s_e)(1+s_i-s_e) + ak(1+3s_i^2 - 2s_i(s_e-2) - s_e(4+s_e)))^2}{16s_i(s_i-s_e)(1+s_i(2+s_i)-s_e(2+s_e))^2} \quad \text{(A16)}$$

Proof of Proposition 3

Plugging Equation (A13) and Equation (A14) into $d_{ds} = \frac{p_i - p_e + p_i s_e}{s_e + s_e^2}$, we can get

$$d_{ds} = \frac{2(1+s_i)(s_i-s_e)(1+s_i+s_e) - ak(1+s_i-s_e)(3+s_i+s_e)}{4s_i(1+s_i(2+s_i)-s_e(2+s_e))} \quad \text{(A17)}$$

From Equation (A17), we have

$$\frac{\partial d_{ds}}{\partial s_e} = \frac{(1+s_i)((s_i-s_e)^2 - 2s_e - 2ak(1+s_e) - 1)}{2s_i(1+s_i(2+s_i)-s_e(2+s_e))^2} \quad \text{(A18)}$$

Solving $\frac{\partial d_{ds}}{\partial s_e} = 0$ generates two solutions. $s_i = -1 < 0$ is invalid. Hence, only one solution left:

$$s_i = s_e - \sqrt{1 + 2ak + 2s_e + 2aks_e} \quad \text{(A19)}$$

As $s_i > 0$, $2s_i(1 + s_i(2 + s_i) - s_e(2 + s_e))^2 > 0$ and $1 + s_i > 0$. Hence, when$(s_i - s_e)^2 - 2s_e - 2ak(1 + s_e) - 1 > 0$, $\frac{\partial d_{ds}}{\partial s_e} > 0$, and vice versa.

When$s_i > s_e - \sqrt{1 + 2ak + 2s_e + 2aks_e}$, $(s_i - s_e)^2 - 2s_e - 2ak(1 + s_e) - 1 > 0$. Hence, $\frac{\partial d_{ds}}{\partial s_e} > 0$.

While $0 < s_i < s_e - \sqrt{1 + 2ak + 2s_e + 2aks_e}$, $(s_i - s_e)^2 - 2s_e - 2ak(1 + s_e) - 1 < 0$.

Hence,$\frac{\partial d_{ds}}{\partial s_e} < 0$

Proof of Proposition 4

From Equation (A10), we have

$$\frac{\partial \pi_e}{\partial s_e} = \frac{1}{16s_i(s_i-s_e)^2(1+s_i(2+s_i)-s_e(2+s_e))^3}\Big(2(1+s_i)(s_i-s_e)(s_i(1+s_i)-(2+s_i)s_e)((1+s_i)^2 - 2(1+2s_i)s_e + 3s_e^2) + ak\Big(s_i(1+s_i)^3(1+3s_i) - 2s_i(1+s_i)\big(5+3s_i(2+s_i)\big)s_e +$$

$$2\left(2+s_i\left(5+s_i(5+4s_i)\right)\right)s_e^2-2s_i(3+5s_i)s_e^3+(4+5s_i)s_e^4\Big)\Big)(2(1+s_i)(s_i-s_e)(1+s_i-s_e)+ak\left(1+3{s_i}^2-2s_i(-2+s_e)-s_e(4+s_e)\right)) \quad \text{(A20)}$$

Solving $\frac{\partial \pi_e}{\partial s_e}=0$ generates solution that

$$k_0=\frac{-((2(1+s_i)(s_i^2+3s_i^3+3s_i^4+s_i^5-3s_is_e-10s_i^2s_e-13s_i^3s_e-6s_i^4s_e+2s_e^2+11s_is_e^2+23s_i^2s_e^2+12s_i^3s_e^2-4s_e^3-19s_is_e^3-10s_i^2s_e^3+6s_e^4+3s_is_e^4))}{a(s_i+6s_i^2+12s_i^3+10s_i^4+3s_i^5-10s_is_e-22s_i^2s_e-18s_i^3s_e-6s_i^4s_e+4s_e^2+10s_is_e^2+10s_i^2s_e^2+8s_i^3s_e^2-6s_is_e^3-10s_i^2s_e^3+4s_e^4+5s_is_e^4))} \quad \text{(A21)}$$

For notational convenience, let $\Gamma=16s_i(s_i-s_e)^2(1+s_i(2+s_i)-s_e(2+s_e))^3$, $K=2(1+s_i)(s_i-s_e)(1+s_i-s_e)+ak\left(1+3{s_i}^2-2s_i(-2+s_e)-s_e(4+s_e)\right)$ and $\Pi=2(1+s_i)(s_i-s_e)(s_i(1+s_i)-(2+s_i)s_e)((1+s_i)^2-2(1+2s_i)s_e+3s_e^2)+ak\Big(s_i(1+s_i)^3(1+3s_i)-2s_i(1+s_i)\left(5+3s_i(2+s_i)\right)s_e+2\left(2+s_i\left(5+s_i(5+4s_i)\right)\right)s_e^2-2s_i(3+5s_i)s_e^3+(4+5s_i)s_e^4\Big)$. Hence, $\frac{\partial \pi_e}{\partial s_e}=\frac{K*\Pi}{\Gamma}$.

As mentioned in paper $s_i>s_e$, $\Gamma=16s_i(s_i-s2)^2(1+2(s_i-s_e)+(s_i^2-s_e^2))^3>0$, and $K=2(1+s_i)(s_i-s_e)(1+s_i-s_e)+ak\left(1+\left(s_i^2-s_e^2\right)+2s_i(s_i-s_e)+4(s_i-s_e)\right)>0$. Thus to ensure $\frac{\partial \pi_e}{\partial s_e}>0$, $\Pi>0$ and vice versa.

When $k>\ k_0$, $\Pi>0$, thus $\frac{\partial \pi_e}{\partial s_e}>0$. When $k<k_0$, $\Pi<0$, thus $\frac{\partial \pi_e}{\partial s_e}<0$.

REFERENCE

Aggarwal, C. C. 2008. On Unifying Privacy and Uncertain Data Models, 2008 IEEE 24th International Conference on Data Engineering, pp. 386–395.

Aguiar, L., Claussen, J., and Peukert, C. 2018. “Catch Me If You Can: Effectiveness and Consequences of Online Copyright Enforcement,” *Information Systems Research* (29:3), pp. 656–678.

Brynjolfsson, E., and Hitt, L. 1996. “Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending,” *Management Science* (42:4), pp. 475–627.

Chellappa, R. K., and Shivendu, S. 2005. “Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets,” *Information Systems Research* (16:4), pp. 400–417.

Cheng, H. K., Liu, Y., and Tang, Q. 2011. “The Impact of Network Externalities on the Competition between Open Source and Proprietary Software,” *Journal of Management Information Systems* (27:4), pp. 201–230.

Cybersecurity, U. 2018. “The Netflix Sharing Problem,” *Magid*. (https://magid.com/mentions/the-netflix-sharing-problem/ accessed on April 28, 2020).

Davis, M. 2019. “4 Damaging After-Effects of a Data Breach,” Cybint. (https://www.cybintsolutions.com/4-damaging-after-effects-of-a-data-breach/ accessed on April 28, 2020).

Dehning, B., Richardson, V. J., and Zmud, R. W. 2003. “The Value Relevance of a Nnouncements of Transformational Information,” *MIS Quarterly* (27:4), pp. 637–656.

Despotakis, K. 2019. “Netflix Co-Founder on a Potential Password Sharing Crackdown: ‘You’re Not Scared, Are You?,’” *Yahoo Finance*. (https://finance.yahoo.com/news/password-sharing-crackdown-netflix-hbo-210830952.html accessed on April 28, 2020).

Galbreth, M. R., Ghosh, B., and Shor, M. 2012. “Social Sharing of Information Goods: Implications for Pricing and Profits,” *Marketing Science* (31:4), pp. 603–620.

Heinzman, A. 2019a. “5 Ways to Save Money on Your Netflix Account,” *How To Geek*. (https://www.howtogeek.com/425375/5-ways-to-save-money-on-your-netflix-account/ accessed on April 28, 2020).

Heinzman, A. 2019b. “Do Video Streaming Services Care If You Share Your Account?,” *How To Geek*. (https://www.howtogeek.com/414990/do-video-streaming-services-care-if-you-share-your-account/ accessed on April 28, 2020).

Housman, D. 2020. “Is It OK to Share Your Netflix Account? The Legal Lowdown on Log-in Sharing at Nine Popular Services,” *The Simple Dollar*. (https://www.thesimpledollar.com/save-money/is-it-ok-to-share-your-netflix-account-the-lowdown-on-log-in-sharing-at-nine-popular-services/ accessed on April 28, 2020).

Lewis, S. 2019. “Parks Associates Forecasts $12.5B in Lost Revenue in 2024 Due to Pay-TV and OTT Piracy and Account Sharing,” *CISION*. (https://www.prnewswire.com/news-releases/parks-associates-forecasts-12-5b-in-lost-revenue-in-2024-due-to-pay-tv-and-ott-piracy-and-account-sharing-300885598.html accessed on April 28, 2020).

Li, B., and Kumar, S. 2018. “Should You Kill or Embrace Your Competitor: Cloud Service and Competition Strategy,” *Production and Operations Management* (27:5), pp. 822–838.

Lu, S., Wang, X. (Shane), and Bendle, N. 2020. “Does Piracy Create Online Word of Mouth? An Empirical Analysis in the Movie Industry,” *Management Science* Forthcoming.

Novos, I. E., and Waldman, M. 1984. “The Effects of Increased Copyright Protection,” *Journal of Political Economy*, pp. 236–246.

ParksAssociates. 2019. “Account Sharing and Digital Piracy.” (http://parksassociates.com/marketfocus/account-sharing-video accessed on April 28, 2020).

Statista. 2019. “Video Streaming (SVoD).” (https://www.statista.com/outlook/206/100/video-streaming--svod-/worldwide accessed on April 28, 2020).

Swaminathan, A. 2019. “Password Sharing Costs Netflix Billions — and That’s before the Price Increase,” *Yahoo Finance*. (https://finance.yahoo.com/news/password-sharing-costs-netflix-billions-thats-price-increase-160028395.html accessed on April 28, 2020).

Szalai, G. 2020. “Studio Profit Report 2019: Disney Reigns, Netflix Gains,” *Hollywood Reporter*. (https://www.hollywoodreporter.com/news/studio-profit-report-2019-disney-reigns-netflix-gains-1281114 accessed on April 28, 2020).

Tunca, Tunay I., and Wu, Q. 2013. “Fighting Fire with Fire: Commercial Piracy and the Role of File Sharing on Copyright Protection Policy for Digital Goods,” *Information Systems Research* (24:2), pp. 436–453.

Vincent, J. 2019. “Sharing Passwords for a Video Streaming Site? This Company Will Use AI to Track You Down,” *The Verge*. (https://www.theverge.com/2019/1/8/18174161/netflix-hbo-hulu-shared-password-account-synamedia-machine-learning-ai accessed on April 28, 2020).

Welch, C. 2019. “Disney Will Tolerate Password Sharing on Disney+ — for Now,” *The Verge*. (https://www.theverge.com/2019/11/12/20960332/disney-plus-password-sharing-streaming-accounts accessed on April 28, 2020).

Weprin, A. 2020. “Streaming Services Reckon With Password-Sharing ‘Havoc,’” *Hollywood Reporter*. (https://www.hollywoodreporter.com/news/streaming-services-prepare-password-sharing-havoc-1267728 accessed on April 28, 2020).

Williams, B. 2019. “The Dangers of Password Sharing at Work,” *Techradar*. (https://www.techradar.com/news/the-dangers-of-password-sharing-at-work accessed on April 28, 2020).

Zahavi, T., and Lavie, D. 2013. “Intra-Industry Diversification and Firm Performance,” *Business* (34:March), pp. 978–998.

DECISION SCIENCES INSTITUTE

Related Technological Knowledge in Buyer-Supplier Networks and its Effect on Innovation

Shubhobrata Palit
Georgia Institute of Technology, Scheller College of Business
shubhobrata.palit@scheller.gatech.edu

Manpreet Hora
Georgia Institute of Technology, Scheller College of Business
manpreet.hora@scheller.gatech.edu

Soumen Ghosh
Georgia Institute of Technology, Scheller College of Business
soumen.ghosh@scheller.gatech.edu

ABSTRACT

Supplier networks are an important source of technological knowledge. We analyze the role of available technological knowledge in a firm's supplier network on the former's innovation performance. Using a panel data of buyer-supplier relationships and the patenting activities of the respective firms, we examine how related technological knowledge between a firm and its supplier network, technological breadth of a firm's supplier network, and extent of global sourcing of a firm influence the firm's innovation performance.

KEYWORDS: Innovation, Buyer-supplier networks, Technological knowledge

DECISION SCIENCES INSTITUTE

Does extended producer responsibility improve eco-innovation: an empirical study of product take-back programs

Yuqi Peng
University of South Carolina - Darla Moore School of Business
1014 Greene St, Columbia, SC 29208
Email: Yuqi.Peng@grad.moore.sc.edu

Yan Dong
University of South Carolina - Darla Moore School of Business
1014 Greene St, Columbia, SC 29208
Email: yan.dong@moore.sc.edu

Sriram Venkataraman
University of South Carolina - Darla Moore School of Business
1014 Greene St, Columbia, SC 29208
Email: sriram.venkataraman@moore.sc.edu

Sining Song
University of Tennessee, Knoxville - Haslam College of Business
453 Haslam Business Building, Knoxville, TN 37996
Email: sining.song@utk.edu

ABSTRACT

Extended producer responsibility (EPR) is an important policy instrument to address environmental risks in solid waste disposal. Whether EPR incentivizes a producer's eco-innovation is an ongoing debate among operations researchers and policymakers. Using firm-level take-back program as an important initiative of EPR and eco-patent as a proxy of eco-innovation, we leverage a difference-in-differences approach to identify the causal relationship between EPR and eco-innovation. We find that adopting take-back programs significantly motivates producers to develop eco-innovation. We also find that producers' program scales and industries play significant roles in the take-back effect. Our research provides implications for both producers and policymakers.

KEYWORDS: Extended producer responsibility, Take-back, Eco-innovation, Difference-in differences

DECISION SCIENCES INSTITUTE

Design and Analysis for A Smart Post-Discharge Assistance Applications

Pan Zou
Georgia Institute of Technology
Email: zoupzp123@gatech.edu

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Yiyun (Cindy) Fei
Georgia Institute of Technology
Email: yfei38@gatech.edu

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Shaojie (Emily) He
American Heritage School
Email: pl241111@ahschool.com

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

This paper presents a general system for post-discharge care assistance (PDA) applications to reduce unnecessary cost of home health care services without potential health risk and thus to benefit the recovery of the diseases of the discharged patients. This paper mainly discusses the design and analysis of the proposed system. A pilot application of system development for smart urine tract catheter (UTC) drainage bag proves the proposed general system's effectiveness.

KEYWORDS: Health care, Medicare accessories, Decision-making

INTRODUCTION

Medicare for discharged patients has been the main problem in the healthcare area. A survey conducted in 2004 revealed that though most of the surveyed discharged patients knew the purpose and course of their medication, they did not know about the necessary lifestyle, side

effects of the medicine, and recommended medication schedules, which benefit their recoveries (Kerzman et al., 2004). Elderly discharged patients tend to receive proficient home care services at a relatively high cost. In 2012, around seventy percent of adults with an average age of seventy referred for home health care service after discharge from the hospital to home (Jones et al., 2015). However, there is not enough coordination between the hospital and home health care, and this care coordination process could cause readmissions and medication errors (Jones et al., 2017). Despite the traditional medical care system, the booming technology in smart sensing and information has achieved various mobile application-based practices in medical consultation of certain diseases. There is still a lack of an integrated system can provide automated patient-centered guidance to discharged patients and establish reliable care coordination between hospital and other medical consultants or doctors.

There are three main technical challenges in the design of such a system. The first technical challenge is the management of acquired data through several sensors. The second technical challenge is in the data analysis area since the system needs to determine the post-discharge patients' health condition. The third technical challenge comes from the assistive action achieving. The system needs to provide different services to the user based on different use cases. To overcome the technical challenges, the design of a general system for post-discharge care assistance (PDA) applications is proposed in this paper. The system implements smart sensing, predictive learning with intelligent reasoning, and a mobile platform to provide patient-centered remote medical care services to post-discharge patients.

A pilot application of a smart urine tract catheter drainage bag system is also included in this paper. The pilot application contains three main components: a machine-learning model trained on historical data that can predict the urination frequency of each patient, a real-time flow tracker that collects the real-time urine flow rate of the patients, and a mobile application platform that can manage all the information.

LITERATURE REVIEW

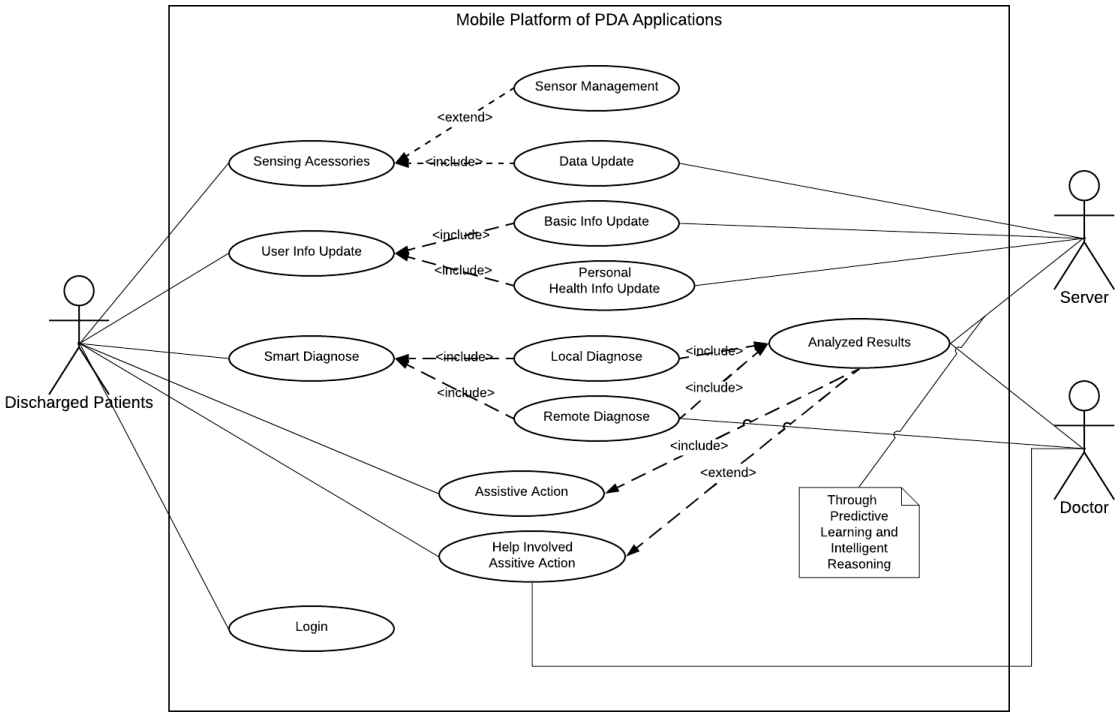
It has been found that social media application on the mobile platform is a viable option for monitoring the condition of discharged patients with head and neck tumor (Lyu et al., 2016). Furthermore, they concluded that the doctor-led follow-up model is able to provide useful health instructions to discharged patients. Several researchers have proposed a short message system to provide health education to discharged patients (Zheng et al., 2011). It can be concluded that the proposed system can solve the nursing problem effectively for discharged patients.

Lin et al. (2019) were able to successfully predict urine output in sepsis patient's eXtreme Gradient Boosting algorithm Decision trees are used successfully to detect patients with a high risk of bladder cancer (Tun et al., 2018). They were able to do so by using data extract from immunofluorescence images and data collected from patients. Some researchers developed a Deep Feature Pyramid Network (DFPN) to detect sediments in the urine (Liang et al., 2018). Their proposed method achieved an accuracy of 87% and can be used to automate the diagnosis of kidney disease. Garapati and other researchers used neural networks, support vector machines, and linear discriminant analysis to detect bladder cancer staging (Garapati et al., 2017). Their results showed that the algorithms could predict the cancer stages with high accuracy regardless of the types of features (texture, morphological, etc.). Some researchers also used random forests and XGBOOST to test for cancer noninvasively. Their proposed methodology achieved 94% accuracy and was statistically better than the current techniques for detecting cancer (Sokolov et al., 2018).

USE CASE ANALYSIS AND SYSTEM DESIGN

The UML case diagram of the proposed system is shown in figure 1. The post-discharge patients are the primary users of the proposed system. The general physical location of employing the system is designed to be home or places that are lack of skilled facilities. The system provides two services to users, which are smart diagnosis and assistive action. When patients use the system initially, post-discharge patients need to input some one-time basic information like sex, weight, and age manually. Users also need to input some specific information like fluid intake, medicine intake, and diet several times a day. Furthermore, personal health information is required for the system. This information could be inputted with assistance from skilled facilities or transferred from medical records in the hospital. Despite some information input from the user, the system requires users to wear smart sensing accessories to collect real-time data like heart rate and body temperature. The real-time data is uploaded to the server for analysis. This data is collected at several specific time intervals to reduce the size of raw data. The server processes the raw data from sensors and extracts features like trend and stability from it. The extracted features from raw data and other input information from the user will be employed as input for the predictive learning model. The output from the predictive learning model needs further processing through intelligent reasoning. The goal of intelligent reasoning is to detect the recurrence of diseases, caused side-effects of the taken medicines and lifestyles of the post-discharge patients.

Figure 1: UML Case Diagram of the System for PDA Applications



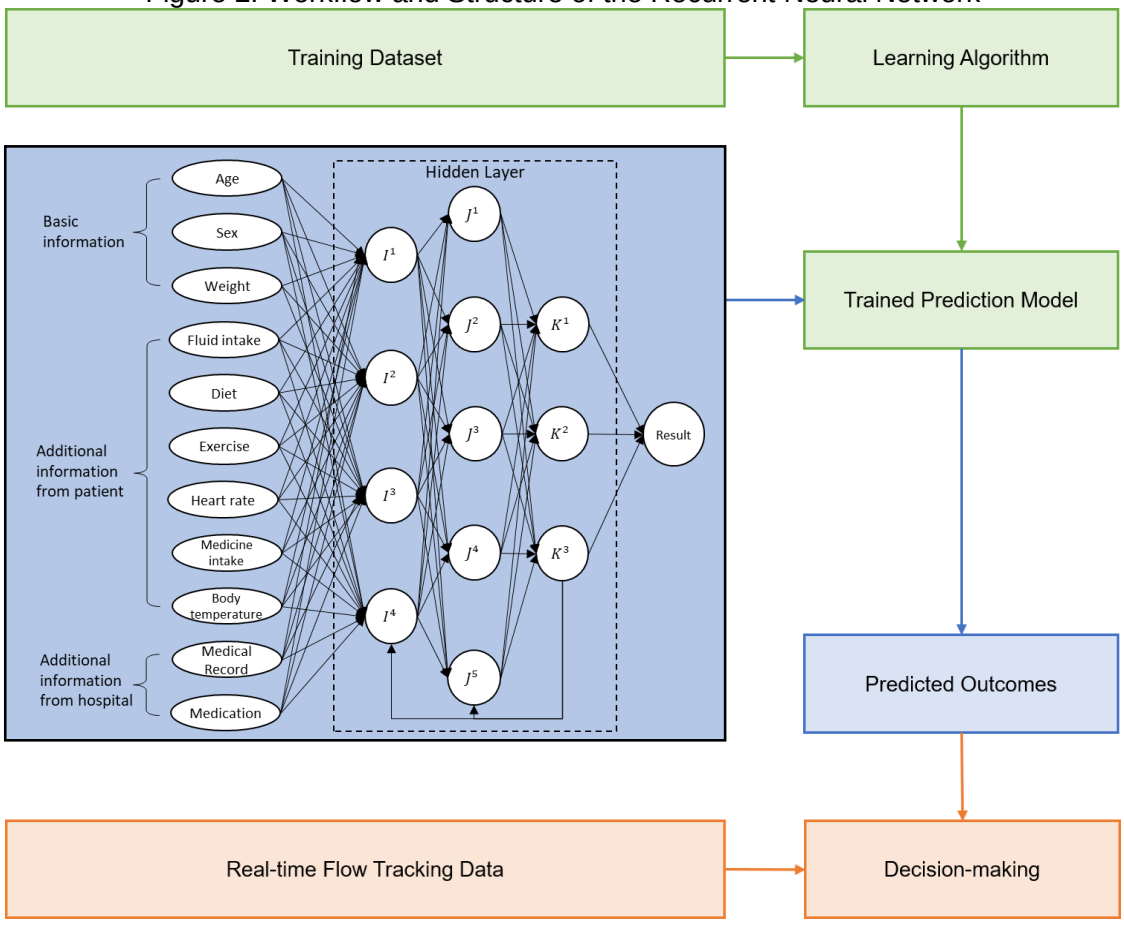
Analyzed results from the server are downloaded to the mobile platform. Those results are necessary for smart diagnoses and assistive actions. The smart diagnoses provide two services to the user, which are local diagnosis and remote diagnosis. The local diagnosis could provide lifestyle guidelines for medicine intake, diet to patients to benefit the recovery or reduce the side effects of intake medicines. The remote diagnosis concentrates on the possible recurrence of the diseases, which requires the involvement of a professional doctor. The analyzed results that

collected data from the sensors and user input information are also sent to the doctor for the purpose of remote diagnosis. Assistance action services provided in the system are more specific and depend on specific diseases. For example, for discharged patients who have diabetes, assistance action provided to patients could be guidelines for the injection of insulin. However, in some complex care situations, the platform can schedule health care services from skilled facilities.

RECURRENT NEURAL NETWORK MODEL FOR PREDICTIVE LEARNING

The prediction model is first trained with the training dataset, including statistical medical record data without identifiable patient information by utilizing a recurrent neural network. When a new patient joins the system, a patient case of a trained prediction model can be initialized and start generating prediction outcomes based on the information of a certain post-discharge patient, as shown in figure 2. The necessary information for initializing the patient case is categorized in type. Basic information requires one-time input which includes age, sex and weight of the out-discharge patient. Additional information from the patient includes the processed data collected by sensors like heart rate and body temperature. It also includes information like fluid intake, diet, exercise and medicine intake which requires the patient to update several times per day. Additional information from the hospital includes medical records and medication. Once inputted into the RNN model, they go through several layers feedforward and returns the prediction outcomes as output results.

Figure 2: Workflow and Structure of the Recurrent Neural Network



The prediction outcomes would also be updated once the real-time data is collected from the post-discharge patients and be preprocessed. It can be used as a reference for the decision-making process. The input information and actual outcomes of this specific case can be added to the training dataset to update, which can improve the accuracy of the prediction model. Thus, the size of the training dataset size can be increased, and the accuracy of the prediction model can be improved as time goes by.

INTELLIGENT REASONING FOR HEALTH CONDITION DETECTION

The health condition detection is processed locally in the mobile platform of DPA applications. After receiving analyzed results from the recurrent network model, a set of rules will be applied based on the possibility of each health condition case from the prediction outcomes. Rough Petri net model (RPN) is employed for applying rules. The general steps include representing the attributes of the rules, construction of input object distribution transitions, the calculation of firing strength of rules and determination of the winning rule. Health conditions could be detected based on generated rules and prediction outcomes. Then, the corresponded assistive actions are generated for post-discharge patients.

PILOT APPLICATION: SMART URINARY TRACT CATHETER DRAINAGE BAG ASSISTIVE AND PREDICTIVE NOTIFICATION SYSTEM

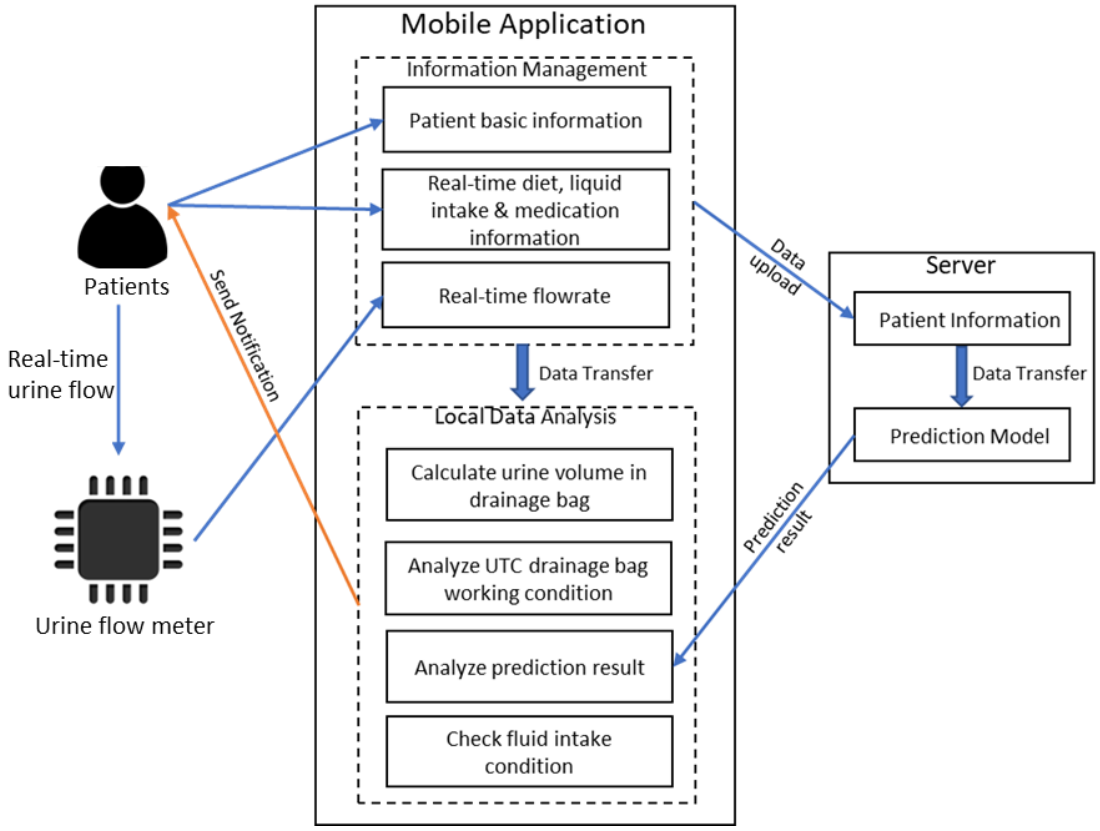
Today, an estimated 15-25% of hospital visits involve patient use of a urinary tract catheter (UTC), whether for a patient receiving anesthesia for surgery or an outpatient in need of one for their daily life (2015, CDC). Masses of post-discharge patients rely on the urinary tract catheter drainage bag in their daily lives. The urinary tract catheter usually connects to drainage bags, which generally are attached either to a rolling stand or to the leg of the patient. Typically, patients in need of a catheter wear a leg bag, as it attaches to the upper thigh or calf and allows for semi-normal daily activities. Completely filling a single drainage bag will exacerbate the development of bacteriuria and increase the possibility of Catheter-associated urinary tract infection and even sepsis. It also may result in the catheter being pulled out, thus requiring reinsertion by a nurse. Other issues include bag leakage, the risk of backflow, and urine tract catheter blockage. The proposed system in the pilot application is designed to solve the two main problems of using a UTC drainage bag. First, without physically inspecting, patients have no information about the working condition of the urine tract catheter and the occupancy percentage of their bag. Second, patients can not notice the development of urinary tract infections before feeling uncomfortable. The proposed system in the pilot application utilizes the mobile application to analyze the results from the prediction model and the urine flowmeter to return the working condition of the drainage bag and detection result of urinary tract infection to the user.

System Analysis and Design

In the case of pilot application, patients and hospitals provide additional information to the system and receive reminders regarding the notification of drainage bag status and the result of infection detection. In figure3, information management in the mobile application will receive three types of input information. Patient basic information like sex, age, weight, and medical record is input from patients and the hospital. This information will be transferred to the server through the mobile application to initialize the prediction model. The real-time diet, liquid intake, and medication information are provided by the patients and nurses. Another type of real-time data is the urine flow rate data collected from the patients by the real-time monitoring device. Two types of real-time data are transferred to the prediction model located on the server. The prediction model

utilizes the RNN model to output the possibility of urinary tract infection cases. The prediction output is analyzed in the mobile application to determine the urinary infection status of post-discharge patients. Urine volume and working condition of the drainage bag is also analyzed in the mobile application locally based on the real-time urine flow. The mobile application can notify the patient about the results to reduce the workload and stress of the patients.

Figure 3: System Architecture of Smart Urinary Tract Catheter System



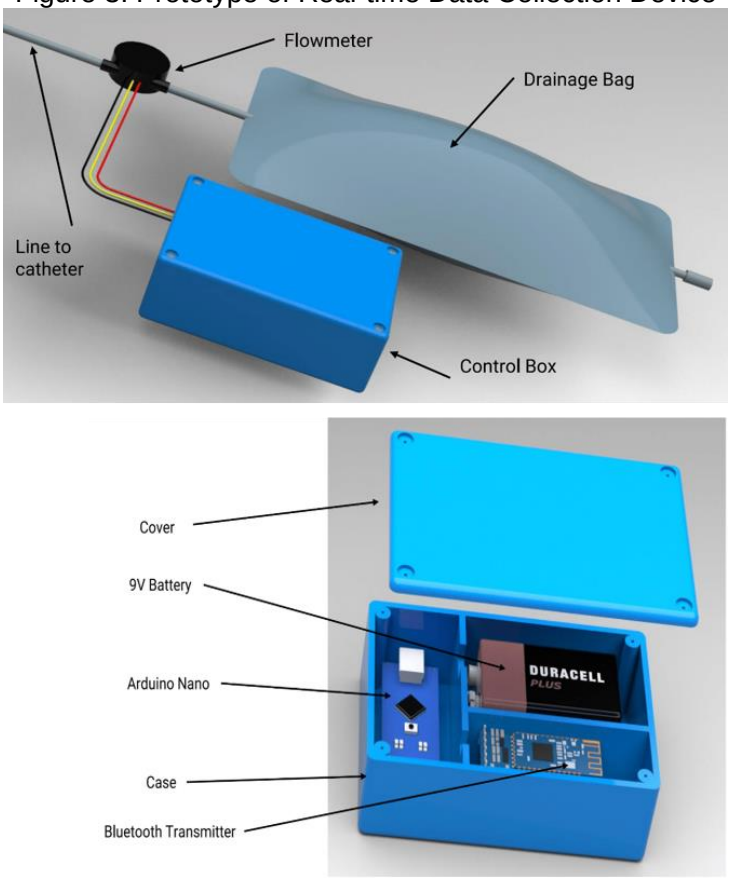
Flowmeter Design for Data Acquisition

In order to advance technology associated with the UTC drainage bag systems, a product must be developed that will accurately monitor the capacitance of the drainage bag and send out signals regarding the status of the bag in a timely manner. Our design resulted in a modular device that utilized an internal fluid flow sensor embedded into the catheter tube in order to measure the rate of drainage into the attached bag. The device then, upon the bag reaching its threshold volume (half of the total capacity), sent a signal via Bluetooth. The CAD model of a real-time data collection device prototype is shown in Figure 3.

The real-time fluid capacity was monitored and analyzed compared to the fluid flow sensor product specifications, ensuring that the accuracy and precision are within a five percent error during the calibration process. In addition, the implementation of the fluid flowmeter in the catheter tubing was tested to ensure that there is no leakage between the components of the prototype. The Bluetooth sensitivity was also tested through distance checks on the connectivity between a smartphone and the microcontroller to identify the radius at which the reliability faltered. Finally,

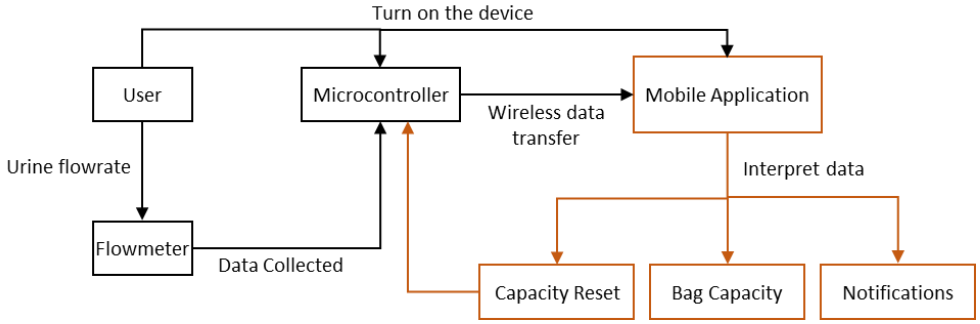
the pulse signal and real-time fluid analysis were compared to the fluid capacity levels of the leg bag, ensuring the accuracy and precision of the overall device.

Figure 3: Prototype of Real-time Data Collection Device



The real-time fluid capacity was monitored and analyzed compared to the fluid flow sensor product specifications, ensuring that the accuracy and precision are within a five percent error during the calibration process. In addition, the implementation of the fluid flowmeter in the catheter tubing was tested to ensure that there is no leakage between the components of the prototype. The Bluetooth sensitivity was also tested through distance checks on the connectivity between a smartphone and the microcontroller to identify the radius at which the reliability faltered. Finally, the pulse signal and real-time fluid analysis were compared to the fluid capacity levels of the leg bag, ensuring the accuracy and precision of the overall device.

Figure 4: Dataflow Diagram of the Urine Flowmeter



The data flow diagram of the urine flow meter is shown in figure 4. The microcontroller receives the raw data from the flowmeter and transfers it to the mobile application. Patients can see the drainage bag capacity on the mobile application and reset the capacity counter while emptying the drainage bag. Notifications are sent when the patients do not empty the bag on time.

CONCLUSIONS

The proposed system design for post-discharge patient care assistance applications utilizes smart sensing, predictive learning and intelligent reasoning to provide patient-centered health care services to discharged patients. Patient-centered health care services could reduce the cost of healthcare and the workload and stress of patients. The sensing accessories in the system collects data and sends the signal, together with the data-drive predictive model and mobile application platform, that contributes to the smart decision-making process of the checking activities of the post-discharge patients. Future work could concentrate on improving the efficiency of the data analysis model.

ACKNOWLEDGMENTS

Anviksha Busa, Christopher Chao, Haider Jamal, Achint Lehal, Chukwuanugo Mojekwu, and Ricardo Riveraare acknowledged for their valuable contributions to prototype development and running the experiments.

REFERENCES

Centers for Disease Control and Prevention. (2015). *"Catheter-Associated Urinary Tract Infections (CAUTI)"* Centers for Disease Control and Prevention. www.cdc.gov/hai/ca_uti/uti.html.

Enshaeifar, Shirin, et al. *Machine learning methods for detecting urinary tract infection and analyzing daily living activities in people with dementia*. PloS one 14.1 (2019): e0209909.

Garapati, Sankeerth S., et al. *Urinary bladder cancer staging in CT urography using machine learning*. Medical physics 44.11 (2017): 5814-5823.

Hana Kerzman, Orna Baron-Epel, Orly Toren,. (2005). *What do discharged patients know about their medication?*. Patient Education and Counseling,Volume 56, Issue 3, Pages 276-282.

Jones, C.D., Ginde, A.A., Burke, R.E., Wald, H.L., Masoudi, F.A. & Boxer,. (2015). *Increasing Home Healthcare Referrals upon Discharge from U.S. Hospitals: 2001–2012*. J Am Geriatr Soc, 63: 1265-1266.

Jones, C.D., Jones, J., Richard, A. et al. (2017). *"Connecting the Dots": A Qualitative Study of Home Health Nurse Perspectives on Coordinating Care for Recently Discharged Patients.* J GEN INTERN MED 32, 1114–1121.

Liang, Yixiong, et al. *Object detection based on deep learning for urine sediment examination*. Biocybernetics and Biomedical Engineering 38.3 (2018): 661-670.

Lin, Pei-Chen, et al. *A machine learning approach for predicting urine output after fluid administration*. Computer methods and programs in biomedicine 177 (2019): 155-159.

Lyu, K. X., Zhao, J., Wang, B., Xiong, G. X., Yang, W. Q., Liu, Q. H., Zhu, X. L., Sun, W., Jiang, A. Y., Wen, W. P., & Lei, W. B. (2016). *Smartphone Application WeChat for Clinical Follow-up of Discharged Patients with Head and Neck Tumors: A Randomized Controlled Trial*. Chinese medical journal, 129(23), 2816–2823.

Sokolov, I., et al. *Non-invasive diagnostic imaging using machine-learning analysis of nanoresolution images of cell surfaces: Detection of bladder cancer*. Proceedings of the National Academy of Sciences 115.51 (2018): 12920-12925.

Tun, William, Ognjen Arandjelovic, and Peter David Caie. *Using machine learning and urine cytology for bladder cancer prescreening and patient stratification*. Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

Zheng, B., He, H., Lin, C. (2011). *Development and application of short message system for follow-up in discharged patients.* Journal of Nursing Science

DECISION SCIENCES INSTITUTE

Detecting Shifts in Distributions of Search Inquiries: A Simplified Tool to Assess Influenza Averages

Matthew Lindsey
Stephen F. Austin State University
Email: lindseymd@sfasu.edu

Robert Pavur
The University of North Texas
Email: pavur@unt.edu

ABSTRACT

The ability to build a model to identify above average flu seasons can provide hospital managers important information to schedule limited resources effectively. In addition business managers and educators can anticipate when absenteeism will be above historical averages due to flu sickness. While the model is built using Google searches for "influenza", it could possibly be applied to other Influenza Like Illnesses (ILI)s that are searched for on the search engine. Many sophisticated methodology has been used in modeling ILI outbreaks. The approach in this paper is designed to address assessments what is considered "average" for a flu season.

KEYWORDS: Forecasting, Single Exponential Smoothing, Influenza Like Illnesses

INTRODUCTION

Influenza cases severely impact businesses and health communities in the United States is well documented by the Center for Disease Control (CDC). The CDC estimates that influenza (flu) was associated with more than 35.5 million illnesses, more than 16.5 million medical visits, 490,600 hospitalizations, and 34,200 deaths during the 2018–2019 flu season (https://www.cdc.gov/flu/index.htm). The flu is responsible for increased emergency room visits and hospitalizations which provide additional stress on our medical system (Dugas, Jalapour, Gel, Levin, Torcaso, Igusa & Rothman, 2013;). The GDP of the United States is burdened by flu cases each year as a result of hospital costs and lost productivity from missing work (Molinari, Ortega-Sanchez, Messonnier, Thompson, Wortley, Weintraub, & Bridges, 2007). Understanding influenza models can provide a basis to develop and understand models for Influenza-Like Illnesses (ILI) like the COVID—19. It is important to be able to determine how early the flu season might arrive or how late the flu season will be. The research in this paper illustrates an approach that is simplified and does not require additional data such as economic or metrological data.

Identifying peak flu timing and peak flu intensity precisely is elusive and thus broad measures of the flu infected population are used in planning and resource allocation. By knowing when demand for resources will be higher than average, managers can proactively schedule personnel and equipment to respond accordingly with robust resource management strategies. The procedure presented in this paper is novel in that it adapts a widely used forecasting

technique mostly used for inventory management to assessing shifts in the distribution of disease data.

LITERATURE REVIEW

The ability to produce current and accurate forecasts of annual flu cases are a valuable tool for businesses and health managers to plan strategies to counteract the negative impacts of flu outbreaks. Producing timely and accurate forecasts to predict the flu due to the mechanics of flu transmission challenge many forecasting tools due to limited historical data that appears to be "noisy" with unpredictable upsurges (Osthus, Gattiker, Priedhorsky and Del Valle; 2019).

Forecasting the Flu

Flu forecasts by the CDC are based on reports of clinical symptoms, virology results, admits to hospitals and mortality and are widely reported with a publication of data lagging real time data by only a few weeks (Dugas et al. 2013). Indirect methods, such as Google Flu Trends have been used to detect surges in influenza-like illnesses as well, with much less lag and provide hidden information that may be underreported. The importance of flu forecasting is highlighted by the CDC's periodic national flu forecasting competitions occurring since 2013 (Biggerstaff et al, 2016). Comparing Google Flu searches it is clear that it closely follows peaks in hospitalizations reported by the CDC for the flu.

Model Categorizations

Flu forecasting models are generally placed into four groups: Mechanistic models, agent-based models, regression models and dynamic models (Osthus et al, 2019). Araz, Bently and Muelleman (2014) created different time series models to predict emergency department visits during the flu season augmented with Google Flu Trends data as a predictor. The three types of time series models were ARIMA, Exponential Smoothing (Holt-Winters) and Linear Regression. Recent research by Kandula and Shaman (2019) indicate that much of the observed errors in Google Flu Trends data, was correctable and could still be a viable method of forecasting spikes in flu cases. Chen, Hossain, Butler, Ramakrishnan and Prakash (2016) incorporated Twitter tweets into models forecasting flu outbreaks and flu peak predictions.

ADJUSTING THE FORECAST

The quality of a forecast depends on how accurate it is in predicting future values. After an acceptable methodology is adopted, from time to time, different models should be applied to determine if another method will reduce the error term. With the seasonal like behavior of the data set, it is reasonable to suspect that a different methodology is appropriate during "high" flu season than during "low" flu season. The proposed methodology looks for a "trigger" to indicate when the "high" flu season methodology is appropriate and when the "low" flu season methodology is appropriate.

SINGLE EXPONENTIAL SMOOTHING

Single Exponential Smoothing (SES) is a weighted-moving average technique conventionally used to forecast demand for inventoried items and can be referenced in any basic statistics text.

DUAL CROSTON MODEL: TWO LEVELS OF COUNTS

If all observations are positive and there are no intermittent data, then Croston's method and SES yield the same result. Typically, intermittent data is problematic and Croston's method was intended to alleviate the shortcomings of SES with intermittent demand or sales data. Boylan and Syntetos (2007) correct for a bias using Croston's method by multiplying the demand per period by 1−*α*/2, that is, by 1 minus the smoothing constant divided by 2. Teunter and Sani (2009) advocate that in some cases when only a limited periods have no demand, Croston's method excels and when most periods have no demand, Syntetos and Boylan's bias corrected technique is better. Croston's (1972) suggested that ideal results would be attained with small alpha values between 0.1 and 0.2. Willemain et al (1994) supports that lower values generally performed better as well.

The application in this paper is flu cases and the data consists of scaled counts. Since the distribution of the counts appear to vary according to times when the flu season is strongest, there is a rationale for estimating the data when the observation counts are high separately from when they are low. This method assumes that a value T_0 can partition the time series into a "high count" series and a "low count" series. Two Croston models are described below, one for counts of the time series when the count is above T_0 and one for when the time series is below T_0. The variable Q in the following model is a count of the time periods between demand Y_t. Let T_0 be the separator between the high and low partitions of the time series data.

Croston Model for "High Partition":

If $Y_t < T_0$, $Z^H_t = Z^H_{t-1}$
$P^H_t = P^H_{t-1}$ $Q^H = Q^H + 1$
Else $Y_t > T_0$, $Z^H_t = Z^H_{t-1} + \alpha(Y_t - Z^H_{t-1})$
$P^H_t = P^H_{t-1} + \alpha\,(Q^H - P^H_{t-1})$ $Q^H = 1.$
Total_Count_H = Total_Count_H + 1 (1)

Croston Model for "Low Partition":

If $Y_t > T_0$, $Z^L_t = Z^L_{t-1}$
$P^L_t = P^L_{t-1}$ $Q^L = Q^L + 1$
Else $Y_t < T_0$, $Z^L_t = Z^L_{t-1} + \alpha(Y_t - Z^L_{t-1})$
$P^L_t = P^L_{t-1} + \alpha\,(Q^L - P^L_{t-1})$ $Q^L = 1.$
Total_Count_L = Total_Count_L + 1 (2)
Total_Count = Total_Count_H + Total_Count_L (3)
The mean demand per period is then
$(\text{Total\_Count\_H} * Z^H_t / P^H_t + \text{Count\_Fast} * Z^L_t / P^L_t)/(\text{Total\_Count})$ (4)

DATA DESCRIPTION

The data for this study was obtained from the Google Trends site on June 2, 2020. Google trends provides scaled search term results. The term "influenza" was used as the key search term. However, "flu" provided similar trend results. The data are provided from 2004 to present. The results are scaled based on the month in the search period with the highest number of

searches. The data is available by country, state and city region. The peak period of searches in this study was in April 2009, closely followed by March 2020. Graphs are provided below of the results for the search term “influenza” since 2004 for the United States, Texas and the Dallas-Ft. Worth metropolitan area.

As seen in the figures, in this particular case, it is informative to have the results at differing levels of granularity. It is interesting that in this particular case, the peak on the national level is in 2020, but on a state and metropolitan area level, the peak was in 2009. Another interesting observation is that the peaks of particular “flu seasons” vary. It can be as early as October or November or as late as March or April. This is valuable to allow local resources to be deployed optimally which might not be the case based on national data.

Figure 1. “Influenza” search results for the United States 2004 to present

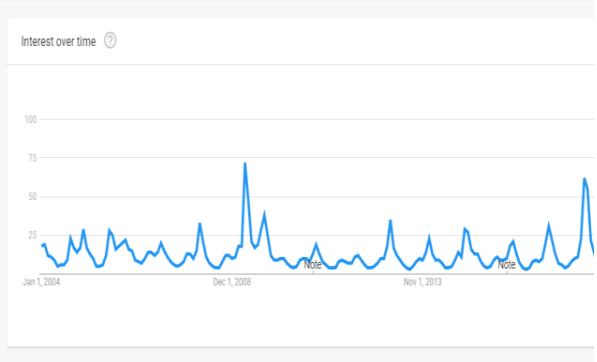


Figure 2. “Influenza” search results for Texas 2004 to present

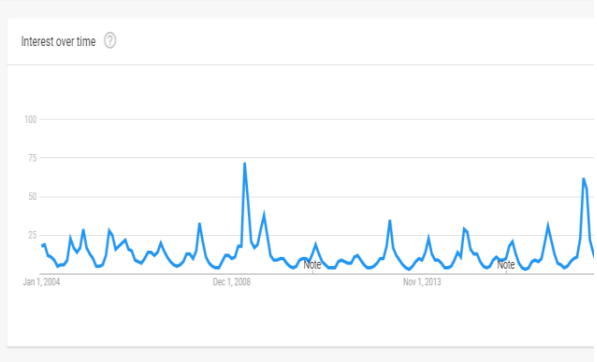
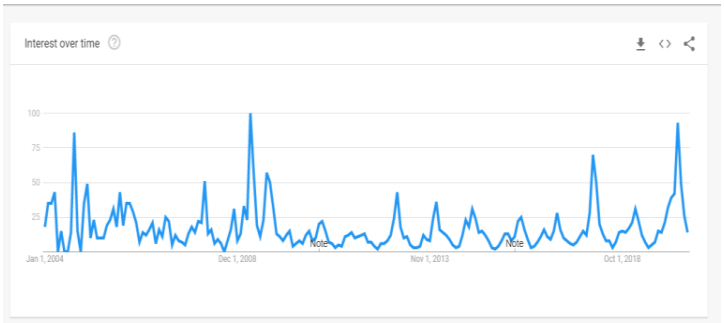


Figure 3. “Influenza” search results for the Dallas-Ft. Worth metro area 2004 to present



COMPARATIVE STUDY

Data were collected using Google Trends for searches for the term “influenza” on Google in the Dallas Ft. Worth Metro area from 2004 to present. The data are scaled data, with 100 being the month with the highest searches. A value of 10 indicates that 10% of the highest month searches were performed in that period. The raw number of searches could easily be used by multiplying by the actual number of searches for the peak month.

The series was split into high levels of searches and low levels of searches. Two cut off rates were examined of 9.5 and 15.5. In each case, if the level was above 9.5 or 15.5 it was considered a high rate (active flu concerns) and if below, a low rate (flu concerns exists, but

nothing exceptional). Forecasts for the average rate were then made using Single Exponential Smoothing (SES) and the Dual Croston methodology.

For the Dual Croston methodology, two different forecast models were used. One for the "High" and one for the "Low" partitions of the time series. The same value of the smoothing constant is used for the Dual Croston methodology and the SES model. When the data is split into a high group and a low group, smoothed values can be more accurate for each partition more so then the entire time series is used.

The objective in this study is to estimate the average count or scaled count for the long term. To plan for the longer term, an accurate ***average*** count per period measure is more beneficial for planning resources then being able to forecast the current count of individuals with the flu. In addition, the smoothed average provides a more realistic comparison to the long term historical average with regard to intensity of the present flu environment. The Root Mean Squared Error (RMSE) is computed to allow for comparison between SES and the Dual Croston.

RESULTS

The prediction error (RMSE) was computed for two different split levels of 9.5 and 15.5. Table 1 illustrates that the lowest RMSE was achieved with the smoothing constant of 0.1 and the separator of 15.5 using the Dual Croston methodology that adapts for different levels of flu activity. Figure 6 shows forecasts from SES and the Dual Croston method. SES appears to be more responsive to changes than the other method. The "High" partition is referred to a "Fast" and the "Low" partition is referred to as "Slow."

Table 1. Root Mean Square Error over three Smoothing Constants

Hi/Lo Separator = 9.5				Hi/Lo Separator = 15.5		
Smoothing Constant	RMSE for SES	RMSE for Hi/Lo Split		Smoothing Constant	RMSE for SES	RMSE for Hi/Lo Split
0.1	4.9205	3.0184		0.1	4.9205	1.9284
0.2	6.9161	4.2304		0.2	6.9161	2.7497
0.3	8.4491	5.1439		0.3	8.4491	3.3548

DISCUSSION AND CONCLUSIONS

By splitting data into high and low groups, a more effective forecast can be obtained than using SES alone for the entire series. Applying different forecasting methodologies to upper and lower partitions of the data provides more accurate forecasts. The high group tends to have more variability and fluctuates more than the low group. The low group is bounded by zero and has a limited range. Many periods during the low period are near zero.

Forecasting models that use seasonality may hold promise for patterns in counts of flu infected individuals. The difficulty with seasonality with this type of data is that the flu season starts and stops at different points during the year and may have an early start or linger much longer than typical. As illustrated in the results of our analysis, the accuracy with regard to estimating the average flu count lies in selecting the T_0 separator value. That is, the challenge lies in picking the value that separates the high and low groups. Defining what levels are truly high and what levels are low will require an additional study.

REFERENCES

Araz, O. M., Bentley, D., & Muelleman, R. L. (2014). Using Google Flu Trends data in forecasting influenza-like–illness related ED visits in Omaha, Nebraska. *The American journal of emergency medicine,* 32(9), 1016-1023.

Biggerstaff, M., et al (2016). "Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge." *BMC Infectious Diseases*, 16(1): 357.

Boylan, J. E. and Syntetos, A. A., (2007). The accuracy of a modified Croston procedure. *International Journal of Production Economics, 107*(2), 511-517.

Center for Disease Control and Prevention: 2018-2019 Influenza. Available: https://www.cdc.gov/flu/about/burden/2018-2019.html. Accessed May 20, 2020.

Chen, L., Hossain, K. T., Butler, P., Ramakrishnan, N., & Prakash, B. A. (2016). Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. *Data mining and knowledge discovery*, 30(3), 681-710.

Croston, J. D., (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3), 289-303.

Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. E. (2013). Influenza forecasting with Google flu trends. *PloS one*, 8(2).

Kandula S & Shaman J (2019) Reappraising the utility of Google Flu Trends*. PLoS Comput Biol* 15(8): e1007258. https://doi.org/10.1371/journal.pcbi.1007258

Molinari, N.-A. M., Ortega-Sanchez, I. R., Messonnier, M. L., Thompson, W. W., Wortley, P. M., Weintraub, E., and Bridges, C. B. (2007). "The annual impact of seasonal influenza in the US: measuring disease burden and costs." *Vaccine*, 25(27): 5086–5096.

Osthus, D., Gattiker, J., Priedhorsky, R., & Del Valle, S. Y. (2019). Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Analysis, 14*(1), 261-312.

Teunter, R.H., & Sani, B., (2009). On the bias of Croston's forecasting method, *European Journal of Operational Research, 194*(1), 177-183.

Willemain, T. R., Smart, C. N., Shockor, J. H. and DeSautels, P. A., (1994). Forecasting intermittent demand in mfg: A comparative evaluation of Croston's method. *IJF, 10*(4), 529-538.

Data obtained from:
https://trends.google.com/trends/explore?date=all&geo=US-TX-623&q=influenza
2004 to present Scaled to 100-

DECISION SCIENCES INSTITUTE
A Preliminary Assessment of the IHME Covid-19 Model

Thomas R. Robbins
East Carolina University
Email: robbinst@ecu.edu

ABSTRACT

The Covid -19 forecasting model published by the Institute for Health Metrics and Evaluation (IHME) is one of the most influential, consequential, and controversial forecasting models ever published. The model has been widely cited by policy makers, but it has also been widely criticized. In this paper we make an objective, external review of the model. We evaluate the outputs of the model, how they changed, and how they compared to actual results. Overall, we find the accuracy of the model was poor and the modeling objectives may have been overly ambitious.

KEYWORDS: Covid-19, Forecasting, Forecast Accuracy

INTRODUCTION

The Covid -19 forecasting model published by the Institute for Health Metrics and Evaluation (IHME) is one of the most influential, consequential, and controversial forecasting models ever published. The IHME model is the model most widely cited by US government policy makers. It has greatly influenced policy makers at both the federal and state levels. The projections coming out of the model have been used to help justify drastic interventions that have shut down vast sections of the US economy and led to an unprecedented economic contraction that led to a record rise in unemployment. While the IHME model is widely cited by policy makers, it has also been widely criticized from all sides.

In this paper we conduct an external review of the model; external in the sense that we do not concern ourselves directly with the methodology or algorithm used be the model, only with the output of the model. We effectively take the point of view of a policy maker and treat the model as a black box. Our focus is on evaluating the outputs of the model, how they varied over time, and how they compared to actual reported data. While the model forecasts multiple quantities, including hospitalizations, ICU usage, and ventilator usage, we focus primarily on forecasted fatalities. We also limit the analysis to the United States, examining both national and state level forecasts. Our review is preliminary in the sense that it includes data through May 31, 2020. While the wave one peak of has passed by this time, the pandemic is far from over and additional waves are possible. The analysis is limited based on the timing of the deadline for DSI conference submission for which this paper has been prepared.

In the remainder of this paper we review the IHME and the model they have put forward including the key criticisms of the model, we then briefly review related literature, and summarize the sources for the data used in this analysis. We then focus on our evaluation of the model, including:

- **Evolution of the Model**: in this section we analyze and critique the series of rapid and often significant changes in the model's forecasts. We only evaluate changes in forecasts and do not consider actual results.
- **Daily Deaths**: in this section we analyze day by day forecasts of fatalities. We compare the forecasted fatality count with the actual results recorded. We focus on the predictive accuracy of each model iteration over a 14-day forecasting window.

- **Cumulative Deaths**: in this section we evaluate the overall fatality forecast focusing on the last date available, May 31, 2020. We compare how accurate each iteration of the model was in predicting end of May cumulative fatalities at the national and state level. We calculate standard forecast metrics summarized at the state and model level.

THE IHME AND THE IHME MODEL

The Institution of Health Metrics and Evaluation (IHME) is an independent population health research center associated with the University of Washington (IHME, 2020). The organization gathers information on the world's largest health issues and evaluates the strategies used to address them. IHME's research stems from three different questions.

1.) What are the world's major health problems?
2.) How well is society addressing these problems?
3.) How do we best dedicate resources to maximize health improvements?

The IHME developed COVID-19 models and projections in response to requests from the University of Washington and other hospitals and government agencies concerned that Covid-19 would overwhelm hospital capacity. Their model forecasts daily and cumulative requirements for key hospital resources, as well as Covid-19 related fatalities. The IHME's stated goal is to release updates as frequently as possible, updating their model with the latest available information. The IHME provides access to their forecasts via an interactive visualization tool on their website. They also allow the model's forecasts to be downloaded from their website in a .csv file format. The first set of projections posted on the website is dated March 25, 2020 and by the end of May 25 different files have been posted. Over a period of 67 days 25 files were posted, an average of one new iteration every 2.68 days.

The first iteration of the IHME model provided a four month forecast on the impact of Covid-19 pandemic (Murray, 2020). That first model included forecasts for total hospital beds, ICU beds, ventilators, and deaths. The original forecast predicted a total of 81,114 deaths with a 95% confidence interval of [38,242 , 162,106]. In that first iteration of the model the pandemic was expected to have been largely over by the end of May and the mean projected death count for that time was 78,320, 96.6% of the total. Note that as of May 31, 2020 the total fatality count stood at 104,229 according to Johns Hopkins; a figure significantly higher than the original mean forecast, but well within the 95% confidence interval.

The IHME model has been widely cited by US public health officials, "Dr. Deborah Birx, the White House coronavirus response coordinator, has repeatedly cited the IHME model at press conferences, and journalists have often asked officials about it, too." (Azad, 2020). Dr. Birx again specifically mentioned the IHME model in early April when the model's forecasted fatalities dropped significantly, from 81,766 to 60,415 a change she justified based on increased social distancing policies (Carvajal, 2020).

Criticism of the IHME Model

The IHME model has been widely criticized for multiple reasons in the mainstream media, on social media, and in the academic literature. Critiques of the model include:

- Overly optimistic projections
- Overly pessimistic projections
- No epidemiologically basis
- Too unstable – large changes in the forecast over short time periods.

- Too uncertain – confidence levels that are too wide for practical use.
- Too short a time horizon – only a four-month time horizon with no allowance for a second wave
- Forecasting zero deaths at the end of the forecast period
- Based on bad data and assumptions

The model has been criticized for being too optimistic, especially after it lowered estimated fatalities in early April and generated forecasts inconsistent with and lower than other models (Wan & Johnson, 2020). The same downward revision caused others to criticize the original model as too pessimistic. Critics pointed out the lack of a national standard model but noted the IHME model was receiving significant support from the White House. Other critics have also identified the disparity across models and the wide range of uncertainty (Boice, 2020).

The model has been criticized because it is not based on a traditional Susceptible-Infectious-Recovered (SIR) model, but rather it is a data fitting model that attempts to extrapolate, at least initially, from Chinese data. Critics charged that the comparison to China was "inherently optimistic because it assumes that all states respond as swiftly as China" (Wan, Dawsey, Parker, & Achenbach, 2020). Dr. Anthony Fauci, the longtime director for the National Institute of Allergy and Infectious Diseases, discounted the use of models in general when he said models are only as good as the assumptions put into them (Hatmaker, 2020). The models are also dependent on the quality of the data they use, especially a curve fitting model like the IHME model. The issues associated with Covid-19 data are significant and are addressed in (Ioannidis, 2020) and reviewed in detail in Robbins (2020). Perhaps the most significant issue is the under-reporting of actual deaths in China, since the early versions of the model relied so heavily on the Chinese data. A non-academic review in early May posted on Vox.com focused on the optimistic nature of the model relative to actuals at that point as well as its divergence from other models and questioned the rationale for continued reliance on the model by policy makers (Piper, 2020). A different critique, posted a few weeks earlier at issueinsights.com, criticized the model for being too pessimistic, noting the significant downgrades in projections in early April (Fumento, 2020). That author questioned all the models and suggested "It's Time to Permanently Dump Epidemic Models."

Harsh criticism of the model for rapidly changing forecasts is documented in Begley (2020). A quote from that article summarizes the view of some in the traditional epidemiological modeling community, "That the IHME model keeps changing is evidence of its lack of reliability as a predictive tool," said epidemiologist Ruth Etzioni of the Fred Hutchinson Cancer Center, "….That it is being used for policy decisions and its results interpreted wrongly is a travesty unfolding before our eyes."

A paper posted on the non-peer reviewed site Medium.com, but co-authored by Wesley Pedgen of Carnegie Mellon, was highly critical of the Covid-19 models in general, and the IHME model in particular, for the short term forecasting horizon, in particular the assumption of a zero death-period and no second wave (Chikina & Pegden, 2020). This paper argues that the strict social distancing imposed by lockdowns across the United States may flatten the curve, but are likely pushing deaths into a second wave outside the forecast window of the models. Another report, written by academics but posted without peer review, criticized the IHME model for the high number of forecasts that fall outside the model's 95% confidence intervals (Marchant, Rosen, Tanner, & Cripps, 2020). This review was dated April 8th, still early in the outbreak. It examined forward predictions of one to four days made at the end of March and found that 49%-73% of

the forecasts are outside the confidence level and they conclude the model dramatically underestimates uncertainty.

RELATED LITERATURE

This paper is not an epidemiological paper or even a modelling paper. We do not consider the internal working of the IHME model. Background information on epidemic models can be found in a review format (Chowell, Sattenspiel, Bansal, & Viboud, 2016) and a tutorial format (Dimitrov & Meyers, 2010). Our focus is more on the assessment of forecasting predictions.

Multiple streams of research address issues of forecasting and forecasting accuracy. In the statistical community forecasting competitions are common. These competitions generally involve the evaluation of specific methodologies, rather than individual models. A somewhat dated review of empirical assessment of forecasting models is provided in Fildes and Makridakis (1995). A more recent evaluation of forecasting methods is the M-3 competition, featured in a special issue of *The International Journal of Forecasting* (Ord, Hibon, & Makridakis, 2000). The results and findings of the competition are summarized and reviewed in Makridakis and Hibon (2000). This was the third in a series of forecasting competitions that compare the performance of a large number of time series forecasting methods. Another forecasting competition, this one related specifically to the tourism industry, is chronicled in Athanasopoulos, Hyndman, Song, and Wu (2011).

A different type of forecasting competition is discussed in Tetlock (2017). Tetlock organized multiple forecasting tournaments over an extended period of time that focused on individual forecasts of major world events. He generally found forecasters did little better than randomized predictions and typically did worse than algorithms. The characteristics of good forecasters are further detailed in (Tetlock & Gardner, 2015). However, Tetlock's work is primarily oriented toward individual human forecasters rather than model-based forecasters.

An alternative stream of research examines the predictive ability of group-based prediction methods. In the book Infotopia, Cass Sunstein compares the predictive ability of group based methods such as polls and prediction markets (Sunstein, 2006). Sunstein makes the case that predictive markets can, under the right circumstances, outperform other predictive methods. Other research examines market-based forecasts, such as the prediction of commodity prices by futures markets (Alquist & Kilian, 2010; Chernenko, Schwarz, & Wright, 2004; Chinn & Coibion, 2009). This research tends to find that futures markets are subject to significant errors but are more accurate than other model-based approaches.

We were unable to find any peer reviewed detailed evaluations of specific models in general, or of the Covid-19 models in particular. Only the general press and non-peer reviewed papers identified previously. But a peer-reviewed paper in JAMA does outline the general principle and value of projections for Covid-19 models (Jewell, Lewnard, & Jewell, 2020). The paper argues that epidemiological models are best suited to evaluate "the relative effect of various interventions in reducing disease burden rather than to produce precise quantitative predictions about extent or duration of disease burdens." The paper does identify, and criticize, specific aspects of the IHME model. The paper states "[IHME model] has received considerable attention and has been widely quoted by government officials. On the surface, the model yields specific predictions of the day on which COVID-19 deaths will peak in each state and the cumulative number of deaths expected over the next 4 months (with substantial uncertainty intervals). However, caveats in these projections may not be widely appreciated by the public or policy makers because the model has some important but opaque limitations." It goes on to

criticize the IHME model for the volatility of fatality forecasts, citing the large change in NY projections at the beginning of April.

This paper provides an important contribution to the literature by providing a detailed, and fact-based assessment of the IHME model. Rather than pointing out individual flaws, out of context, we provide an objective assessment of the model, looking at all the different iterations provided, and comparing those projections to actual reported results. Any model as impactful as the IHME Covid-19 deserves careful scrutiny and evaluation so as to provide both modelers and policy makers with insights that can be applied in future forecasting situations.

DATA AND DATA SOURCES

Our analysis in this paper relies on three primary data sources. We utilize data on the IHME model projections along with actual data reported concerning the impacts of the pandemic. IHME is highly transparent and provides data on their model on their website (IHME, 2020). Each iteration of the model is provided as a separate downloadable .csv file, and release notes are posted for each version. Our data set includes every model posted by the IHME from March 25th to May 29th. Each model is identified with the date identified on the Covid-19 estimate download page. The data provided by the model has evolved along with the model, and while initial versions of the model focused only on the United States, later versions added international data. For our purposes we use only data focused on the US. The IHME projections include data for each state as well as the US as a whole. Later iterations of the model added data for some US territories such as Puerto Rico, but we restrict our analysis to the 50 states and the District of Columbia. Early versions of the model provided projections through August 4th. Beginning with the May 4th iteration of the model the projections for some metrics extend to August 24th, though the forecasts for US deaths only extended through August 4th. While additional data has been added over time, for this analysis we focus on the following data; deaths: the average number of deaths projected by day, by location, and totaldea: the cumulative number of deaths. Secondary data we use includes admis: total number of hospital admission, newICU: new admissions to Intensive care units, and InvVen: daily patients on ventilators. Each forecast parameter has a mean, and 95% confidence intervals14 days for each location each day. Additional measures we do not analyze in this paper include projected overages for beds and ICUs. We compare these models to actual data and we use data sources for actuals consistent with the data used by the IHME modelers.

Our second major source of data is the Johns Hopkins Coronavirus Resource Center (Johns_Hopkins, 2020). The resource center provides a number of different analysis and visualization tools. For our purposes we focus on the time series data they publish on Github (GitHub, 2020). The downloadable data sets include time series data for US counties. For this analysis we use only the deaths file, which documents total deaths. We use this data and calculate new deaths by a simple difference operation. We also aggregate this data to the state and national level. The site provides additional data on recoveries, as well as international cases that we do not include in this analysis.

Our third data source is the Covid Tracking Project Website (CovidTracking, 2020). Like the Johns Hopkins page, the Covid Tracking project provides data reporting, visualization and downloading services. The page provides a rich set of national and international data on a wide range of data points. In this analysis we use their data on hospitalizations. Note that Covid Tracking also provides data on case infections and deaths. For the most part that data is consistent with the John's Hopkins data with relatively minor differences, with Johns Hopkins numbers mostly larger. A notable exception is the state of New York where the two data series

differ significantly after mid-March with Johns Hopkins providing a number larger by about 5,000.

EVOLUTION OF THE IHME MODEL

Criticisms of the IHME model include the frequency with which the model has been updated and the major changes that occur from version to version. Between March 25th and May 29th, 25 different version of the model have been released. During these updates the mean estimate of peak deaths for the US as a whole varied considerably, from a low of 60.3K in the April 17th version to a high of 147K in the May 12th version. The peak death range has also fluctuated up and down over this time frame. Figure 1 shows the daily forecasted deaths for 6 selected versions of the model, while Figure 2 shows the cumulative forecasted deaths for the same 6 models. These graphs show the familiar bell curve shape standard for epidemic forecasts. In Figure 1 we see that the main changes in the models are the height of the peak, and the thickness of the tail. The peak grew in the first iterations. In the April iterations the peak was reduced, and the tail thinned out. The May versions show a significant thickening of the tail with an effective elimination of the zero-death region.

Figure 1- Mean Daily Death Count Forecasts

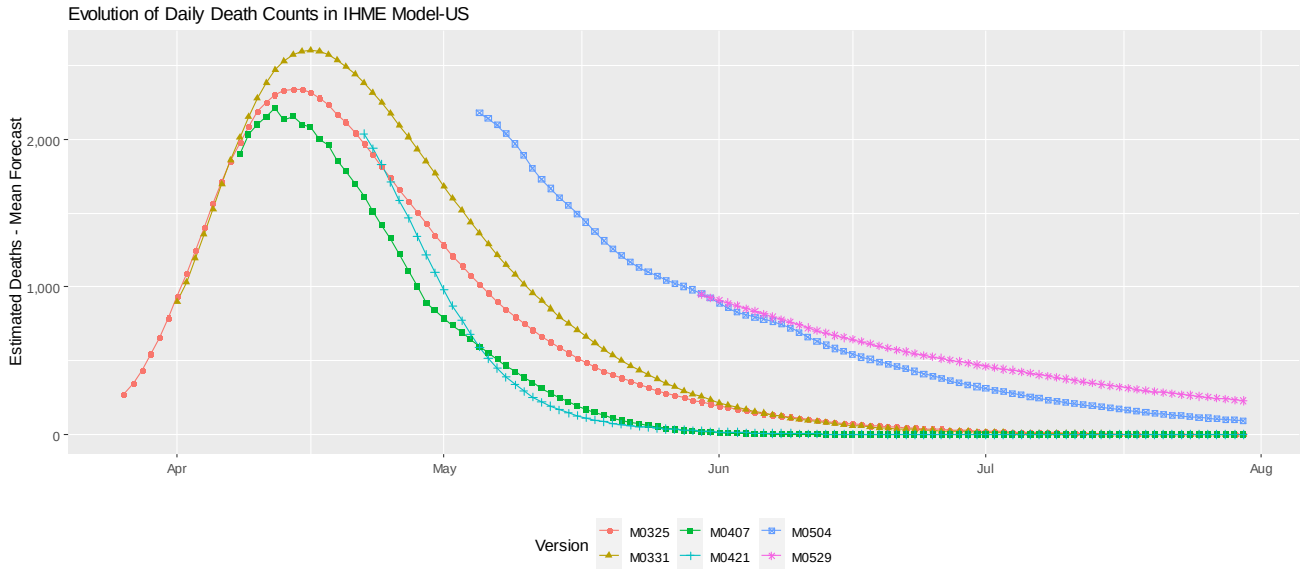


Figure 2 more clearly illustrates the changes in total predicted deaths. The total fatality forecast increased, then dropped significantly, the change in May increased the forecast from 72 to 134 thousand in a five-day period.

Figure 2- Mean Total Death Count Forecasts

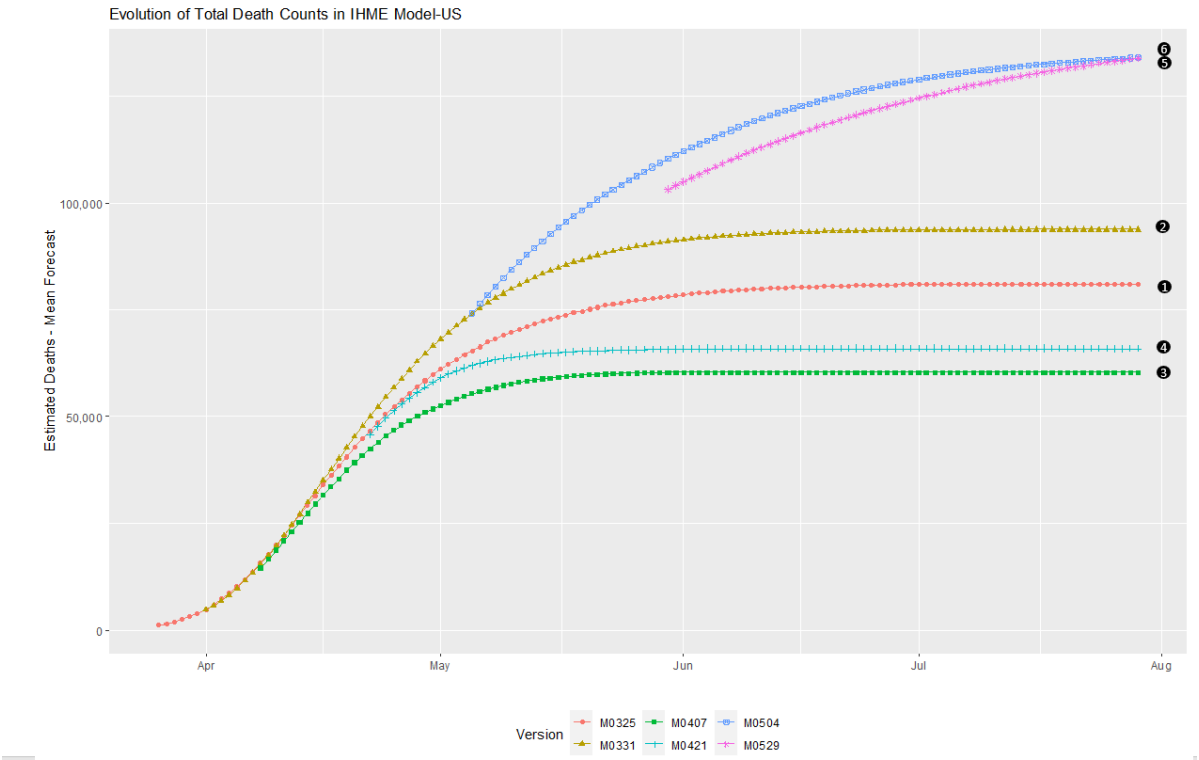
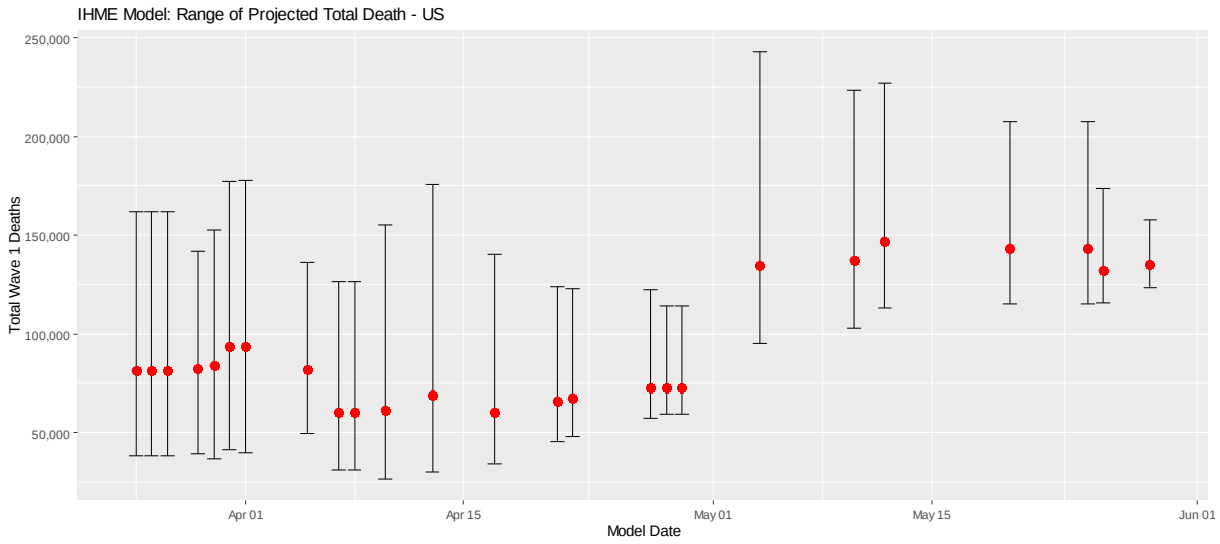


Figure 3 plots the mean and 95% confidence interval for the total death forecast for each model version. As the graphic shows, the uncertainty range has remained relatively large, while the mean estimate drifted downward before a major upward shift in May. The forecasted fatalities more than doubled from 72K at the end of April, to more than 147K mid-May.

Figure 3- Range of Projected Deaths for the US

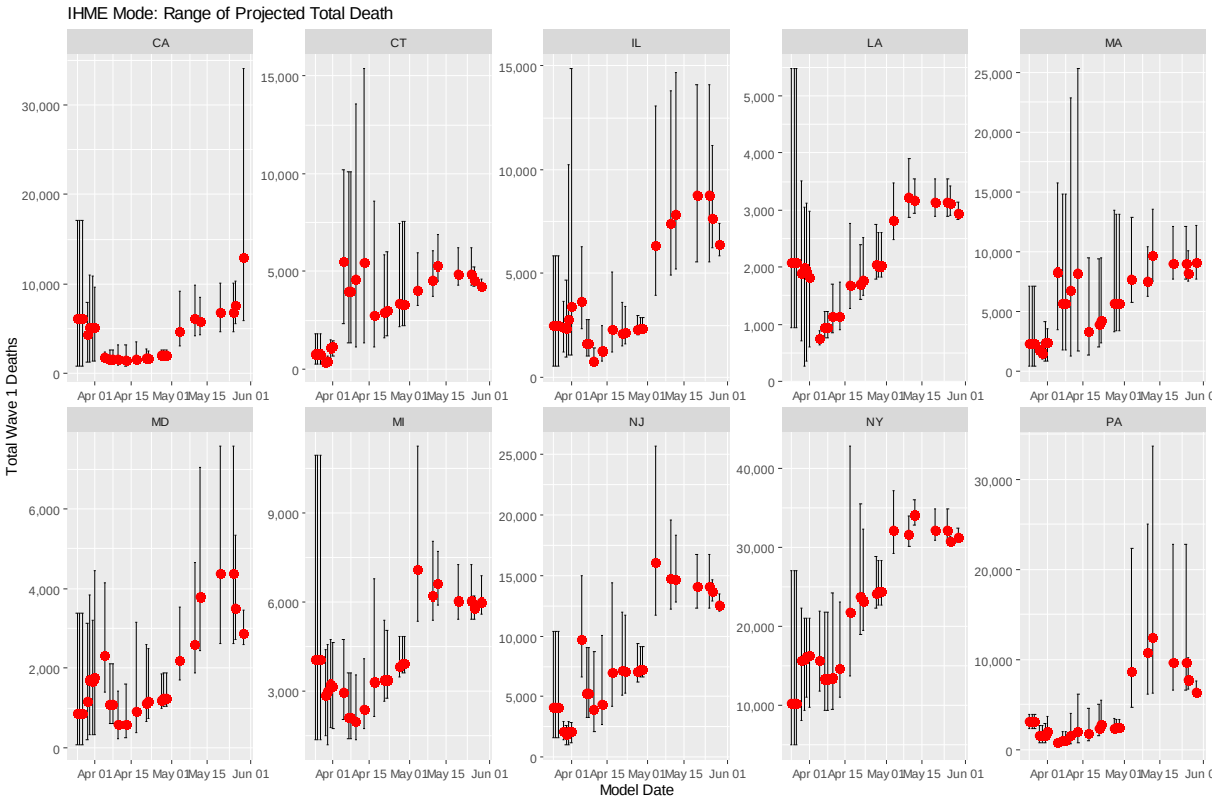


The predictions at the state level are somewhat more erratic. Table 1 shows the mean forecast for fatalities in the US as a whole, and for the 10 states with the highest death count as of late-May. Table 7 in the appendix shows this data for all states. The range of forecasts for these same states is shown Figure 4. As the states with the highest fatalities, these are also the most data rich states, the states for which the models have the most data to work with.

Table 1- Mean Death Predictions for Selected States

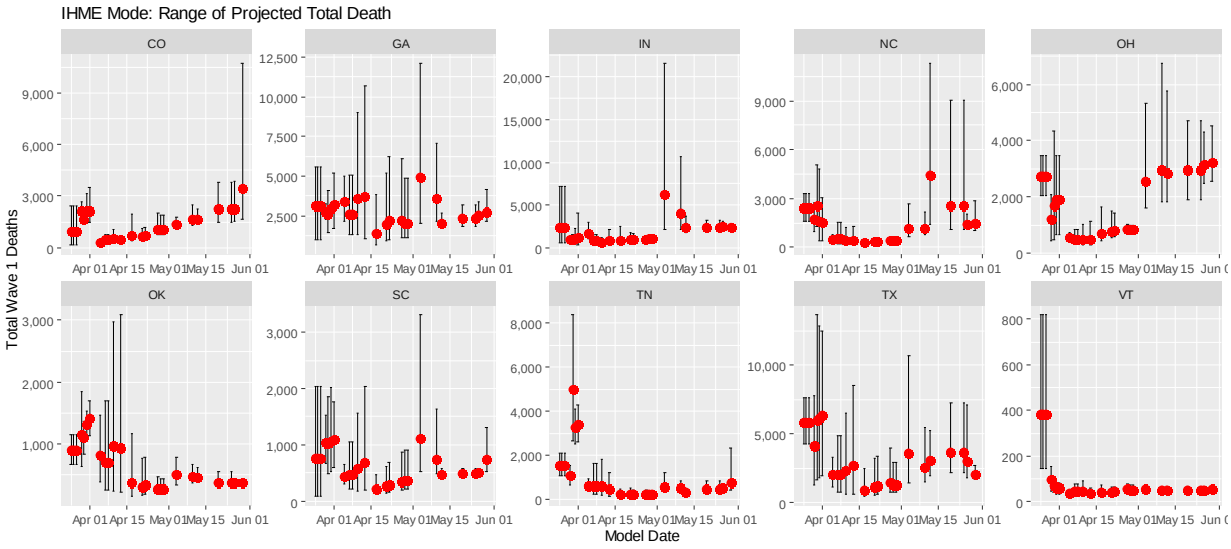
Model	US	NY	NJ	MA	PA	MI	IL	CA	CT	LA	MD
M0325	81,114	10,244	4,109	2,231	3,094	4,061	2,454	6,109	774	2,081	858
M0326	81,114	10,244	4,109	2,231	3,094	4,061	2,454	6,109	774	2,081	858
M0327	81,114	10,244	4,109	2,231	3,094	4,061	2,454	6,109	774	2,081	858
M0329	82,141	15,546	2,096	1,782	1,579	2,862	2,360	4,306	378	1,897	1,171
M0330	83,967	15,788	1,844	1,507	1,574	3,007	2,326	5,086	414	1,978	1,700
M0331	93,765	16,090	2,129	2,357	1,562	3,235	2,789	5,161	1,092	1,934	1,691
M0401	93,531	16,261	2,117	2,382	2,023	3,169	3,386	5,068	1,144	1,834	1,766
M0405	81,766	15,618	9,691	8,254	782	2,963	3,629	1,783	5,474	746	2,326
M0407	60,415	13,307	5,277	5,625	969	2,103	1,588	1,611	4,003	946	1,094
M0408	60,415	13,307	5,277	5,625	969	2,103	1,588	1,611	4,003	946	1,094
M0410	61,545	13,463	3,915	6,739	1,503	1,977	777	1,616	4,614	1,125	598
M0413	68,841	14,542	4,407	8,220	2,005	2,373	1,248	1,483	5,426	1,141	596
M0417	60,308	21,812	6,952	3,236	1,707	3,304	2,259	1,658	2,732	1,685	914
M0421	65,976	23,741	7,116	3,898	2,323	3,361	2,093	1,743	2,884	1,704	1,112
M0422	67,641	23,232	7,058	4,243	2,770	3,379	2,138	1,719	3,006	1,780	1,171
M0427	72,860	24,088	7,051	5,633	2,318	3,818	2,292	2,035	3,367	2,033	1,197
M0428	72,433	24,314	7,246	5,634	2,400	3,920	2,337	2,104	3,315	2,026	1,245
M0429	72,433	24,314	7,246	5,634	2,400	3,920	2,337	2,104	3,315	2,026	1,245
M0504	134,475	32,132	16,044	7,697	8,607	7,080	6,353	4,666	4,047	2,822	2,190
M0510	137,184	31,621	14,752	7,545	10,742	6,217	7,395	6,086	4,575	3,233	2,606
M0512	147,040	34,068	14,692	9,629	12,420	6,621	7,830	5,832	5,263	3,169	3,799
M0520	143,357	32,136	14,070	8,949	9,623	6,043	8,781	6,842	4,860	3,141	4,363
M0525	143,357	32,136	14,070	8,949	9,623	6,043	8,781	6,842	4,860	3,141	4,363
M0526	131,967	30,823	13,654	8,232	7,680	5,801	7,628	7,558	4,548	3,119	3,510
M0529	135,109	31,295	12,577	9,003	6,418	5,986	6,371	12,951	4,251	2,947	2,883
Minimum	60,308	10,244	1,844	1,507	782	1,977	777	1,483	378	746	596
Average	92,555	21,214	7,664	5,499	4,051	4,059	3,746	4,328	3,195	2,065	1,808
Maximum	147,040	34,068	16,044	9,629	12,420	7,080	8,781	12,951	5,474	3,233	4,363
Range	86,732	23,825	14,200	8,123	11,638	5,104	8,004	11,468	5,096	2,487	3,767
Std Dev	31,019	8,160	4,647	2,715	3,539	1,536	2,561	2,789	1,722	758	1,153
CV	0.335	0.385	0.606	0.494	0.874	0.378	0.684	0.644	0.539	0.367	0.638

Figure 4- Fatality Forecasts for Peak States



The data reveals significant shifts from version to version, both in the forecast mean as well as the width of the confidence interval. States such as CA, and IL saw large reductions in the mean and range between early and late April, only to see significant increases in mean and range in early May. Conversely CT levels jumped in early April, followed by a decline and then a rise again, increasing by a factor more than 14, dropping by nearly half in three days and then doubling again. LA saw a very large decline in early April, followed by a steady rise throughout April and May. The forecast dropped suddenly from 4,748 to 1,300, then rose again to more than 3,000. States such as MI, NJ, PA, and NY saw very large increases in both the mean and range of the forecast with the May versions of the model that coincided with the start of re-opening in parts of the country, though these states are not among those reopening. If we turn to the lower volume states, the models become even more erratic. Figure 5 shows the forecasts for 10 other states.

Figure 5-Fatality Forecasts for Other Select States

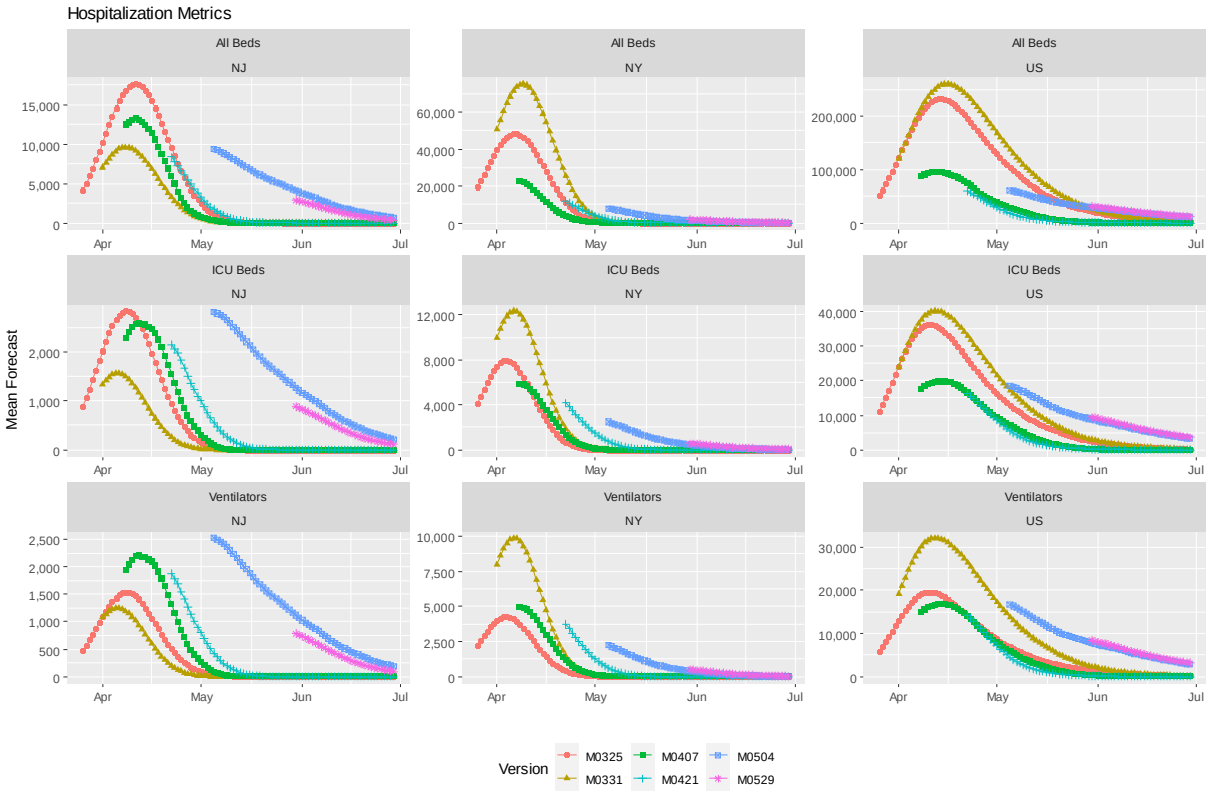


In CO we see a sudden drop from 2,153 to 302 along with a significant reduction in uncertainty. The forecast then steadily increases to more than 1,600 by mid-May while the level of uncertainty varies from version to version, and jumps significantly in the final version of the model. States such as OK, TN, TX, and SC saw very large reductions in the mean and range though these changes took place at different times. VT forecast was very high at first, but both the mean and range dropped dramatically in early April and stayed low. Conversely states such as NC and OH saw major increases in early April, in both the mean and range. IN saw a similar increase, though the level dropped significantly as did the range in early May. GA is an interesting case in that GA quite controversially began reopening business at the end of April. The forecasted fatalities shot up and became highly uncertain, only to drop down with a significant reduction in uncertainty in early May.

Hospitalization Metrics

While our focus in this paper is on forecasted fatalities, we will briefly examine the issue issues related to hospitalization requirements. Figure 6 shows 3 key hospitalization metrics; the total number of beds projected, the total number of ICU beds projected, and the total number of ventilators projected, for the US as a whole and for the states of NY and NJ. We select these two states as they represent the early epicenter of the outbreak in the US and the areas where hospitalization over utilization was considered a major risk. We see here the same type of erratic predictions we saw for fatalities.

Figure 6- Hospitalization Metrics for US and Key States



In NJ we see major shifts over a few weeks at the end of March/beginning of April. We see the peak demand for ventilators shift from about 1,500 down to about 1,250 and then up to 2,250 in two weeks while the peak date also shifted back and then forward. The early May model dramatically shifts the May requirements upward for all resources. Conversely in NY over the same two weeks demand shift up dramatically and then back down even more dramatically. For the US as whole demand increased somewhat from March 25th to the 31st, then dropped by half or more in the April 7th model, total hospitalizations were reduced from more than 225,000 to less than 100,000 over the course of that week. These metrics then stayed within a narrow range until the early May increase.

DAILY DEATHS

In this section we investigate the accuracy of the model in terms of forecasting daily deaths. One of the criticisms levelled against the model was the number of daily death observation outside the 95% confidence interval. To evaluate this problem, we conducted an analysis whereby we compared the reported fatality figure for each state with the confidence interval for each state. For each model we looked at predictions over the next 14 days after the forecast was published and computed the proportion of times the reported death count was within the interval. Table 2 shows the results for the 10 highest death states along with five other selected states for any model with at least a 14-day prediction window. Cells for which the number of predictions in bounds is less than 95% are highlighted. The table also includes averages by model and by state. Also note that with 14 observations, any prediction outside the confidence limit will result in a failure to achieve the 95% goal.

Table 2- Proportion of Actual Daily Fatalities within CI over 14-day Horizon

	M0325	M0326	M0327	M0329	M0330	M0331	M0401	M0405	M0407	M0408	M0410	M0413	M0417	M0421	M0422	M0427	M0428	M0429	M0504	M0510	M0512	Avg
CA	92.9%	92.9%	92.9%	28.6%	35.7%	21.4%	42.9%	100.0%	92.9%	85.7%	78.6%	92.9%	85.7%	57.1%	50.0%	35.7%	42.9%	35.7%	92.9%	71.4%	64.3%	66.3%
CO	28.6%	28.6%	28.6%	14.3%	7.1%	7.1%	0.0%	100.0%	85.7%	78.6%	78.6%	57.1%	92.9%	78.6%	71.4%	100.0%	100.0%	92.9%	35.7%	42.9%	35.7%	55.4%
CT	64.3%	71.4%	71.4%	0.0%	0.0%	21.4%	21.4%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	92.9%	92.9%	92.9%	92.9%	92.9%	78.6%	64.3%	64.3%	72.5%
FL	35.7%	35.7%	28.6%	42.9%	35.7%	42.9%	50.0%	92.9%	64.3%	57.1%	64.3%	64.3%	100.0%	92.9%	92.9%	100.0%	100.0%	100.0%	71.4%	71.4%	64.3%	67.0%
GA	35.7%	28.6%	35.7%	7.1%	14.3%	7.1%	35.7%	78.6%	57.1%	57.1%	50.0%	78.6%	100.0%	100.0%	71.4%	92.9%	100.0%	100.0%	50.0%	64.3%	35.7%	57.1%
IL	92.9%	92.9%	92.9%	21.4%	35.7%	57.1%	64.3%	100.0%	100.0%	100.0%	28.6%	50.0%	85.7%	57.1%	57.1%	35.7%	21.4%	14.3%	85.7%	78.6%	57.1%	63.3%
IN	35.7%	42.9%	42.9%	21.4%	28.6%	35.7%	50.0%	92.9%	92.9%	92.9%	78.6%	100.0%	85.7%	71.4%	64.3%	35.7%	28.6%	21.4%	35.7%	64.3%	50.0%	55.8%
LA	50.0%	50.0%	57.1%	85.7%	92.9%	100.0%	92.9%	57.1%	42.9%	35.7%	35.7%	50.0%	64.3%	64.3%	57.1%	85.7%	64.3%	57.1%	78.6%	64.3%	42.9%	63.3%
MA	85.7%	85.7%	85.7%	21.4%	21.4%	28.6%	28.6%	85.7%	78.6%	78.6%	78.6%	78.6%	92.9%	100.0%	100.0%	100.0%	100.0%	100.0%	85.7%	92.9%	57.1%	75.5%
MD	78.6%	85.7%	85.7%	14.3%	42.9%	50.0%	50.0%	85.7%	85.7%	85.7%	71.4%	78.6%	92.9%	78.6%	71.4%	35.7%	35.7%	28.6%	92.9%	100.0%	85.7%	68.4%
MI	64.3%	64.3%	64.3%	71.4%	92.9%	78.6%	78.6%	100.0%	85.7%	78.6%	57.1%	57.1%	100.0%	78.6%	78.6%	50.0%	42.9%	35.7%	28.6%	50.0%	28.6%	66.0%
NC	64.3%	57.1%	50.0%	28.6%	7.1%	28.6%	21.4%	92.9%	85.7%	85.7%	92.9%	100.0%	57.1%	50.0%	57.1%	35.7%	42.9%	35.7%	100.0%	42.9%	28.6%	55.4%
NJ	57.1%	57.1%	57.1%	14.3%	21.4%	21.4%	21.4%	100.0%	100.0%	92.9%	71.4%	64.3%	85.7%	78.6%	71.4%	42.9%	35.7%	28.6%	57.1%	64.3%	50.0%	56.8%
NY	28.6%	35.7%	42.9%	50.0%	50.0%	57.1%	64.3%	92.9%	64.3%	57.1%	50.0%	57.1%	92.9%	92.9%	85.7%	64.3%	57.1%	50.0%	78.6%	71.4%	21.4%	60.2%
PA	7.1%	7.1%	14.3%	35.7%	21.4%	14.3%	28.6%	71.4%	64.3%	71.4%	85.7%	100.0%	71.4%	71.4%	78.6%	42.9%	35.7%	35.7%	57.1%	57.1%	57.1%	49.0%
TX	7.1%	0.0%	0.0%	78.6%	50.0%	28.6%	21.4%	100.0%	85.7%	85.7%	100.0%	78.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	85.7%	71.4%	57.1%	69.0%
US	64.3%	71.4%	78.6%	64.3%	64.3%	78.6%	85.7%	92.9%	100.0%	100.0%	100.0%	100.0%	100.0%	92.9%	92.9%	100.0%	92.9%	85.7%	92.9%	78.6%	57.1%	85.4%
Avg	52.5%	53.4%	54.6%	35.3%	36.6%	39.9%	44.5%	90.8%	81.5%	79.0%	71.9%	76.9%	88.7%	79.8%	76.0%	67.7%	64.3%	59.7%	71.0%	67.7%	50.4%	63.9%

The table indicates the results are relatively poor. Of the 357 trials, only 50 instances met the stated target (14.0%). No state made the 95% target, and neither did any model. Only a total of 63.9% of the predictions fell within the 95% limit.

Table 3-Model Bias

	M0325	M0326	M0327	M0329	M0330	M0331	M0401	M0405	M0407	M0408	M0410	M0413	M0417	M0421	M0422	M0427	M0428	M0429	M0504	M0510	M0512	Avg
CA	-91.8%	-96.3%	-97.1%	-82.6%	-87.9%	-89.5%	-91.1%	7.1%	1.2%	24.6%	51.2%	92.2%	87.5%	98.3%	98.9%	100.0%	100.0%	100.0%	43.3%	15.7%	-19.9%	7.8%
CO	94.9%	100.0%	100.0%	-98.3%	-94.4%	-100.0%	-100.0%	94.6%	72.2%	84.8%	75.2%	98.6%	67.5%	86.4%	82.9%	-15.4%	3.7%	25.7%	68.3%	50.5%	40.1%	35.1%
CT	43.5%	53.6%	62.0%	97.1%	100.0%	93.5%	95.3%	-81.1%	-44.9%	-46.5%	-49.8%	-68.2%	0.5%	17.7%	3.4%	-17.4%	4.0%	25.4%	57.8%	-3.7%	-72.4%	12.9%
FL	88.6%	89.9%	90.6%	-27.6%	-0.9%	-46.6%	-52.6%	-100.0%	-100.0%	-100.0%	-100.0%	-97.8%	32.5%	39.1%	47.3%	39.9%	47.3%	50.9%	17.2%	-95.1%	-100.0%	-13.2%
GA	-31.7%	-43.2%	-51.8%	-44.0%	-32.7%	-49.7%	-63.7%	-88.8%	-96.8%	-99.1%	-98.6%	-95.4%	10.7%	-48.1%	-76.6%	-84.7%	-41.9%	-55.5%	-99.3%	-59.9%	53.6%	-57.0%
IL	24.8%	23.7%	21.2%	-21.9%	14.0%	-19.9%	-52.6%	-100.0%	18.4%	26.2%	100.0%	97.9%	58.6%	98.8%	98.4%	100.0%	100.0%	100.0%	23.2%	-2.1%	-44.0%	31.6%
IN	-100.0%	-100.0%	-100.0%	54.7%	57.5%	33.2%	45.2%	-100.0%	-28.7%	-32.7%	66.5%	39.7%	75.5%	64.2%	73.9%	77.5%	78.1%	78.2%	-76.8%	-86.3%	16.5%	6.5%
LA	-87.5%	-89.3%	-97.1%	-84.2%	-86.7%	-68.5%	-44.2%	97.2%	100.0%	100.0%	97.8%	100.0%	73.6%	90.6%	87.7%	95.8%	99.3%	99.3%	29.0%	-34.8%	-35.4%	21.1%
MA	-57.8%	-65.1%	-66.5%	55.5%	86.4%	-15.2%	23.9%	-100.0%	-58.3%	-57.0%	-21.3%	-45.3%	76.7%	91.5%	88.1%	33.5%	30.2%	33.1%	-54.4%	-14.7%	-82.9%	-5.7%
MD	86.9%	91.1%	94.5%	100.0%	95.4%	92.6%	53.5%	-100.0%	-32.6%	-32.3%	69.8%	83.1%	60.4%	60.1%	60.2%	73.8%	74.3%	75.2%	85.9%	85.5%	-91.4%	51.7%
MI	-98.7%	-98.9%	-100.0%	-43.9%	-64.9%	-87.1%	-82.3%	-34.9%	80.7%	88.9%	100.0%	98.1%	68.9%	97.5%	97.0%	100.0%	99.8%	99.8%	-94.3%	-30.4%	-61.8%	6.4%
NC	-94.9%	-96.5%	-97.4%	-95.3%	-100.0%	-93.4%	-94.2%	-98.4%	-92.8%	-92.8%	-55.6%	-4.0%	97.5%	100.0%	98.5%	100.0%	100.0%	100.0%	-46.0%	-10.5%	-61.7%	-25.6%
NJ	38.9%	42.6%	46.5%	95.8%	99.0%	95.5%	96.3%	-100.0%	9.4%	23.8%	84.7%	99.0%	55.7%	70.8%	74.0%	97.9%	98.1%	98.1%	-92.2%	-93.1%	-80.5%	41.0%
NY	100.0%	100.0%	100.0%	100.0%	98.1%	95.8%	93.6%	60.6%	79.6%	80.3%	81.1%	79.0%	-35.0%	33.4%	68.8%	88.3%	90.7%	90.3%	-57.1%	-34.8%	-100.0%	57.7%
PA	-57.0%	-67.4%	-67.9%	18.8%	27.7%	47.8%	29.9%	99.7%	97.0%	96.9%	83.4%	36.8%	86.8%	64.8%	55.1%	99.0%	99.2%	99.1%	-38.8%	-80.4%	-81.7%	30.9%
TX	-100.0%	-100.0%	-100.0%	-62.4%	-87.2%	-85.2%	-90.0%	-100.0%	-100.0%	-100.0%	-100.0%	-98.7%	-3.6%	-29.7%	-35.7%	-43.5%	29.0%	31.4%	-58.6%	4.9%	-53.9%	-61.1%
US	73.7%	68.4%	58.2%	63.6%	50.3%	-30.4%	-46.1%	-97.0%	16.5%	25.8%	81.7%	79.9%	58.4%	85.7%	87.6%	100.0%	100.0%	100.0%	-72.2%	-63.8%	-94.6%	30.8%
Avg	-9.9%	-11.0%	-12.0%	1.5%	4.3%	-13.4%	-16.4%	-43.6%	-4.7%	-0.5%	27.4%	29.1%	51.3%	60.1%	59.4%	55.6%	65.4%	67.7%	-21.5%	-26.6%	-51.2%	10.0%

To track the bias of the forecast we use a measure similar to the tracking signal but scaled on the interval [-1,1]. We define the bias (*B*) of the forecast as the ratio of the mean deviation to the mean absolute deviation.

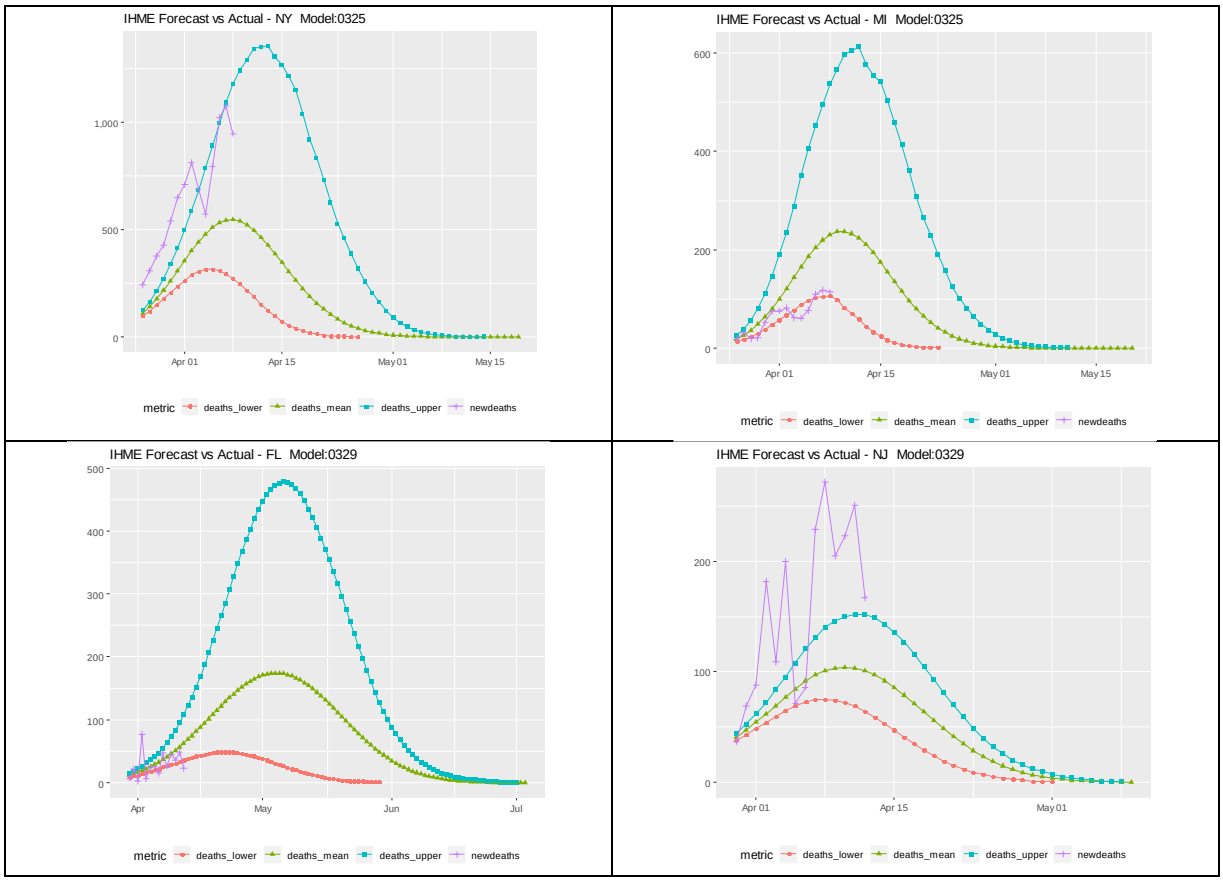
$$B = \frac{\sum_{i=1}^{W}\left(A_i - F_i\right)}{\sum_{i=1}^{W}\left|\left(A_i - F_i\right)\right|} \quad (1)$$

Where A_i is the observed actual value in period *i*, F_i is the forecast mean in period *i*, and *W* is the number of periods in the forecast window. If all actuals are above the forecasted value *B* will be positive 1, if all actuals are below the forecast then *B* will be negative 1. If the deviations are evenly weighted above and below the forecast, then the bias will be 0. Table 3 lists the bias for each model for the same states shown in Table 2. The table highlights any model that is 100% biased. Of the 357 predictions 51 (14.3%) are 100% biased, with 25 consistently over predicting and 26 consistently underpredicting.

Figure 7 illustrates some examples. The NY 03/25 model consistently under forecasts the death count with a 100% bias. Forecasts are above the confidence level for the first 8 days and are only in the 95% limit 28% of the time in the 14-day horizon. The MI 03/25 model is within the CI 64% of the time but are mostly below the mean forecast with a bias score -99%. The FL

03/29 model seems to have significant noise possibly due to batching/timing issue. This model is considerably less biased with a bias score of -27.6%. Only 42.9% of the observations are within the limits, with observations both above and below the limits. The final example, NJ 03/29, is the most problematic. It shows significant noise issues, with the two observations inside the bounds standing out as low side outliers. The data is overall significantly above the forecast and a smoothed model would likely have no observations within the bounds. Given the high side bias it is somewhat surprising the forecasts for NJ did not rise significantly for several iterations. As shown in Table 1 it was not until the 04/05 model that NJ projections rose significantly. They then dropped down again before rising again significantly, more than doubling, in the 05/04 model.

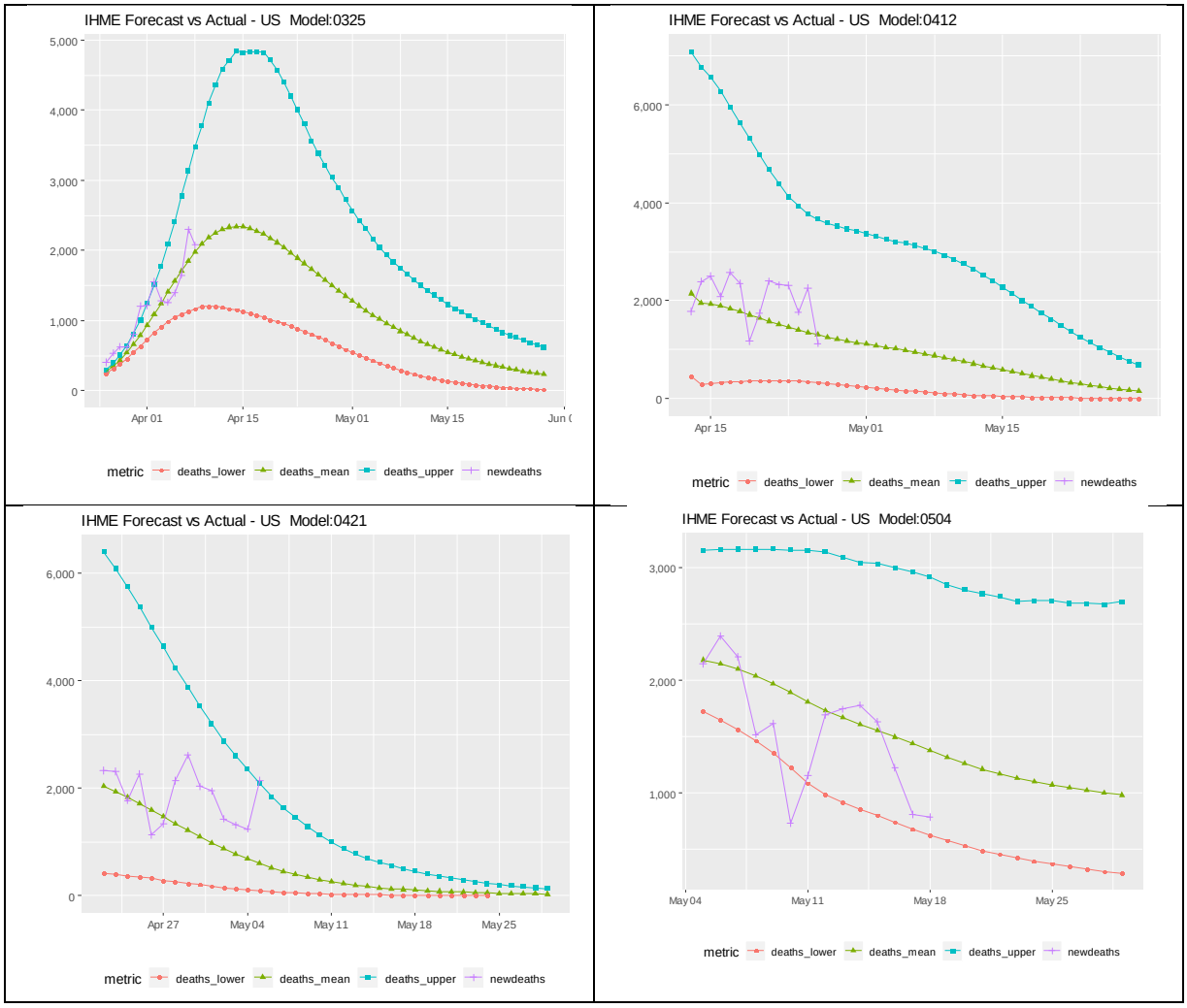
Figure 7- Various Forecast vs Actual Graphs



We next examine the forecast vs. actual performance at the national level. Figure 8 shows 14-day forecasts for four different iterations of the national model. In the early stage (3/25) model the actual deaths generally exceed the forecast and the model is 74% biased with only 64% of the predictions within the 95% interval. This is in the early stage of the outbreak when forecasts and actual deaths were increasing rapidly. The next three models are for the period where deaths are declining nationally. If we fast forward a few weeks to the 4/12 model forecast accuracy is much better. The 14-day forecast is all within the confidence bounds and quite close to the mean forecast. The model is however somewhat biased to underpredict the death count with a bias metric of 70.8%. Moving ahead a bit to the 4/21 model and the forecast is again relatively accurate, all but the last forecast is in the confidence bounds, though the model

continues to under forecast with a bias of nearly 86%. The last example is for the 5/4 model. Recall from Figure 3 and Table 1 that in this model the forecast has shifted sharply higher, more than double the 4/21 model. Consequently, the model is now biased to over predict deaths. The model has a bias measure of -72%, but all but 1 forecast is within the 95% interval which appears to possibly be a noise issue.

Figure 8-National Forecast vs Actual Graphs



In summary the day to day prediction accuracy for the model at the state level is relatively poor. The model does not come close to forecasting within its stated confidence intervals at the projected 95% confidence level. The model is however much better at the aggregate level than at the individual state levels. Over forecasting and under forecasting at the state level tend to somewhat cancel out. In terms of the criticism leveled at the model, the overly optimistic and overly pessimistic critiques are both valid depending on the particular forecast entity and model iteration. The criticism that the forecast confidence intervals are too wide, does not appear to be valid, given the number of predictions that fall outside the confidence bounds. It can be argued that the limits are in fact too narrow given the model's ability to accurately forecast day by day fatalities.

CUMMULATIVE DEATHS

While the ability to forecast hospital requirements and daily fatalities is very important, to many observers the most important numbers coming out of the model are the overall death predictions. These numbers, in particular the US total, have been the most analyzed and most scrutinized numbers reported.

At the time of the writing of this paper the model has not fully played out, that is we have not come to the end of the forecasting period. But what we can look at is the most current count of fatalities compared to the forecasted cumulative forecast on that date. We currently have actual fatalities reported through the end of May. As shown in Figure 1 many of the early version of the model had forecast a near zero death rate by then, a prediction which has not materialized. We begin with the United State in aggregate.

Figure 9- Comparison of Forecasts to Actual for US

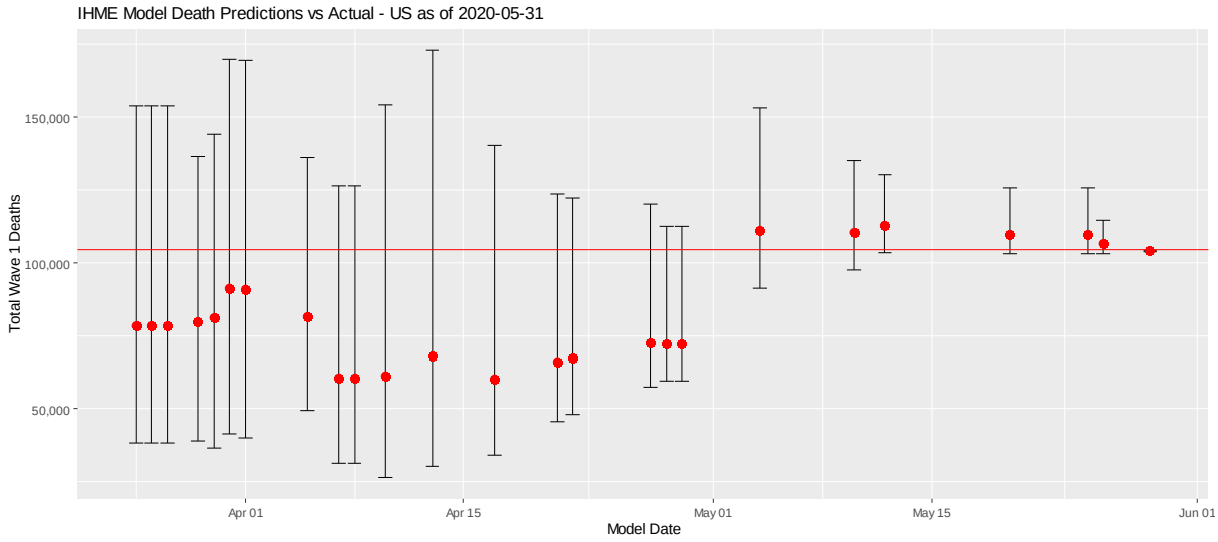


Figure 9 shows the range of deaths from each model projected and the actual deaths reported through May 31st. For the early models that is a forecast a little more than 2 months out. For the later models it is less than 2 weeks out, for the final model just 2 days out. The early models are also making projections well before the peak, while the later models are predicting during the down slope of the pandemic, as death rates are slowly declining. For both these reasons it would seem to make sense for the later models to be more accurate as well as more precise, i.e., smaller confidence intervals. The confidence intervals do become much smaller and the later stage models are more accurate, but the model does show some systemic bias. All models prior to May 1 under-predicted the death count. The models beginning in early May significantly raised the death count but all over-predicted the actual count. Table 4 lists key statistics for each model version at the aggregate US level for 5/25. It was the strong downward revision of the models in early April that created the least accurate predictions. The model from 4/17, 44 days out generated an error of more than 42%. With the benefit of hindsight this downward revision, cited by Dr. Birx, was premature and inaccurate.

Table 4- US Ending Death Predictions

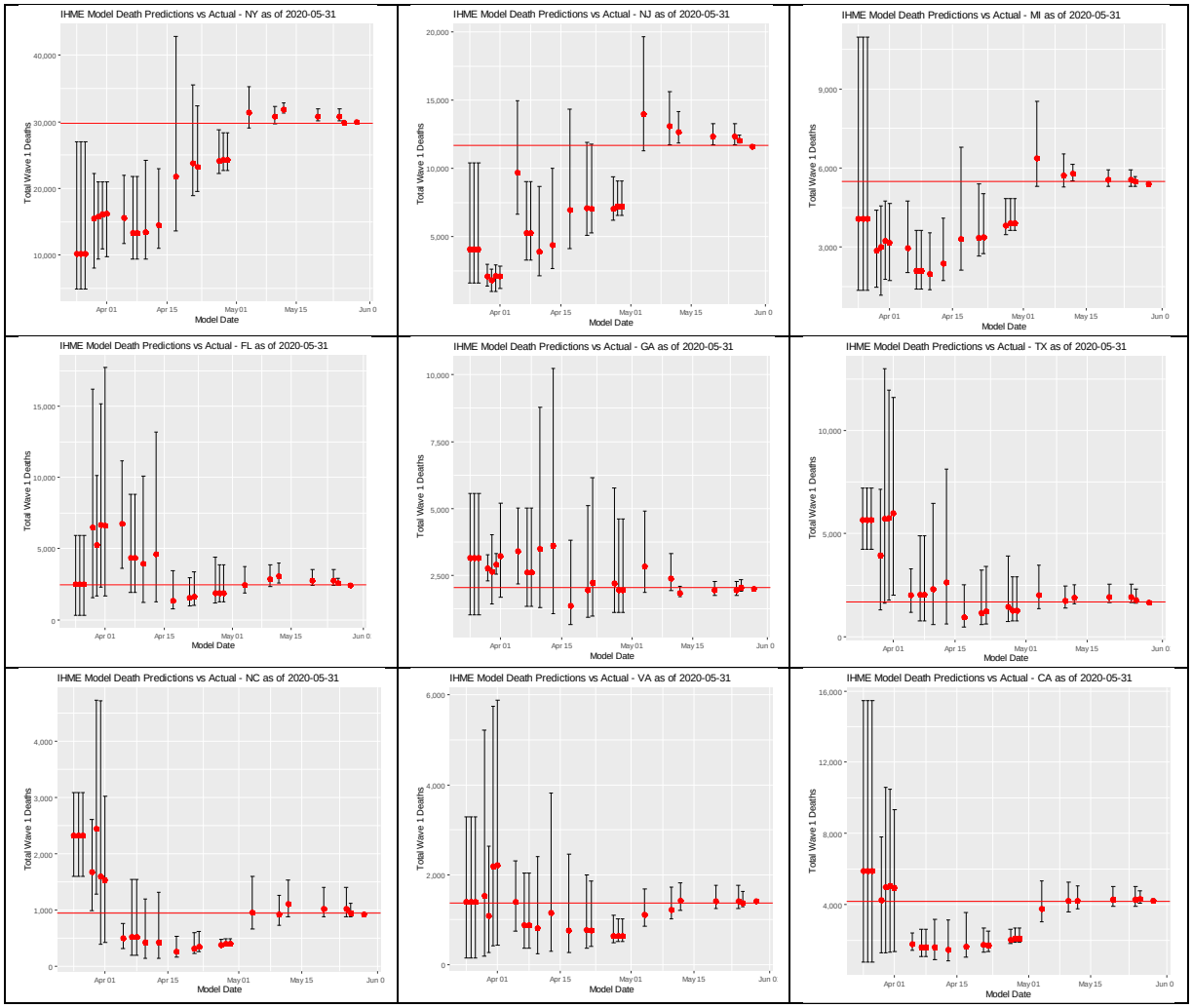
Model	Lower	Mid	Upper	Range	Actual Deaths	Error	ABS Error	Sq Errror	ABS Pct Error	In Interval	Forecast window
M0325	38,191	78,115	153,166	114,975	104,365	26,250	26,250	689,046,750	25.2%	100%	67
M0326	38,191	78,115	153,166	114,975	104,365	26,250	26,250	689,046,750	25.2%	100%	66
M0327	38,191	78,115	153,166	114,975	104,365	26,250	26,250	689,046,750	25.2%	100%	65
M0329	39,087	79,684	135,692	96,604	104,365	24,682	24,682	609,176,442	23.6%	100%	63
M0330	36,545	81,001	143,288	106,744	104,365	23,364	23,364	545,885,842	22.4%	100%	62
M0331	41,374	90,936	169,170	127,796	104,365	13,429	13,429	180,327,298	12.9%	100%	61
M0401	39,912	90,612	168,684	128,772	104,365	13,754	13,754	189,158,762	13.2%	100%	60
M0405	49,429	81,626	136,083	86,654	104,365	22,739	22,739	517,071,217	21.8%	100%	56
M0407	31,221	60,333	126,431	95,211	104,365	44,032	44,032	1,938,781,799	42.2%	100%	54
M0408	31,221	60,333	126,431	95,211	104,365	44,032	44,032	1,938,781,799	42.2%	100%	53
M0410	26,487	61,218	153,634	127,147	104,365	43,148	43,148	1,861,706,756	41.3%	100%	51
M0413	30,182	67,997	172,223	142,041	104,365	36,368	36,368	1,322,631,424	34.8%	100%	48
M0417	34,063	60,245	140,009	105,946	104,365	44,120	44,120	1,946,609,696	42.3%	100%	44
M0421	45,375	65,811	123,377	78,002	104,365	38,554	38,554	1,486,395,494	36.9%	100%	40
M0422	48,058	67,388	121,942	73,885	104,365	36,977	36,977	1,367,283,738	35.4%	100%	39
M0427	57,451	72,354	119,953	62,502	104,365	32,011	32,011	1,024,710,523	30.7%	100%	34
M0428	59,343	72,036	112,133	52,790	104,365	32,329	32,329	1,045,177,173	31.0%	100%	33
M0429	59,343	72,036	112,133	52,790	104,365	32,329	32,329	1,045,177,173	31.0%	100%	32
M0504	91,164	110,259	150,246	59,081	104,365	(5,894)	5,894	34,740,415	5.6%	100%	27
M0510	97,003	109,272	132,764	35,761	104,365	(4,907)	4,907	24,078,649	4.7%	100%	21
M0525	102,352	108,569	123,290	20,938	104,365	(4,204)	4,204	17,669,412	4.0%	100%	6
M0526	102,352	105,383	112,744	10,392	104,365	(1,018)	1,018	1,036,935	1.0%	100%	5
M0529	102,971	103,106	103,260	289	104,365	1,259	1,259	1,584,829	1.2%	0%	2
Average	53,891	80,632	136,651	82,760	104,365	23,733	25,126	833,266,332	24.1%	96%	43.0

The actual death count is within the 95% confidence interval for all but one iteration of the model, although actuals are quite close to the lower limit in later models. Interestingly, it is the final iteration of the model with just a 2 day forecast window, where the actual outcome is outside the interval. The uncertainty in the model, as expressed by the width of the confidence interval started high but became even larger in the mid-April time frame before declining significantly. As expected, the later models are fairly accurate given the short prediction interval. Overall the models exhibit a reasonably high error, with the Mean Absolute Percentage Error at 20.7%. Somewhat surprisingly the earliest models were not the least accurate. The two oldest models, forecasting more than 2 months out, were as accurate as the average model.

We now turn from the aggregate model to individual state by state forecasts. Figure 10 shows the various model predictions and the actual fatality levels for various states. Several patterns are apparent. For the first three states, NY, NJ and MI, the pre-May revision models significantly under forecast actual deaths, with many, but not all actuals outside the model's confidence interval. With the major upward revision in May the models consistently over forecast the death rate. The next three states, FL, GA and TX have a different pattern. Here the early models significantly over forecasted fatalities. Each was revised downward sharply in mid-April and from that point forward have been fairly accurate with relatively narrow confidence levels. These states all had a lower fatality rate than the first group. The last group, NC, VA, and CA had a sharp drop in the forecast in early April. The forecasts then steadily rose over

time. In the case of NC and CA, the final death count came in well below the original forecasts, while in VA the final forecasts are close to the original forecasts.

Figure 10- State Level Forecasts vs Actual



To make a more comprehensive assessment we look at the end point forecast for each state from each model and compare that to the actual reported death count, giving us one point per model per state. We calculate error rates for each prediction and then summarize those errors by model in Table 5 and by state in Table 6. We calculate the following summary error measures, the mean deviation (MD), the mean absolute deviation (MAD), the root mean square error (RMSE), the mean absolute percentage error (MAPE) and the proportion of the predictions within the confidence interval (Prop in CI).

$$MD = \frac{\sum_{i=1}^{n}(A_i - F_i)}{n} \tag{2}$$

$$MAD = \frac{\sum_{i=1}^{n} \left| \left(A_i - F_i \right) \right|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(A_i - F_i \right)^2} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| A_i - F_i \right|}{A_i} \quad (5)$$

Table 5- Long Term Prediction Error: Model Level

Model	Actual	Avg Forecast	MAD	RMSE	MAPE	Prop In CI	Forecast Window
M0325	104,229	78,115.4	1,290.0	3,203.0	220.7%	49.0%	67
M0326	104,229	78,115.4	1,290.0	3,203.0	220.7%	49.0%	66
M0327	104,229	78,115.4	1,290.0	3,203.0	220.7%	49.0%	65
M0329	104,229	79,683.7	1,396.6	2,837.1	267.0%	33.3%	63
M0330	104,229	81,000.7	1,374.0	2,873.8	234.7%	51.0%	62
M0331	104,229	90,936.1	1,441.1	2,910.6	246.9%	39.2%	61
M0401	104,229	90,611.6	1,394.8	2,812.5	241.4%	41.2%	60
M0405	104,229	81,626.0	1,001.1	2,343.0	167.7%	47.1%	56
M0407	104,229	60,333.5	1,080.4	2,747.4	108.1%	64.7%	54
M0408	104,229	60,333.5	1,080.4	2,747.4	108.1%	64.7%	53
M0410	104,229	61,217.8	1,145.0	2,814.3	89.1%	72.5%	51
M0413	104,229	67,996.5	1,137.3	2,651.4	68.2%	78.4%	48
M0417	104,229	60,244.9	884.3	1,712.8	77.2%	70.6%	44
M0421	104,229	65,810.9	779.7	1,468.0	73.8%	68.6%	40
M0422	104,229	67,388.2	761.4	1,473.2	64.1%	68.6%	39
M0427	104,229	72,253.7	663.9	1,336.2	49.0%	62.7%	34
M0428	104,229	71,717.0	651.9	1,294.0	39.2%	54.9%	33
M0429	104,229	71,717.0	651.9	1,294.0	39.2%	54.9%	32
M0504	104,229	110,259.1	242.8	488.6	18.4%	82.4%	27
M0510	104,229	109,271.8	139.5	308.7	11.6%	82.4%	21
M0512	104,229	111,473.2	178.2	386.3	15.0%	64.7%	19
M0520	104,229	108,568.9	107.3	221.1	10.4%	76.5%	11
M0525	104,229	108,568.9	107.3	221.1	10.4%	76.5%	6
M0526	104,229	105,383.3	47.3	84.1	7.1%	78.4%	5
M0529	104,229	103,106.2	36.8	62.1	3.2%	23.5%	2

The results on a model by model basis are not very good. Noting that this particular analysis weighs each state equally, small absolute errors in low impact state may create rather large percentage errors. None the less, on a state by state basis the early models were off in their

predictions for the end of May death count by more than 200%. Of the 51 forecasts made the early stage models had a best case 26 predictions (51%) land within the 95% confidence interval. As time progressed the epidemic peaked and begin to subside, and the forecast horizon became much shorter, the models became more accurate. In the mid-to late May models, where predictions are only for a few weeks, the MAPE dropped below 15%. But even with these short time frames the model never predicted more than 42 states (82%) within the 95% interval.

If we instead aggregate across the models, we can assess the predictive accuracy within an individual state. Table 6 presents results select states, while Table 8 in the appendix provides this data for all states.

Table 6- Long Term Prediction Error: State Level

State	Actual	Avg Forecast	MD	MAD	RMSE	MAPE	Prop In CI
CA	4,172	3,404.6	(767.4)	1,386.1	1,728.7	33.2%	56.0%
CO	1,445	1,126.4	(318.6)	528.6	627.1	36.6%	56.0%
CT	3,944	3,018.3	(925.7)	1,285.6	1,800.2	32.6%	68.0%
FL	2,451	3,394.6	943.6	1,316.7	1,945.9	53.7%	88.0%
GA	2,053	2,529.6	476.6	600.5	770.2	29.2%	84.0%
IL	5,390	3,185.1	(2,204.9)	2,348.2	2,728.6	43.6%	48.0%
LA	2,791	1,993.4	(797.6)	827.6	1,026.2	29.7%	52.0%
MA	6,846	5,094.7	(1,751.3)	2,145.6	2,821.7	31.3%	80.0%
MD	2,532	1,518.6	(1,013.4)	1,029.3	1,220.0	40.7%	60.0%
MI	5,491	3,892.8	(1,598.2)	1,707.9	2,034.3	31.1%	40.0%
NC	937	1,013.5	76.5	552.9	706.9	59.0%	48.0%
NJ	11,698	7,168.2	(4,529.8)	4,988.9	5,932.5	42.6%	24.0%
NY	29,784	20,833.1	(8,950.9)	9,445.9	11,662.8	31.7%	24.0%
PA	5,555	3,198.2	(2,356.8)	2,782.3	3,111.3	50.1%	12.0%
TX	1,675	2,770.5	1,095.5	1,313.7	2,060.7	78.4%	76.0%
VA	1,375	1,183.4	(191.6)	334.0	447.1	24.3%	88.0%

The results at the state level are not very good either. Consider the high impact states of NY and NJ. Less than 25% of the model predictions for the late May death count contained the actual value within the confidence interval. The MAPE for NY models is 31.7%, for NJ 42.6%. For several small, low impact states such as MT, HI, AK and WY the MAPE was greater than 450%. For no state did all the models contain the final death count in the confidence interval.

CONCLUSIONS

The objective of the IHME model, to make accurate forecasts of various quantities related to the Covid-19 pandemic, and to update those forecasts as new data became available and as conditions changed was a difficult one. The Covid-19 epidemic is unprecedented, so the situation contains many unknowns. Additionally, given the non-linear characteristics of a pandemic, small errors in a model can lead to very large errors. Finally, the data upon which the IHME modelers relied upon was poor, and in some cases may have been purposely mis-stated. Accurately predicting the impact of the Covid-19 pandemic was without a doubt an

extremely difficult task. With the benefit of hindsight, we can now point out all the deficiencies and inaccuracies of the predictions.

While the model can be said to have been reasonably accurate at the national level over the short term, it was far less accurate at lower levels of aggregation or over the long term. State level forecasts were consistently outside the confidence interval boundaries both in terms of the short-term daily counts, measured over a 14-day window, as well as for the long term measured to the last date on which actuals are available. The models were at times overly optimistic, and at other times overly pessimistic. While the widths of the confidence intervals were criticized in general for being too imprecise, the fact is that the intervals were clearly too narrow based on the model's overall predictive accuracy. From an external point of view, the 95% confidence levels provided with the estimates provided nothing close to that level of confidence. While the confidence ranges may have been far wider than desired, they were clearly not wide enough to provide the intended level of certainty.

In terms of usefulness to policy makers, the most significant issue may be the relative instability of the model's predictions. Large swings in forecasts occurred over short time frames that in many cases cannot be reasonably justified based on changes in the external environment. Recall the case of PA where over the course of two weeks projected fatalities started at 2023, dropped as low as 782, rose back to 2005 and then dropped again to 1707. Later in May the forecast would jump by a factor of more than 5 between April 28 and May 12. This level of instability is difficult to reconcile with policy level changes observable on the ground. The largest single change to the model occurred between April 27th and May 4th when the aggregate death forecast nearly doubled, jumping from 72.4K to 134.5K in less than a week. However, the change associated with the May 4th release was not based on changes in the external environment, it was based on changes to the underlying forecasting algorithm. It was in that release the model changed to a multi-stage hybrid model. The estimation update for this model release states "Overall, these *modeling improvements* have resulted in considerably higher projections of cumulative COVID-19 deaths through August, primarily due to longer peaks and slower declines for locations that have passed their peaks" (emphasis added).

The significant and rapid shifts in the model seem to indicate a model that is overly sensitive to changes in data and or method, like an overamplified control system prone to overshoot its target. In retrospect the policy of updating the model with such a high frequency and making changes that shifted predictions so significantly may not have been the best decision. Given that the earliest models, those issued in late March and early April, where more accurate than the models issued in mid to late April, the rapidly evolving set of models may have been counterproductive.

The model's predictions were also more accurate at the national level than they were at the individual state level. Even, in the states with the most significant outbreaks and therefore the most data to model, the state level predictions were often quite inaccurate. In states such NY and NJ, the major epicenters of the outbreak, the models consistently and significantly under forecast the fatalities through the end of April. In other states such as NC and CA, the models dramatically over forecasts for a period, then under forecast for a period, before finally settling it at a reasonable level in May, after the pandemic had peaked and was in decline.

Perhaps the most significant issue with this model is that it was overly ambitious and unrealistic in its goals. The idea that a model could be developed and implemented in a short time, for a highly contagious but poorly understood virus, that could accurately predict fatalities at a low level of geographic and temporal granularity, at the same time that massive public health

interventions were being rolled out, seems unrealistic. While we will likely never encounter a situation quite like this again, a note of caution seems justified. Modelers and policy maker's alike may be better off with less ambitious goals. They may want to consider developing models at higher levels of aggregation, with even broader levels of uncertainty, that they update in a more deliberate, controlled and less frequent fashion. In short, a little more humility as to what is possible may be in order.

Limitations

While our analysis is comprehensive it does have several limitations. First and foremost, it is a preliminary assessment conducted before the pandemic has fully played out. Additionally, by necessity it is focused on selected aspects of the model. We examine only the United States, ignoring all international projections. The IHME model is not a single model, it is an integrated model making forecasts for each state and the US as a whole, forecasting multiple variables in 25 different iterations of the model. It is not practical to examine every aspect of the model and we chose what we believe to be the most relevant items. We focus primarily on fatalities, only briefly looking at hospitalization metrics. Additional analysis of the hospitalization forecasts is warranted, but beyond the scope of this preliminary analysis. When examining daily forecast accuracy, we somewhat arbitrarily chose 14 days as the forecast period. While we found relatively poor forecast accuracy, a more complete analysis would examine how accuracy varies along different forecast horizons. Similarly, when examining cumulative forecasts, we focused on the end date projection, with the end date chosen based on an arbitrary deadline. A more complete analysis might examine intermediate forecasts. Finally, one key forecast metric we failed to examine is the forecasted peak of the pandemic. While this is a critical quantity, the stochastic nature of the actual data makes it difficult to determine when the peak actually occurred. While these limitations are real, we believe the analysis provided does give us a fair and objective sense of the model's predictive ability. We also believe further analysis at the end of the forecast period is warranted.

APPENDIX

Table 7-Mean Death Predictions for All States

		M0326	M0327	M0329	M0330	M0331	M0401	M0405	M0407	M0408	M0410	M0413	M0417	M0421	M0422	M0427	M0428	M0429	M0504	M0510	M0512	M0520	M0525	M0526	M0529	Average	Range	Net Change
AK	Alaska	154	154	129	149	151	148	315	131	131	39	25	24	17	16	15	11	11	5	5	4	3	3	3	10	69	312	-144
AL	Alabama	1,155	1,155	1,732	1,173	7,334	5,516	923	634	634	431	351	295	290	306	309	294	294	2,308	1,554	796	1,208	1,208	882	833	1,317	7,044	-323
AR	Arkansas	707	707	762	729	562	619	297	230	230	199	195	158	131	126	164	149	149	171	213	116	115	115	140	608	316	646	-99
AZ	Arizona	1,687	1,687	1,613	1,575	1,380	1,387	571	775	775	974	1,005	267	480	583	767	734	734	1,043	2,987	2,871	6,174	6,174	4,335	2,665	1,802	5,907	978
CA	California	6,109	6,109	4,306	5,086	5,161	5,068	1,783	1,611	1,611	1,616	1,483	1,658	1,743	1,719	2,035	2,104	2,104	4,666	6,086	5,832	6,842	6,842	7,558	12,951	4,253	11,468	6,843
CO	Colorado	940	940	2,151	1,683	2,135	2,154	302	457	457	540	456	715	649	720	1,052	1,039	1,039	1,336	1,676	1,641	2,281	2,281	2,238	3,457	1,347	3,155	2,517
CT	Connecticut	774	774	378	414	1,092	1,144	5,474	4,003	4,003	4,614	5,426	2,732	2,884	3,006	3,367	3,315	3,315	4,047	4,575	5,263	4,860	4,860	4,548	4,251	3,296	5,096	3,478
DC	District of Columbia	132	132	132	384	220	218	58	84	84	75	117	170	213	228	255	249	249	556	536	648	577	577	540	698	297	640	566
DE	Delaware	228	228	827	240	163	166	50	39	39	43	74	143	136	146	157	161	161	542	413	416	540	540	521	413	266	788	185
FL	Florida	3,342	3,342	6,766	5,568	6,937	6,897	6,770	4,357	4,357	3,999	4,748	1,363	1,537	1,620	1,921	1,898	1,898	3,971	5,440	5,819	4,721	4,721	3,900	3,707	4,150	5,573	366
GA	Georgia	3,165	3,165	2,777	2,648	2,916	3,232	3,413	2,631	2,631	3,564	3,718	1,369	1,981	2,254	2,259	2,009	2,009	4,913	3,596	2,062	2,381	2,381	2,544	2,759	2,766	3,544	-405
HI	Hawaii	352	352	390	374	376	372	155	112	112	101	82	38	26	20	23	20	20	24	19	18	19	19	17	18	127	373	-335
IA	Iowa	742	742	138	777	1,367	1,488	420	565	565	744	618	625	442	365	330	302	302	1,513	915	710	1,043	1,043	1,146	1,040	747	1,375	297
ID	Idaho	388	388	84	459	397	397	69	57	57	61	73	63	61	64	66	68	68	83	75	73	81	81	88	98	142	402	-290
IL	Illinois	2,454	2,454	2,360	2,326	2,789	3,386	3,629	1,588	1,588	777	1,248	2,259	2,093	2,138	2,292	2,337	2,337	6,353	7,395	7,830	8,781	8,781	7,628	6,371	3,800	8,004	3,917
IN	Indiana	2,440	2,440	940	906	1,083	1,160	1,681	857	857	606	860	903	931	971	1,018	1,041	1,041	6,248	4,091	2,429	2,397	2,397	2,567	2,350	1,759	5,641	-90
KS	Kansas	669	669	687	687	627	640	265	299	299	426	555	187	273	319	354	307	307	264	411	421	327	327	235	227	408	500	-442
KY	Kentucky	585	585	605	1,162	815	821	1,750	1,017	1,017	933	1,067	407	466	529	517	421	421	537	703	728	508	508	480	540	713	1,343	-45
LA	Louisiana	2,081	2,081	1,897	1,978	1,934	1,834	746	946	946	1,125	1,141	1,685	1,704	1,780	2,033	2,026	2,026	2,822	3,233	3,169	3,141	3,141	3,119	2,947	2,064	2,487	866
MA	Massachusetts	2,231	2,231	1,782	1,507	2,357	2,382	8,254	5,625	5,625	6,739	8,220	3,236	3,898	4,243	5,633	5,634	5,634	7,697	7,545	9,629	8,949	8,949	8,232	9,003	5,635	8,123	6,772
MD	Maryland	858	858	1,171	1,700	1,691	1,766	2,326	1,094	1,094	598	596	914	1,112	1,171	1,197	1,245	1,245	2,190	2,606	3,799	4,363	4,363	3,510	2,883	1,848	3,767	2,026
ME	Maine	334	334	3,013	373	368	364	115	108	108	61	63	51	53	51	67	63	63	85	92	75	80	80	87	167	261	2,962	-166
MI	Michigan	4,061	4,061	2,862	3,007	3,235	3,169	2,963	2,103	2,103	1,977	2,373	3,304	3,361	3,379	3,818	3,920	3,920	7,080	6,217	6,621	6,043	6,043	5,801	5,986	4,059	5,104	1,925
MN	Minnesota	1,281	1,281	2,146	1,039	1,061	932	625	456	456	442	656	195	301	360	728	741	741	2,183	2,073	2,049	1,814	1,814	1,611	1,712	1,112	1,988	431
MO	Missouri	2,977	2,977	1,055	1,219	1,283	1,290	352	548	548	1,161	1,713	362	386	438	638	470	470	1,243	1,984	2,348	1,506	1,506	1,022	996	1,187	2,625	-1,981
MS	Mississippi	675	675	2,292	1,223	1,098	918	237	301	301	407	371	369	399	400	380	307	307	483	1,236	1,508	986	986	836	947	735	2,055	272
MT	Montana	251	251	282	268	270	266	22	81	81	50	22	17	14	18	27	20	20	13	14	14	12	12	11	16	85	271	-235
NC	North Carolina	2,411	2,411	1,721	2,538	1,597	1,534	496	522	522	416	415	251	310	342	375	394	394	1,169	1,190	4,413	2,524	2,524	1,397	1,505	1,307	4,161	-907
ND	North Dakota	163	163	171	168	171	169	677	519	519	369	32	149	277	356	259	136	136	38	81	199	132	132	120	96	218	645	-68
NE	Nebraska	437	437	424	442	448	447	459	289	289	273	281	127	245	347	381	266	266	380	218	184	243	243	331	475	330	347	38
NH	New Hampshire	319	319	351	337	338	331	32	66	66	60	79	55	70	70	72	71	71	168	228	424	315	315	377	350	203	392	32
NJ	New Jersey	4,109	4,109	2,096	1,844	2,129	2,117	9,691	5,277	5,277	3,915	4,407	6,952	7,116	7,058	7,051	7,246	7,246	16,044	14,752	14,692	14,070	14,070	13,654	12,577	7,813	14,200	8,468
NM	New Mexico	513	513	514	529	526	525	603	394	394	192	155	81	105	109	210	196	196	430	382	378	520	520	577	556	380	522	43
NV	Nevada	801	801	568	506	921	799	916	446	446	254	272	257	240	247	259	258	258	480	511	529	610	610	581	692	511	681	-109
NY	New York	10,244	10,244	15,546	15,788	16,090	16,261	15,618	13,307	13,307	13,463	14,542	21,812	23,741	23,232	24,088	24,314	24,314	32,132	31,621	34,068	32,136	32,136	30,823	31,295	21,672	23,825	21,051
OH	Ohio	2,733	2,733	1,203	1,672	1,886	1,898	544	489	489	489	482	716	771	808	836	857	857	2,552	2,960	2,844	2,963	2,963	3,160	3,227	1,672	2,745	494
OK	Oklahoma	898	898	1,149	1,100	1,312	1,419	813	697	697	965	929	359	306	335	271	271	271	494	470	457	368	368	359	361	649	1,148	-537
OR	Oregon	584	584	470	566	551	558	171	172	172	203	260	131	120	122	125	128	128	193	204	321	231	231	197	177	275	464	-407
PA	Pennsylvania	3,094	3,094	1,579	1,574	1,562	2,023	782	969	969	1,503	2,005	1,707	2,323	2,770	2,318	2,400	2,400	8,607	10,742	12,420	9,623	9,623	7,680	6,418	4,091	11,638	3,324
RI	Rhode Island	245	245	306	259	265	259	966	848	848	871	984	438	578	611	594	468	468	485	675	922	776	776	881	1,160	622	915	915
SC	South Carolina	768	768	1,043	1,034	1,060	1,095	442	470	470	572	680	217	261	283	351	361	361	1,112	744	469	485	485	502	734	615	895	-35
SD	South Dakota	201	201	204	203	206	204	178	189	189	357	181	94	98	93	45	17	17	279	167	137	126	126	95	81	154	339	-120
TN	Tennessee	1,551	1,551	1,067	4,985	3,259	3,422	587	617	617	612	481	231	235	233	248	227	227	563	507	323	489	489	509	734	990	4,759	-817
TX	Texas	5,847	5,847	4,150	6,029	6,128	6,392	2,025	2,042	2,042	2,350	2,704	957	1,166	1,241	1,447	1,288	1,288	3,632	2,567	3,093	3,710	3,710	2,985	2,025	3,111	5,434	-3,822
UT	Utah	619	619	502	582	586	580	186	163	163	188	179	202	223	241	422	317	317	86	150	228	150	150	146	220	301	533	-400
VA	Virginia	1,543	1,543	2,054	1,897	3,073	3,152	1,402	891	891	830	1,188	763	771	762	634	638	638	1,198	1,396	1,944	1,980	1,980	1,956	2,576	1,487	2,518	1,033
VT	Vermont	386	386	95	69	58	58	35	47	47	48	36	40	43	46	54	53	53	54	52	51	49	49	50	57	80	351	-329
WA	Washington	1,429	1,429	1,670	1,621	1,233	978	632	700	700	842	855	694	779	813	860	877	877	1,159	1,120	1,193	1,262	1,262	1,176	1,200	1,057	1,038	-230
WI	Wisconsin	853	853	1,309	935	926	951	644	424	424	357	338	302	340	356	331	336	336	717	673	791	734	734	697	799	632	1,007	-54
WV	West Virginia	460	460	543	496	495	487	182	104	104	74	70	22	35	39	38	41	41	121	79	68	107	107	109	107	183	522	-353
WY	Wyoming	136	136	132	143	142	140	119	67	67	43	34	243	243	161	64	36	36	6	6	5	4	4	9	38	84	239	239
US	**US**	**81,114**	**81,114**	**82,141**	**83,967**	**93,765**	**93,531**	**81,766**	**60,415**	**60,415**	**61,545**	**68,841**	**60,308**	**65,976**	**67,641**	**72,860**	**72,433**	**72,433**	**134,475**	**137,184**	**147,040**	**143,357**	**143,357**	**131,967**	**135,109**	**93,031**	**86,732**	**53,995**

Table 8 - Long Range Prediction Error All States

State	Actual	Avg Forecast	MD	MAD	RMSE	MAPE	Prop In CI
AK	10	70.8	60.8	63.8	99.7	637.6%	32.0%
AL	630	1,140.2	510.2	710.0	1,694.0	112.7%	72.0%
AR	133	299.2	166.2	175.0	288.5	131.5%	60.0%
AZ	907	1,025.3	118.3	353.1	422.2	38.9%	76.0%
CA	4,172	3,404.6	(767.4)	1,386.1	1,728.7	33.2%	56.0%
CO	1,445	1,126.4	(318.6)	528.6	627.1	36.6%	56.0%
CT	3,944	3,018.3	(925.7)	1,285.6	1,800.2	32.6%	68.0%
DC	466	261.1	(204.9)	216.5	258.3	46.5%	44.0%
DE	366	232.1	(133.9)	179.2	213.8	48.9%	32.0%
FL	2,451	3,394.6	943.6	1,316.7	1,945.9	53.7%	88.0%
GA	2,053	2,529.6	476.6	600.5	770.2	29.2%	84.0%
HI	17	130.7	113.7	113.7	180.0	668.8%	20.0%
IA	535	589.2	54.2	182.0	292.4	34.0%	88.0%
ID	82	147.6	65.6	82.8	151.8	100.9%	68.0%
IL	5,390	3,185.1	(2,204.9)	2,348.2	2,728.6	43.6%	48.0%
IN	2,134	1,519.5	(614.5)	823.3	950.4	38.6%	56.0%
KS	215	386.5	171.5	174.2	246.0	81.0%	56.0%
KY	431	625.6	194.6	202.5	358.6	47.0%	92.0%
LA	2,791	1,993.4	(797.6)	827.6	1,026.2	29.7%	52.0%
MA	6,846	5,094.7	(1,751.3)	2,145.6	2,821.7	31.3%	80.0%
MD	2,532	1,518.6	(1,013.4)	1,029.3	1,220.0	40.7%	60.0%
ME	89	251.6	162.6	187.4	572.4	210.6%	52.0%
MI	5,491	3,892.8	(1,598.2)	1,707.9	2,034.3	31.1%	40.0%
MN	1,050	904.8	(145.2)	339.3	442.4	32.3%	80.0%
MO	776	1,001.8	225.8	459.6	810.9	59.2%	88.0%
MS	734	644.9	(89.1)	309.8	432.4	42.2%	72.0%
MT	17	89.1	72.1	74.5	126.0	438.4%	44.0%
NC	937	1,013.5	76.5	552.9	706.9	59.0%	48.0%
ND	61	186.5	125.5	130.2	201.9	213.5%	60.0%
NE	170	286.1	116.1	124.9	157.5	73.5%	68.0%
NH	245	181.7	(63.3)	109.8	129.2	44.8%	60.0%
NJ	11,698	7,168.2	(4,529.8)	4,988.9	5,932.5	42.6%	24.0%
NM	356	340.8	(15.2)	123.3	147.9	34.6%	88.0%
NV	417	483.3	66.3	170.0	231.1	40.8%	80.0%
NY	29,784	20,833.1	(8,950.9)	9,445.9	11,662.8	31.7%	24.0%
OH	2,155	1,492.4	(662.6)	806.6	1,033.9	37.4%	52.0%
OK	334	637.6	303.6	322.8	467.8	96.6%	56.0%
OR	153	262.1	109.1	122.2	196.8	79.9%	68.0%
PA	5,555	3,198.2	(2,356.8)	2,782.3	3,111.3	50.1%	12.0%
RI	718	551.3	(166.7)	237.0	284.4	33.0%	68.0%
SC	494	573.7	79.7	194.0	261.8	39.3%	72.0%
SD	62	130.0	68.0	77.8	103.8	125.5%	60.0%
TN	364	963.7	599.7	667.4	1,322.6	183.4%	52.0%
TX	1,675	2,770.5	1,095.5	1,313.7	2,060.7	78.4%	76.0%
UT	113	282.1	169.1	175.1	249.0	154.9%	68.0%
VA	1,375	1,183.4	(191.6)	334.0	447.1	24.3%	88.0%
VT	55	91.9	36.9	47.3	115.1	85.9%	60.0%
WA	1,118	1,031.2	(86.8)	227.8	276.6	20.4%	68.0%
WI	592	552.8	(39.2)	179.5	238.9	30.3%	76.0%
WV	75	179.2	104.2	124.9	204.5	166.6%	44.0%
WY	16	83.0	67.0	72.1	97.3	450.6%	32.0%

REFERENCES

Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics, 25*(4), 539-573. doi:10.1002/jae.1159

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting, 27*(3), 822-844. doi:https://doi.org/10.1016/j.ijforecast.2010.04.009

Azad, A. (2020). An influential model projects coronavirus deaths will stop this summer, but experts are skeptical. Retrieved from https://www.fox10tv.com/news/coronavirus/an-influential-model-projects-coronavirus-deaths-will-stop-this-summer-but-experts-are-skeptical/article_f308b3c2-9752-5367-9d06-fd078e309ee0.html

Begley, S. (2020). *Influential Covid-19 model uses flawed methods and shouldn't guide U.S. policies, critics say*. Retrieved from https://www.statnews.com/2020/04/17/influential-covid-19-model-uses-flawed-methods-shouldnt-guide-policies-critics-say/

Boice, J. (2020). Best-Case And Worst-Case Coronavirus Forecasts Are Very Far Apart. *FiveThirtyEight*. Retrieved from https://fivethirtyeight.com/features/best-case-and-worst-case-coronavirus-forecasts-are-very-far-apart/

Carvajal, N. (2020). Birx says drop in US death projection is due to Americans changing their behavior through social distancing. Retrieved from https://www.cnn.com/2020/04/08/politics/deborah-birx-social-distancing-models/index.html

Chernenko, S., Schwarz, K., & Wright, J. (2004). The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk. *SSRN Electronic Journal*. doi:10.2139/ssrn.560386

Chikina, M., & Pegden, W. (2020). A call to honesty in pandemic modeling. *Medium*. Retrieved from https://medium.com/@wpegden/a-call-to-honesty-in-pandemic-modeling-5c156686a64b

Chinn, M., & Coibion, O. (2009). The Predictive Content of Commodity Futures. *Journal of Futures Markets, 34*. doi:10.2139/ssrn.1490043

Chowell, G., Sattenspiel, L., Bansal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of life reviews, 18*, 66-97. doi:10.1016/j.plrev.2016.07.005

CovidTracking. (2020). Covid Tracking Project. Retrieved from https://covidtracking.com/

Dimitrov, N. B., & Meyers, L. A. (2010). Mathematical Approaches to Infectious Disease Prediction and Control *Risk and Optimization in an Uncertain World* (pp. 1-25).

Fildes, R., & Makridakis, S. (1995). The Impact of Empirical Accuracy Studies on Time Series Analysis and Forecasting. *International Statistical Review / Revue Internationale de Statistique, 63*(3), 289-308. doi:10.2307/1403481

Fumento, M. (2020). *After Repeated Failures, It's Time To Permanently Dump Epidemic Models*. Retrieved from https://issuesinsights.com/2020/04/18/after-repeated-failures-its-time-to-permanently-dump-epidemic-models/

GitHub. (2020). Johns Hopkins Coronoavirus Time Series data. Retrieved from https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

Hatmaker, T. (2020). Fauci: US can expect more than 100,000 COVID-19 deaths, millions of cases. Retrieved from techcrunch.com website: https://techcrunch.com/2020/03/29/fauci-how-many-coronavirus-deaths-in-us-estimate/

IHME. (2020). Health Data.org. Retrieved from https://covid19.healthdata.org/projections

Ioannidis, J. P. (2020). A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data. *STAT*. Retrieved from https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/

Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA*, *323*(19), 1893-1894. doi:10.1001/jama.2020.6585

Johns_Hopkins. (2020). Coronavirsu Resource Center. Retrieved from https://coronavirus.jhu.edu/

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451-476. doi:https://doi.org/10.1016/S0169-2070(00)00057-1

Marchant, R., Rosen, N. I. S. a. O., Tanner, M. A., & Cripps, S. (2020). *Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions*. Retrieved from https://arxiv.org/abs/2004.04734

Murray, C. J. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv*, 2020.2003.2027.20043752. doi:10.1101/2020.03.27.20043752

Ord, K., Hibon, M., & Makridakis, S. (2000). The M3-Competition. *International Journal of Forecasting*, *16*(4), 433-436. doi:https://doi.org/10.1016/S0169-2070(00)00078-9

Piper, K. (2020). *This coronavirus model keeps being wrong. Why are we still listening to it?* Retrieved from https://www.vox.com/future-perfect/2020/5/2/21241261/coronavirus-modeling-us-deaths-ihme-pandemic

Robbins, T. R. (2020). *Real Time Tracking of a Global Pandemic: Data Issues and Implications*. Submitted to DSI Annual Meeting 2020.

Sunstein, C. R. (2006). *Infotopia: How Many Minds Produce Knowledge*.

Tetlock, P. E. (2017). *Expert Political Judgment, How Good is it? How can we tell?*

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*.

Wan, W., Dawsey, J., Parker, A., & Achenbach, J. (2020). Experts and Trump's advisers doubt White House's 240,000 coronavirus deaths estimate. *The Washington Post*. Retrieved from https://www.washingtonpost.com/health/2020/04/02/experts-trumps-advisers-doubt-white-houses-240000-coronavirus-deaths-estimate/

Wan, W., & Johnson, C. Y. (2020, 4/7/2020). America's most influential coronavirus model just revised its estimates downward. But not every model agrees. *The Washington Post*. Retrieved from https://www.washingtonpost.com/health/2020/04/06/americas-most-influential-coronavirus-model-just-revised-its-estimates-downward-not-every-model-agrees/

DECISION SCIENCES INSTUTUTE

An Anomaly in Goodman and Kruskal's Lambda Measure of Association in 2x2 Cross-Tabulations

John R. Dickinson
University of Windsor
Email: MExperiences@bell.net

ABSTRACT

With a finding of statistical significance in the common chi-square test of independence, i.e., that two variables are related, the degree to which they are related is often of interest. Goodman & Kruskal's lambda is a well-known statistic that describes that degree of association. Lambda, though, is subject to an anomaly that has been largely unrecognized in research methods textbooks. The present study demonstrates the anomaly in the popular 2x2 cross-tabulation with the objective of reviving its recognition.

KEYWORDS: Goodman & Kruskal lambda, Cross-tabulation, Chi-square

Full text and references available upon request.

DECISION SCIENCES INSTITUTE

Empirical Examination of Nudge Theory among Students of University

Sanjana Srivastava
Ahmedabad University
Email: sanjana.s1@ahduni.edu.in

Siddhartha Saxena
Ahmedabad University
Email: Siddhartha.saxena@ahduni.edu.in

ABSTRACT

This research talks about influencing and change in behaviour, further developing it into a pattern made by right decision. An individual creates meaning and behaves in a particular manner based on the societal meaning imbibed and interpreted by the brain. The rational and irrational decisions so taken are influenced by the system 1 part of the brain which influences the system 2 part of brain and decision making which are said to be rational decision. An experiment was conducted using Nudge theory to experiment the change in a particular behaviour using signage as a medium in the university premises.

KEYWORDS: Nudge theory, Decision making, Signage, Behaviour change, System 1 & System 2 brain

INTRODUCTION

"A growing body of evidence demonstrates that behavioural science insights – research findings from fields such as behavioural economics and psychology about how people make decisions and act on them – can be used to design government policies to better serve the American people."

- ***Obama, 2015***

Behavioural Economics, an experimental study about human decision making based on biases to Bounded rationality, states that people are not always in the want of optimization which means maximizing the output with minimum input. People are always influenced by their surroundings and do not always make rational decisions. To understand the deeper meaning to the study Nudge, (Sunstein & Thaler, 2008) a simple model of behavioural economics, will help in understanding the use of the concept in marketing, policy making, healthcare, education, charity, and the list goes on, basically knowing how to influence decision making. It is important to design choice architecture which will influence the decision-making pattern as the majority of the decision are based on irrationality, contradicting the general assumption among all. These decisions and design architecture for decision-making pattern are composed of the organised systems developed from the community which are assigned meaning grounded with one's idea influenced by the meaning derived from the symbolic meaning assigned by the society (Stryker, 2008).

Symbolic interactionism is the major concepts that aid in carrying forward the decisions by interacting with each other. The interactionism is based on two factors: strong conceptual from

the culture (Hays, 1977; Howard, 1981) and study of culture involve both qualitative and quantitative research (Thomas & Znaniecki, 1918-20). The frame of reference begins with individuals when interacting with each other provide a symbolic meaning and these meaning then are followed as behaviour pattern in the society (Burgess, 1926; Handel, 1985). Symbolic interactionism is formed when people shape their behaviour according to the societal needs and understanding which is derived from the concept of Libertarian paternalism. The latter concept allows people to be free to make their own choices, but their choices are influenced by the set pattern or the ideal state of thinking which have been moulded throughout their lives. Libertarian paternalism leads a person in becoming two sides of the same coin, where a person experiences two expressions and behaviour when they are along with society and another of their own. Blummer (1969), stated that society emerges from “joint action” a combination of what is rightly agreed upon in the society and people behave in a particular pattern to be a part of the society.

These irrational rationalities can be intervened through positive and negative reinforcements where it has been taken into consideration that positive reinforcements have a higher impact in the nonconscious brain of people. An experiment will be carried out in the thesis to have an empirical study of nudge theory to understand the impact among students and change in the environment of the canteen where a change of behaviour can be experienced for picking up their plates after consuming snacks through signage which will be displayed in the premises. The objective of the study uncovers the psychological processes behind the economic judgments we take. The objective of the study will help us in: Understanding the meaning of Behavioural Economics, how it helps people to make decisions based on them and positive reinforcement is a stronger mechanism than negative reinforcement in people.

LITERATURE REVIEW

Dan Ariely (2008) in his book Predictably Irrational says that human beings though categorized as a rational and calculative human being has a tendency to make everyday mistakes by overpaying, underestimating or procrastinating based on their emotions. Any judgment and decisions taken are have been shaped through psychological behaviours influenced by cognitive, emotional, economic and social factors (Samson, 2014). Human behaves in different pattern and it is important to understand how they feel, look or act towards a particular situation. It becomes essential to conduct an experimental economic test to understand such behaviours which help in figuring out on multidisciplinary approach in the various fields like marketing, policy making, healthcare, education. One must understand the economic concepts from a psychological and sociological perspective (Ariely, D., 2008; Thaler, R., 2015) to understand the meanings assigned to monetary values. Economists believe that life and things attached to it are defined by the “Willingness to Pay”. A person's tendency for willingness to pay can be defined by the utility obtained or is perceived to provide, a parameter to define good and bad, and measurement to move to substitute options (Simmel, G., 2004; Zelizar VA., 2007; Foley, D., 2009). It is required to understand emotions and actions taken by human which may be classified as senseless, random and misguided behaviour based on predictably irrational behaviour (Thaler & Sunstein, 2008; Ariely, 2008)

Neo-Classical Economics

Economic theories are based on economic development and cost-benefit analysis (Little, 1982; Ray, 1984). Economic in short studies the resource allocation, social goals, trade-offs and utility

based on the income distribution. The limitations of economics lead to the development of neoclassical economics. It took into consideration social and income distribution for cost-benefit analysis and maximizing the utility (D. Pearce, 1980; Squire & Van der Tak, 1981; Dasgupta, 1982). All neoclassical economic theories put in the centre that all people make a rational decision which was later criticised by Milton Friedman (1970) who believed that it is not to be assumed that humans always make a rational decision. An experiment conducted by Steven Pinker (2003) suggested that a person decided to go on a diet and goes for dinner and the waiter offers a small dessert and instead of avoiding dessert, you tend to order one. So, people tend to make decisions and solve problems through their system 1 brain or the nonconscious brain. Thus, moving beyond the singularity approach of rational choice making to detailed mapping of behaviour based on various contextual factors (Gigerenzer&Selten, 2001). To make the decisions distinguished and know the realism of economic model attention to decision process needs to be taken care of.

Behavioural Economics: Past, Present, Future

It is believed in the world of economics that people make an optimal and logical decision which provides the maximum utility and satisfaction under various conditions. This ideal situation is possible when people are not affected by their surroundings or emotions (Holt &Laury, 2002). Herbert Simon (2000) in his economic model about realistic assumptions to build an economic model and understand the cause and effect of it, actual human behaviour and institutional framework need to be considered (Altmen, 2012). The institutional framework consists of various aspects of peer pressure, religions, differences, gender, legal and past behaviour.Research shows that people have a high preference for pro-social preference, more weight to piece weight incentive over relative rate incentive to avoid negative consequences with peers (Bandiera, Barankay, and Rasul 2005). This portrays that people are capable of taking irrational decisions. Behavioural Economics studies the social, cultural, economic, cognitive and psychological factors (Levingson & peng, 2006; Kahnemen, 2003; Bertrand et al., 2004).

People are said to display inconsistent behaviour in economic studies as they are driven by the course of cognitive bias, emotions and social influence (Thaler & Sunstein, 2009). People thus work on thinking by two brains which are System 1, the Automatic System and System 2 also known as Reflective System. Understanding these systems more on the depth it is known that System 1 brain (Automatic System) is the Unconscious decision-maker of the brain which is fast in taking a decision based on the associations taking place. Similarly, for System 2 brain (Reflective System) it is more about taking rational decisions that can be judged based on the time taken more than 600 milliseconds. These decisions are controlled and deductive in nature and the decision-maker is completely aware of the rationale.

Human Vs. Econs

Econs is the definition of the people as per economic concepts where people are rational decision-maker and is driven by their rational brain after understanding the utility of the choice. On the other hand, behavioural researchers accepted the concept that people are making a decision and are driven by their irrational brain (Kahenman, 2013).In the Nudge by Thaler and Sunstein (2008), provided a step further in differentiating between Humans and Econs, as humans make decisions by the rule of thumb, avoiding loss aversions or status quo biases

based on the social and market norms followed in the society as people are willing to be a part of the group behaviour or known as the herd behaviour. Economic perspective is also important to be known to understand the norms and benefits to be experienced after the change in decisions which can be directed as Econs are rational and about maximizing utility (Bordley et al., 2013). Interesting point is that tough Humans and econs are two schools of thoughts but yet are complementary to maximize the output and will ease the achieving of goal (Castagnoli and Calzi, 1996).

The experiment described in the book- Nudge, where two tables upper view and side view are shown and from one angle it looks a square table and another as rectangle table though both the table are same. But the human will respond as both tables are different but econ will logically focus that angle is different and so the size of the table is the same. It states that econs will mentally account to evaluate and choose options while Humans will have endowment effect people will choose and even pay more that they are willing to buy but does not own them yet and will not take into consideration the sunk cost (Thaler, 2015), Economists like Adam smith, Pigou and Keynes talk about life-cycle of a person which includes incomes, saving rates, investment and consumptions. Further, Hersh gave the concept of “behavioural life cycle hypothesis” which will include removing of limited choices or delay in decision making to achieve a higher goal. People who are irrational thinkers and rational thinkers are bifurcated in system 1 and 2 thinking.

Two System Thinking

System 1 is an automatic system and system 2 is Reflective system. The system 1 is also like a lizard and known as 'reptilian brain' (Abdulkadirog˘lu et al., 2005).An example can be voters who rely on automatic thinking brain which is formed by the first impression or when plane shakes due to bad weather the brain triggers that we will die but our reflective thinking system prompts that we will be safe.

Table 1: Two system thinking

Two cognitive systems	
Automatic System	*Reflective System*
Uncontrolled	Controlled
Effortless	Effortful
Associative	Deductive
Fast	Slow
Unconscious	Self-aware
Skilled	Rule-following

Source: Nudge: Improving Decisions About Health, Wealth and Happiness, 2008

However, system 1 thinking brain have cognitive biases as people are a mix of rational and irrational thinking system (Bhalla, 2014). People usually resolve their problems through short-cuts and make decisions quickly. Heuristics help in developing rule-of-thumb strategy leading to cognitive biases of Priming, Anchoring and Framing (Thaler & Sunstein, 2013).

- Priming is preparing the target person to react in a certain way based on how the situation is set up. People are helped or nudged into doing something by the way it is worded or designed. Do's are focused more than Don't. In an experiment when French music was played at bar more number of French wine were sold while when German music was played German music was sold more which is priming associative technique which is by providing a connected meaning to make the related choice.

- Anchoring is all about placement of the choice or product which we want the person to select. Obvious applications for this in marketing and pricing. Research experiments can help marketer decide how to put the consumer in the desired 'anchor' frame of mind. Example, purchasing a customised laptop. The base model is the default option more people will opt for a basic model but if the fully customized laptop is shown in display there few features which the person will remove but will be higher than the base model making the company earn a profit.

- Framing is a description of a choice can entirely alter the way people notice and perceive the meaning and implications of the choice. A good option can always be described or 'framed' as a poor one, and vice-versa. It is a significant heuristic in how people assess options and make decisions. A doctor may announce that of 100 cases 3% is the failure rate or may announce that 97% is the success rate. It's a matter of orientation, and presentation, or 'framing'.

Choice Architecture

Along with understanding the behaviour economics behind policymaking, and other decisions the act of persuasion also stands important. It is important to understand if people choose because of their willingness to buy or buy because they have to buy. Status quo bias provides an extending hand to persuasion in the behavioural change framework (Insight Association, 2017). People tend to desire the same behaviour if the majority of the people are following the pattern. Persuasion is a method to make a behaviour, pattern or a product look more attractive which

in the long run may lead to change in behaviour making towards the right decision. Intervention through persuasions changes behaviour by providing the right kind of messages and information. The persuasive message refers to

- How information is presented
- How it is framed and communicated.
- What moment is represents
- How the message is delivered
- When will the receiver be more receptive?

Leaders have understood the art of persuasion and often utilize the art in influencing the society, behaviour, pattern, fashion and cultural practices. Where one behaviour becomes a chain for others to adopt similar behaviour in society (Harper, 2015). The US government office by F.D. Roosevelt during the World War-II communicated to the common people that more amount of vegetable production needs to be made at Victoria garden with the sort of persuasion the production exceeded commercial production so a common goal was created in the society

and everyone followed a pattern. Persuasion is required during the beginning phase to push people in taking the right decisions on their own.
Persuasion can happen when a person is initially attracted towards it. Supermarkets, for example, try to attract customers with bright colours (visuals) of fruits and vegetables kept at the visible stand. Bright colour visibility is the source towards persuasion of buying. Another method leading people to persuasion is “messenger effect” (Dolan et al., 2010) where a certain type of message or advice is shared with people. In India, during the pandemic crises of covid-19 government is sending messages to people to stay safe and to stay in-house (Times of India, 2020).In the world of big data, the data related to the searches are collected. People start searching for buying baby products. The companies capture the data and pursuing the potential buyers for their brand and product for the baby. Thus, tempting consumers at the right time to the product (Duhigg, 2012) and trying to change the buying behaviour.

Right Moment

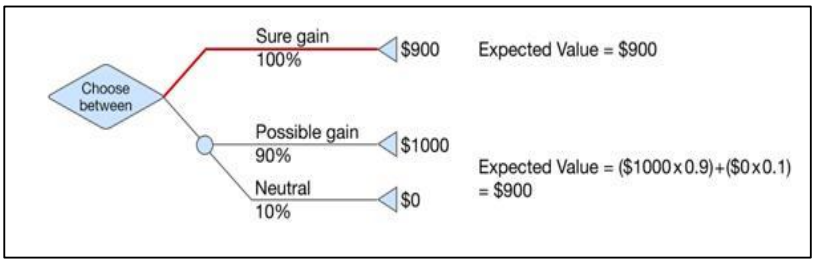
The First mover is always advantageous which impacts the psychological at the maximum. Key moments need to be kept in track and intervention at proper time need (Marcel, 1983) to take place so convincing is done at the right time. The move should be made taking into consideration the preferences of the people and in sync with the current behaviour to modify it to better decision making (Neslin & Heerde, 2009).

Prospect Theory

A way in which people make a decision and how they perceive the options was registered when the weightage of outcome were measured over possible outcomes (Harley, 2016). According to Li & Chapman (2013), people tend to choose options with gains against even when there is a chance of higher gain but also a chance of losing. There are two options provided to the participants:

(1) Chance of winning $ 900 is 100 per cent
(2) Chance of winning $1000 is 90 per cent and getting $0 is 10 per cent

Equation 1: Example of Prospect Theory



Source: Prospect Theory and Loss Aversion: How Users Make Decisions (Aurora Harley, 2016)

The outcomes were such that participants opted where winning $900 was a sure shot. Though the amount for the second option is more than the first option, $1000 but it also had a minimal chance of not winning anything. Participants preferred the option where they had the sure option of winning though less amount.

Social

Humans are known as a social animal, they have a tendency to behave in a particular manner in social gathering so that they are similar to a large number of people around them (Cullen et al., 2006). The surrounding of various people is an important tool of influence. The behaviour of

the crowd will lead to word of mouth and status quo of a particular product or behaviour. The simple example is when we are with someone we are okay to prefer stairs over the elevator or take a revolving door if a person with us or in front of us does so (Halpern, 2010). We constantly try to follow other people with us who can be said as the persuaders for us. As Robert Cialdini states in his experiment that people are more like to leave their flyers on the ground if flyers are already littered on the ground. They are declared 'Social norm' (Cialdini et al., 1990). We approve of the system, behaviour, pattern or decisions based on other decisions and where persuasion is the right mode to influence.

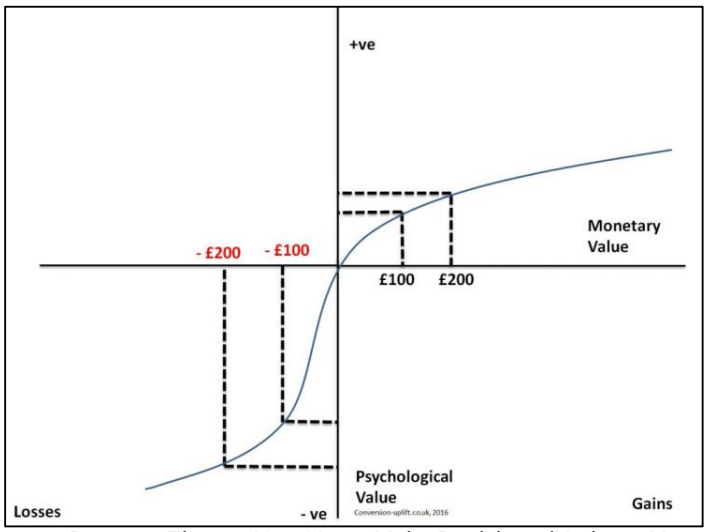
Loss Aversion

People tend to react more on negative messages than the positive messages, this fact has been proven by the neurologists. In other words, people react on losses which they face rather than gain which may happen in future. It can be known by the behaviour when more investments made by the lower-income group is in low-risk investment policies so that there are low chances of losing money (Thaler and Sunstein, 2008).

In conventional economics the concept of loss aversions is not of importance though considered as irrational, it is considered normal behaviour as per behavioural economics (Altmen, 2012). Kahenman in his research has studied that people do often remember their total experience but the most impact is created through the end result through that experience, any loss or gain as a result of the experience. It can be understood by the example that if a patient had undergone a half an hour of operation and was later painless will make a patient feel better rather than an operation which was successful but the pain lasted (Ariely, 2008).

"Pain, I have already had occasion to observe, is, in almost all cases, a more pungent sensation than the opposite and corresponding pleasure. The one almost always depresses us much more below the ordinary, or what may be called the natural state of our happiness, than the other ever raises us above it (Smith, 2012). "In marketing where the product is placed in a manner that buying a product now can help in avoiding paying extra amount later or in other words will save the loss of amount. Example delay in repaying loan which will lead to a surcharge which means paying more amount and thus leading to loss of money. Because of the penalty so charged any person will always prefer to repay the loan on time. Further concepts are related to status quo bias and the endowment effect. People in many stages of their life have a product and value it more than its value or the market price of the product. An experiment conducted by Thaler: A group of people were divided into two groups of which half choose chocolate and other half choose a mug. The possession which the participants felt was

Figure 1: Loss Aversion Example



Source: Prospect Theory: How Users Make Decisions (Nadeem Murad)

high and when the participants were asked to sell the mug and they were demanding more price than the actual price.

The theory by Tversky and Kahneman people make decisions based on psychological values when monetary values are involved. The theory is known as the "prospect theory". The theory caters to three reference points: judgement, emotions and perceptions. The curve suggests that people refer to loss stronger than gains leading to biases of certainty, possibility effect and isolation effect. People gamble on smaller win then a big one which shows that the loss if occurred will be less. People tend to be overconfident in the decision which needs to be taken like insurance company convince clients to buy insurance for critical illness leading to possibility effect. To reduce the features by eliminating common factors and isolating the product a person is willing to choose through psychological value.

Reinforcement and Behaviour (Punishment/reward)

Reinforcement study began with Pavlov's classical conditioning and developed Skinner's operant conditioning which leads to the development of learning and cognitive theory (Luthan et al., 1999). Theory of reinforcement is a widely accepted concept in learning theory, about how students process and retain the behaviour and knowledge gained. The ability of cognitive learning through reinforcements helps in changing the behaviour of people at large (Knud, 2004; Jeanne, 2012). The study of memory shows that it is not about how behaviour is acquired but the main point lies how the behaviour can be retained. This can be performed with a combination of schedule and performance, which is then shaped (Skinner,1958). Behaviour can be classified into two segments: reflective behaviour which is dominated by system 1 brain or automatic brain according to the pattern set in the mind. Another is operant behaviour, which is controlled by the central nervous system which is designed by the frequency of occurrence of influenced by certain events (Burgess and Akers, 1966). There are different types of reinforcements: Positive reinforcement and Negative reinforcement (Luthans et al., 1999; Sidoti, 2008). We take the example of criminal behaviour which are reformulated in the image above. It was found that higher impact on behavioural change experienced positive when coupled with reinforcements (Burgess and Akers, 1996).

Cognitive Hierarchy Model: Experience weighted attraction

Learning in decision making can predict the behaviour of any particular person based on time. The cognitive Hierarchy model was formed by camerer et al. (2002) was experimented in game theory where the time path and the behaviour of the team and players were predicted through experience weighted attraction (EWA). The data collected through EWA is then compared with various random sample groups to understand the positive economic and behavioural values. The game theory in which the concept was put to practice the players understand what they are playing off (Thaler, 2009) when they move on the field. The design of the model id useful in 3 ways:

- The model captures the deviant behaviour which any player shows on the field
- A mixed behaviour which is very well behaved on the field along with few deviant behaviours
- Redefining the equilibrium point in the behaviour of the player and teams.

There will be various play points when where competitive hierarchy model will be where the risk of dominance will pay off and deviant behaviour may be required more than equilibrium

behaviour. The population mix of the players and combination of various equilibrium points will lead towards the pure strategy “purification” or “Fictitious play” (Camerer, Ho & Chong, 2002) Thorndike's Law of effect counts on two effects: positive consequences have higher likeliness to be repeated and behaviour modified based on negative reinforcement will not show a long-lasting impact.

Libertarian Paternalism

Due to majority of the decisions taken based on System 1 approach behavioural economics helps in the prediction of realistic psychological foundations (camerer et al., 2004) which can be applied to economic and non-economic behaviour through oxymoron concept of Libertarian Paternalism, which may be followed by both private and public entities to affect the behaviours along with respecting the freedom of people (Thaler, 2003). The behaviour from the

said concept can be changed or manipulated so that people can have their freedom of choice (Thaler & Sunstein, 2003). People should be free to make their own decision to opt-out or opt for a particular behaviour, pattern or object with the influence of others. This concept leads towards the theory of Nudge, which describes that direct or indirect reinforcements towards influencing a decision that is usually context-dependent.
The new movement of Libertarian Paternalism had its own way of letting people make the right choice and preserving the liberty to make their own choice though pushing people towards the right decision (Thaler & Sunstien, 2008). These decisions can be made for policymaking for private players, saving the environment, for paying timely taxes, education, and governance. People usually try to make their own choice like saving tax by paying them on time or investing them at appropriate modes to reap returns and deduct the amount of taxes to be paid.
The government may levy a tax on a cigarette (Sunstein & Thaler, 2003) in order to discourage buying of it but under certain circumstances like drunk driving, heroin usage or prostitutes government do not provide an option for people but rules are made mandatory.
The concept is defined from the action, rules, and other nudges and regulating Recap, Record, Evaluate, and comparing alternate pricing of the targeted change in behaviour or change in a product.

Resemiotization

Krishnamurti and Bohm (2001) have described culture as “to cultivate”- to grow science, art music, literature, and technology leading to a constructive modification of human beings in the world. For Lotman, is a way of scientific thinking to provide meaning in creating ideological consciousness about current trends (Zylko, 2001). The meaning that any object is associated is through cultural biographies derived from the discourse of human society. Age, gender, social role become the symbol associated with the object have led towards making of symbol and providing a meaning to the object (Valsiner, 2007). These exchange of meanings create significance in the human collection and matures to a norm. This symbolic norm forms the rules of behaviour that need to be followed and performed in society.

Lévi-Strauss (1960), state that these symbols are the collective source of mental representation which are shared through multi-model instruments through a certain art form, visuals, religion, myth beliefs to construct it into a shared belief. Movies are a robust example of transformative cultural resistance (Ashcroft, 2012) this transcultural process which creates a symbolic meaning in the mind of group and society (Bose 2006).

Another research conducted by Eagly et al. (2000) showed the social role of women was defined and restricted to being the lady of the house and her main focus is only family, family happiness and well-being and so linked with common characteristics of being kind and sensitive. Though the traditional role of women had remained unchanged (Barska, 2005; Diekman & Goodfriend, 2006) but since the mid-20th century with the changes in the economic and social situations many developments in the roles have been observed and women have started to move from stereotypical roles as stated in social role theory, which is a result of group member collective behaviour that leads to stereotyping (Diekman & Eagly, 2000), of women's role suitable from nurse to the various positions in corporates, politics, medicine, law, photography, film making, journalist, fashion designer and other (Astin et al., 2002). The changes in the role of women from traditional to non-traditional roles and shifts in the symbolic meaning associated with them carry great importance.

Symbolic Interactionism

Symbolic interactionism in simple words can be implicated as the meanings which emerge from the interaction between people which is derived from the symbols reciprocated through individuals in the society or vice versa. These meanings in the mind of people are evolved through various social actors in their surroundings (Schenk & Holman, 1980). As per an important theorist George Herbert Mead, this process of forming symbolic meaning is possible through various means of communication by creating (Ashworth, 2000). This assigning of meaning communication is two way which is assigned to objects, events which are formed through human interactions

(Blumer, 1969). These arrived symbolic meanings is expressed through language which leads to the developing of attitude at the nonconscious level through an interactive process and interpreting the process to create (Poloma, 1999: 224-225; Tye & Tye, 1992). Thus, O' Shaughnessy (1992) the meanings so formed affect the behaviours and actions which is formed by humans and society.

Nudge

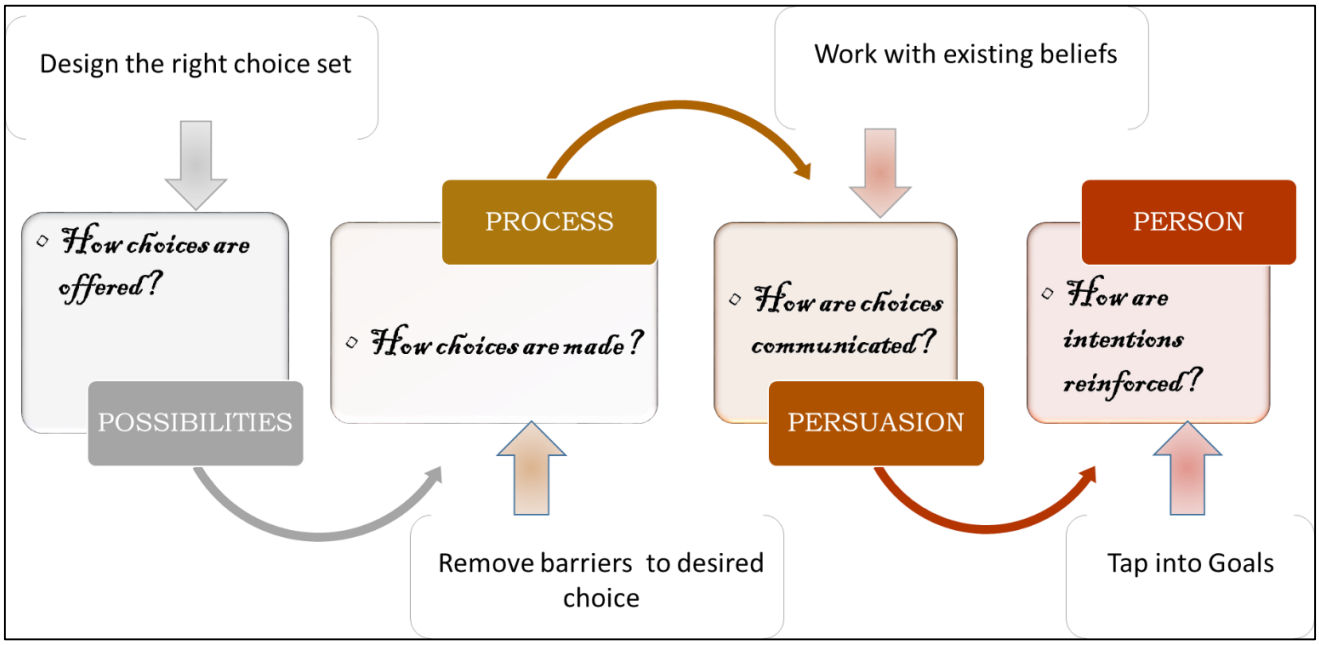
The interdisciplinary concept of nudge is widely used by people knowingly or unknowingly in their day to day life. Thaler and Sunstein (2008) in their book Nudge: Improving Decisions About Health Wealth and Happiness communicate few examples where people think about saving for tomorrow, the doer wants to spend the money earned while a person with planner personality believes in saving for later in life, this is where insurance sellers or the banks can pitch in for their retirement plants where people are nudged to live and spend in their today and save for tomorrow. Another example is related to Health, in case of Organ donation where 95% of people from US showed willingness for organ donation but were not performing the needful consent form during their visit to hospitals and design architect was made, when people had to choose to opt-out as an organ donor the consent increased to 82% than for opting in as organ donor which was recorded as 42%. This Heuristic process can be used by the organization to nudge employees towards the positive behaviour in the organization or drive customers towards the products.

Behavioural economics is critically dependent on psychological assumptions and how people behave while taking decisions through a branch from neo-classical is also known as the "invisible hand" where sentiments are considered along with economic concepts (Camerer &

Loewenstein, 2002). Though Nudge is being one of the modes for behavioural economics there is still lack of evidence on the nudge work (Moseley& Stoker, 2013; Kosters& Van der Heijden, 2015). To understand the impact of nudge clearly Zoë Chance, Margarita Gorlin, and Ravi Dhar from Yale School of Management developed 4P framework to understand the shift in the behaviour.

- Possibilities: The availability of choice is very crucial, the design formed should be in a manner that options are easily available and recognized by the irrational thinking brain of the person. If we discuss the cafeteria experiment performed by Thaler and Sunstien in a cafeteria that students did not have any hesitation regarding healthy food it was just not at the eye level means was invisible for the students to choose.

Figure 2: 4P Framework to understand the shift in the behaviour



- Process: people as stated in the book Thinking Fast and Slow by Daniel Kahneman, decisions are influenced by their irrational or system 1 brain. The system 1 brain is conditioned by the social norms, behaviours in their cognitive capacity which functions on autopilot mode. If the choice availability is visible it becomes easy to process the brain and make them nudge towards right decision making. As per the newspaper experiment carried out by Ariely as mentioned in Predictably Irrational shows that the human brains are anchored to a particular choice because of the priming effect. It becomes important to show a higher choice option, in the experiment where the price for 1 year of online subscription is US$59, the Print subscription is US$125 and print plus web subscription is US$125. 84% of the respondents opted for option 3. While in another part of an experiment, two options of price per year were provided where an online subscription was US$ 59 and other option of print medium charging US$ 125 and 68% opted for online mode. Thus, the behaviour process was being implemented.

- Persuasion: Any person can be persuaded when the right information and choice is provided to them. The choice should push the person to take an appropriate decision. If the decision need to be taken is complex system 2, the rational brain is also known as the 'lazy policeman' as it evaluates the options taken by system 1 brain. In the experiment conducted by Alberto Alemanno regarding behavioural turnabout tobacco consumption, one choice architecture which was performed that cigarettes were kept away from the eye visibility so the buyers were ashamed of asking for the packet and the consumption of cigarettes reduced for the store. Another experiment took place was that brand name of the cigarette manufacturer was not allowed to be mentioned and on the white packet, only the contents were visible which would push people to take decisions of not consuming or reducing consumption of cigarettes due to visibility of harmful content and reduction of status quo due to non-visibility of the brand.

- Person: By taking the right decisions many times it becomes a behaviour of a person and routine to take a similar decision which are or will impact the societal decision. Goals which are also small wins should be measured to maintain the momentum. Temptation can be a mode of bundling the goals. Parents can promise to buy a videogame if the child scores above a certain percentage though the ultimate goal will remain the child completes their education with good marks. A temptation for the cycle is just a temptation for a small victory.

Bounded Rationality

An idea that people make decisions based on limited rationality. The theory is dependent on intuitiveness and accessibility which is an automatic operation of decision making while accessibility is the deliberate reasoning and understanding the consequences behind any decisions which need to take place (Simon, 2004). The drawback experienced with intuitiveness is the biasness that follows. To negate the drawback it is important to make a decision based on both the factors like a pair of scissors where one represents “cognitive limitations” and other “structure of environment” as people are a mix of rational and irrational actions as perfectly rational decisions may not be practical (Gigerenzer & Selten, 2002).

Intuition and Accessibility

Intuition and Accessibility are two different generic methods of cognitive functions. The intuitive mode wherein decisions and choices are made automatically and quickly and a controlled mode, which is intentional and slower. Considering these two factors (Tversky & Kahneman, 1971) has conducted research among the Statistical researches. The instinctive decisions of these Statistical researchers didn't fit in with Statistical Principal with which they were completely aware of. Specifically, their intuitive Statistical reasoning and their assessments of factual Statistical Principal demonstrated an absence of congruence with the Sample Size. There was a clear inconsistency between Statistical Intuition and Statistical Knowledge. Even critical research choices, for example, the decision of Sample Size, are routinely guided by the flawed intuition of individuals.

The overall capacity for mental effort is limited, effortful processes tend to disrupt each other, whereas effortless processes neither cause nor suffer much interference when combining with other tasks (Kahneman, 1973; Pashler, 1998)

Rational Decision making

Herbert A. Simon (Decision Making, August 1993) has described the Decision-making process in three stages. The first stage of the process of finding the problem requires significant range of skill set that which is required for finding and attending the issues. Finding and attending are critical to such an extent which directs towards appropriate insight. The second part of the process is, once we finalize the problem, finding the alternatives and appropriate the proper solution suits the problem. In terms of the psychologists' in this process, Humans elaborate and craft all the possible solutions to the problem. The third and final process, assessing those solutions and selecting among the solutions.

Economic Rationality and Psychological Rationality

There is a significant difference between the psychologist and economist approach to the rationality. In the economic point of view, rationality is considered to be substantive. While in psychology the rationality is considered to be procedural. The theory of economic rationality concentrates on the substance of the decision to the disregard of the procedure and only deals with the third steps in the decision-making process. Economic theory does not focus or consider attention and emphasis where the options of decision originate from. Unlike psychology rationality.

In reality, Humans have to get along by having preferences and taking into consideration action plans that are not appealing to them. We as a human have to implement the mechanism called the aspiration level. We, for the most part, make an evaluation of what the world is probably providing for us. Human's aspiration level accommodate considering our expectations, in the view of our own and other's experience. Herbert A. Simon (Decision Making, August 1993) has explained by giving example. In a survey conducted by Fortune Magazine on US Nationals asking about their expectation on salary hike which helps them to become less anxious about money. Very strangely from middle level to lower level manager trainee has given the same answer which is 10% increment. This is usually a characteristic of human aspiration level which easily adapts from their and other's experience about what is within their reach. And that turns into significant standards by which Human decide that they have come to a satisfactory solution to the problem

The decoy effect

Dan Ariely in his book predictably irrational has explained the decoy effect through relativity. Relativity is straightforward. In any case, there is one characteristic of relativity that persistently confuses humans. Humans not only have the tendency to compare things or experience with another yet additionally will, in general, focuses on comparing at things that are effectively equivalent and refrain to compare things that cannot be compared. In a survey conducted among 100 students at MIT's Solan School of Management for the subscriptions of the Economist'. Students have been given 3 choices.

1. Internet-only subscription for $59.
2. Print-only subscription for $125.
3. Print-and-Internet subscription for $125.

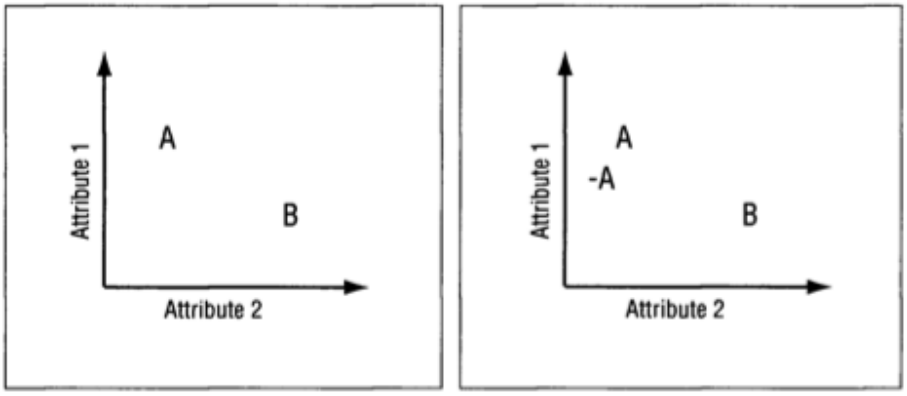


Figure 3: Decoy Illustration

16 Students have opted option 1 and 84 Students have opted option 3. None of them has opted option 2 very surprisingly students of one of the most reendowed management institute were dominated by the option of print-only subscription which is decoy among the three options. When the researcher had removed the decoy, 68 students had opted for option 1 and only 32 students opted option 2. Human rarely nominates things in absolute terms. They do not have any internal measure that reveals credit or worthiness of things or experience or maybe focus on the overall preferred advantage of one thing over another and measure their value accordingly. This behaviour pattern is completely irrational. The following figure explains how relativity works.

In the above illustration, supposedly there are product A and B, having the different attribute. Product A is better in attribute 1 and Product B is good at attribute 2. Left Illustration gives very straightforward options. Consider the right side of the illustration –A is the decoy which constitutes a comparison between all 3 options. Introducing decoy(-A0) will make a better option among all. Hence nothing is rational in presence of Decoy. The human mind will always look for a better alternative. Many people don't have slightest of an idea what they want unless they see in context.

Everything is relative, and that's the point he might be a mere curiosity, but for the fact that it mirrors the way the mind is wired: we are always looking at the things around us concerning others. We can't help it. This holds true not only for physical things—toasters, bicycles, puppies, restaurant entrées, and spouses—but for experiences such as vacations and educational options, and for ephemeral things as well: emotions, attitudes, and points of view. We not only tend to compare things with one another but also tend to focus on comparing things that are easily comparable—and avoid comparing things that cannot be compared easily.

The economic theory of valuation, by and large, presume that cost of resources are derived from the underlying essential values. One cannot deny the fact once the initial prices (arbitrary) institute in Human's mind they will form not only present prices but also the future prices (Coherence) for

the commodities. The existence of arbitrary coherence plays a major role in the decision-making process.

Herd Behaviour

Humans are social creatures, the communication is the mode of influence and then behave in a similar pattern which is also known as “Herd behaviour” (Shiller, 1995). To behave and take decisions in a similar pattern is irrational along with influencing and motivating people to make similar decisions (Jost, 1995) so being socially nudged by other humans. There are various behaviours in which people obtain information, one is through observing others and taking action accordingly, “information cascading” model (Banerjee, 1992; Bikhchandani et al, 1992). Other is a transfer of information through conversations in intergroup, "conversation analysis" model (Goodwin & Heritage, 1990) and study of cognition on the societal level, “Socio-cognition” model (Levine & Resnick, 1993).

Information cascading

Asch's theory proved that an answer which is given by the first person in the panel will be considered correct answer or decision and rest members in the panel will support a similar decision. For instance, as mentioned by Shiller (1995) if given an option of dining in two restaurants, the first person may use their rational brain to make the decision or it may be irrational which may be close to their culture of memories and might choose restaurant 1 but other people are also like to choose restaurant 1 being influenced by the first decision-maker and everyone can end up in a bad restaurant and this creates a bad equilibrium arising from the herd mentality. Similar is the example of drugs and alcohol as the researchers mention. One person in the group will decide to drink and other people in the group will adapt to the social norm. The first-mover advantage also players an important role in influencing the group.

Conversation Analysis

Speech is an essential mode which impacts the emotional and interpersonal behaviours which leads to influence towards a new behaviour. The flow of ideas from person to another id through the mode of conversation. One of the power tools of conversation and impact creation is media which provides us with information but a better mode is always face to face conversation which lead towards deviant behaviour pattern (McGuire, 1995).

The stimulating conversation in the group is usually known by the members. A large part of the information is still transmitted through conversations, leading to influencing of other members in the group which further influence other people of other groups (Levine, et al., 1993). The ideas then get associated with the group. As a result of observing the behaviour of other members and listening to the opinions about others, it gives a priming effect of association and no member would want to feel differentiated from the group or the social norm (Thaler & Sunstein, 2008). Each person has a choice and space to their own decisions but ultimately humans are influenced from the outside information and observations. If the nudge happens even for one person in the society the recurring effect will take place. Though cascading can be fragile but can also lead people towards the right behaviour Economic concepts when working simultaneously with realistic Psychology boost the understanding of the people and boost behavioural economic concepts (Camerer & Loewenstein, 2002).

PROBLEM DEFINITION

Looking at the cost the canteen has to manage, it becomes a cost centre when “Shetty Hospitality” the contractor of the canteen has to employ more people to clean the tables than in serving students. Students in GICT building are known for having limited time between the sessions and the serving time if observed is around 10 minutes per dish if the students are willing to buy a hot snack. As a result, the preference for students' food is unhealthy packet food, tea or coffee which takes very less serving time.

The observation was made when students are offered healthy food and the contractor have almost 11 people as employees the reason for students preferring packet food over hot healthy snack which they are even willing to buy. In an informal discussion with students sitting in canteen stated: " canteen food is good but who will wait for long hours" another student stated that " Canteen is not capable in serving in less time, maximum time we get between sessions is 15 minutes:10 minutes in the canteen and 5 minutes for going up and down the stairs."
Even though it was observed that canteen keeps a limited amount of stock of packet food which usually gets over by 2:00 pm if the stock kept is more than other days. There are two purposes of the canteen to keep less stock of unhealthy packet food:

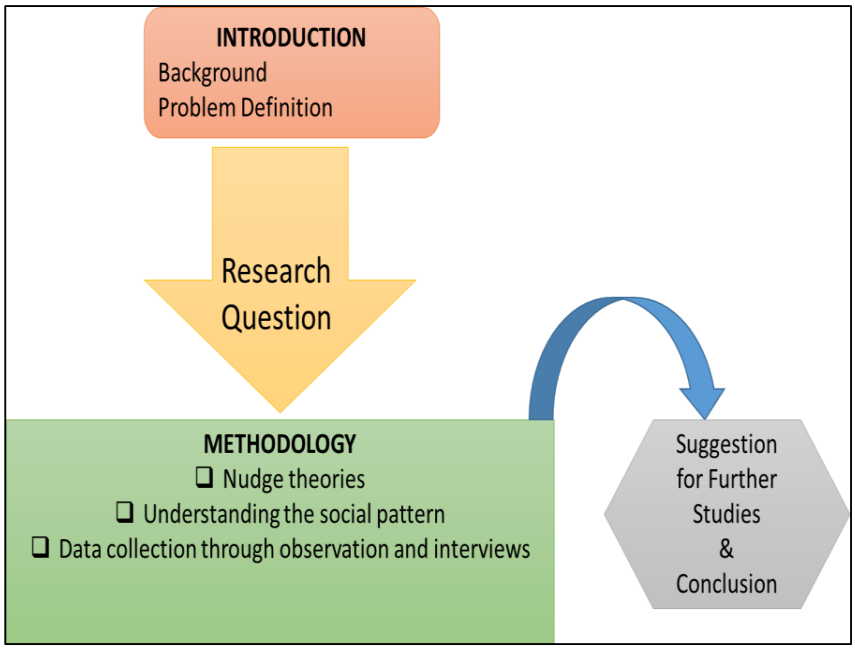
1. Encourage students to consume hot and healthy food
2. Increase their profit margin if students order healthy food.

Another student added to the conversation that “Canteen food is tastier than before, but they are not able to meet serving time, we need to bring food from home.” To understand the issue to benefit both the entities, students and Shetty Hospitality, observation for a week was made and it was known that on an average 300 students are served per day. Though 11 people active in the canteen premises should not take more operational time but further it was known that 2 people from the staff who are dedicated to picking up plates from the table and washing the dishes. So, behaviour change had to be introduced for students to pick up their plates and keep them on the counter. A research was conducted to observe if a nudge sign, which prompts the decision-maker to take the right decision. Provoking students to develop a better habit of making the right choice of picking up their plates and keeping them at the washing area which is in the general interest of student behaviour and Shetty hospitality running the canteen.

PROJECT DESIGN

Initially, background research was done on the canteen and the behaviour of students in the canteen. It was then studied that student behaviour of not picking up their plates and putting at the designated place was creating a problem for students in the long run which impacted them ending up opting for junk/packet food even when they were not willing for the same. Nudge theories were used with transparency. Thaler and Sunstien introduced a concept of 'visibility' long with 'monitoring' to have a tab on the change in the behaviour. Brian Wansink’s experiment on the size of the plate was: If a restaurant or the cafeteria is willing to earn more profit in buffet system by wanting customers to eat less, the size of the plates is then reduced so that customers will take less food in the plate, thus the cost incurred on food is less. These nudge interventions are neither transparent nor visible to monitor. Any changes in the behaviour cannot be recorded in such non-transparent conditions.

Figure 4: Project Design



In this current research, A transparent approach was used where signage was put up designed in a manner to impact the system 1 brain and nudge students to take appropriate decisions to keep their plates at the allocated places. Though the sign was visible always at the eye level with particular colour schemes and fonts students consciously or non-consciously will be expected to put their dishes as instructed leaving room for next batch of students to sit on the particular table with neatness and ease of time. The expected behaviour was recorded and analysed.

ARCHITECTURE AND DESIGN

Analysis performed in the oral and written format utilized for social context which is beneficial in studying real-life situations. It helps in bridging the understanding between cultural rules, values, beliefs historical context communicated in different languages (Luo, 2019). Thus, discourse analysis is taken place when there is the involvement of both verbal and non-verbal interaction in the form of verbal communication which may include face to face communication and nonverbal interaction which may include signage, symbols, colours or documents as a mode of communication.
Discourse analysis is performed to understand the social life and community practices or behaviour and its implication helps in providing in-depth study from micro-level face to face study to macro-level of community practices (Shaw & Bailey, 2009). The study does not ease the data analysis but explores the in-depth analysis of the impact of the ideas and ideologies, cultural gender studies impact, the influence of interdisciplinary there is various research which took place to understand various practices in the society.

Case studies on Discourse Analysis

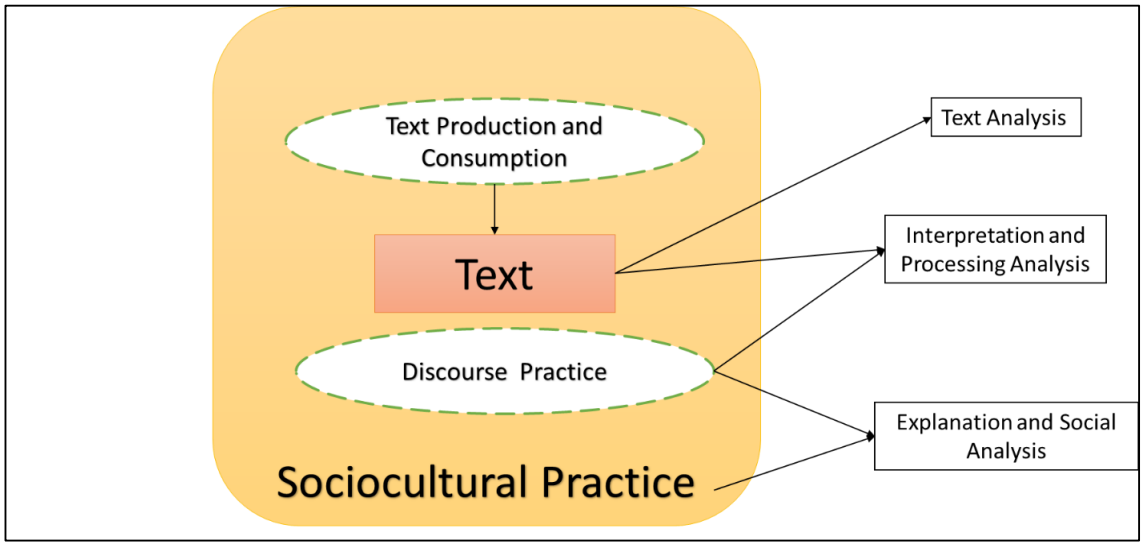
1. ***Study 1 (Roberts et al., 2005)***
   Misunderstandings can occur due to lack of appropriate mode of communication. The research carried out an analysis on 232 videotapes and 37 tapes were transcribed in detail to understand why misunderstanding may occur between the patients and general practitioner. Any kind of miscommunication led to misunderstanding in communication which meant analysing of stress, presentation style, tone, and connotation. The general practitioners trained in a certain community will lead to misunderstanding if dealing with the patients of other community.

2. ***Study 2 (Wilkinson &Kitzinger, 2000)***
   Different cultures have a different mechanism of talking about a different coping mechanism for a life-threatening disease. There are many brave women in the world who suffer from breast cancer but have different methods of coping up and trying to talk and think positive about their life. Focus group discussion of women who had suffered from breast cancer, the FGD was taped and transcribed where they discussed the first time they felt when they heard the news and how they coped up with positive words and context to help them cope up in bad times. The talking about positive comments was a coping mechanism, the therapy which they undertook which focused on thinking positive so the conversation followed was the positive thoughts, illness or death which leads to the bonding of the group.

3. ***Study 3 (Brandes&Lachover, 2015)***
   A study in Israel took place after Dove had started a campaign about real women, the idea of the concept was meant to be understood by the researchers that the company was trying to portray. The agenda of the company was to change the stereotype image of women that the country was carrying in society. Two mediums: internet and media were used to perform discourse analysis from the perspective of the Body, context and title of the image that was being portrayed. The non-linguistic context like the colour and font and the impact that was being created in the minds of readers was also analysed. It was analysed that the efforts made by the company on the right platform were at least able to address the issues that the country was facing which might lead towards a cultural shift of redefining "Real Women".

Critical Discourse Analysis (CDA), an interdisciplinary in approach states that sociocultural practices are defined by simple talk or text. The changes in the behaviour pattern formed previously or being made are due to the reinforcement of the linguistic and non-linguistic usage lead to changes in the structural inequities in society (Blommaert & Bucean, 2000). CDA a step for providing insights related to the socio-cultural context along with manipulating impression given to the audience. A mixture of common-sense and social practice is a language of analysis

Figure 5: Three dimensional model



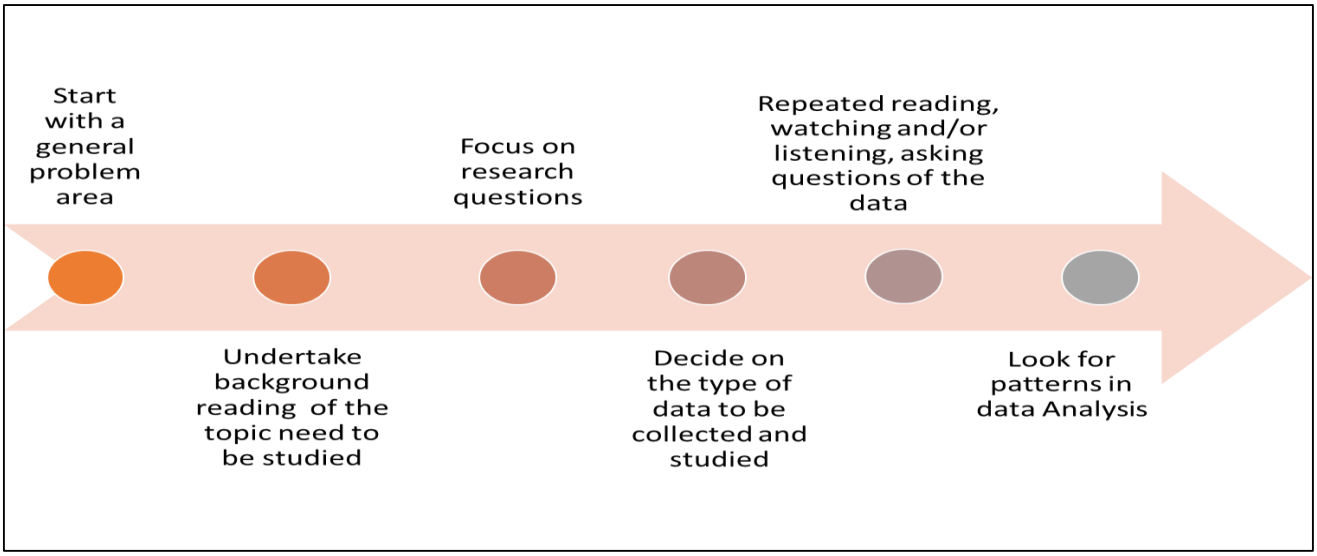
for both oral and written communication as stated by Norman Fairclough (1989) in his book Language and Power.
Fairclough developed a three-dimensional discourse analysis framework where the text produced and consumed were utilized for text analysis and further used for interpreting the practices which were being followed in the social scene in a cultural context. The framework is a combination of three levels: micro, meso and macro-level of social context and the impact which is being created in the scenario (Fairclough & Norman, 1995; 2001).

The three-dimensional model goes in the flow of Text analysis, Processing analysis and Social analysis, the analysis between the impact of text leading to discursive practices and change in the social & cultural behaviour (Weiwei & Weihua, 2015). Discourse analysis is said to be the medium of power and ideology which is being followed or need to be followed in interdisciplinary fields which may be analysed at a multifunctional level at three levels of analysis conceiving the relationship between society and culture.

Steps to perform Discourse Analysis

If placed in a broader sense the research starts with an understanding of the general problem area and defining the research question and the content which need to be focused on. The analysis can be applied to both small and large sample size and at any particular time frame. The literature review and past researches are studied along with the theoretical frameworks used in past references to help in guiding the real-life experiences. The interdisciplinary method then decides on the research question which needs to focus on deciding how the data will be collected through verbal or non-verbal ways which may include vocabulary, Genre, conversational codes, non-verbal communications, and structure. After which the data can be analysed and understanding of pattern might take place related to the research question and conclusion of the real-time research.

Figure 6: Steps for Discourse Analysis



RESEARCH METHODOLOGY

Research method can be in three folds: Exploratory, Descriptive and Explanatory.

- Exploratory Research will study the 'what' part of every research and will be performed in-depth seeking new depths by asking a lot of questions which may be performed by a focused group, in-depth literature review and conducting an interview of field experts. This method has a broad focus on the topic and is open towards the "what" of the research question (Adams and Schvaneveldt, 1991)

- Descriptive Research methodology ‘to portray an accurate profile of persons, events or situations’ (Robson 2002:59). It can be performed on both qualitative and quantitative data which identifies the trends, frequencies, and correlations between variables or characters. Various method of performing the method is through case studies, observations, surveys and other techniques which can probe in-depth collection of data.

- Explanatory research, on the other hand, explains the causal relationship between the variables which is helpful to understand the reason behind the task performed. This technique is performed on quantitative data, but the reasons can be explained through the qualitative data collected.

In this Research, Descriptive research technique was used where observation was performed on the participants. The data about the representative's behaviour will be gathered in the canteen premises of the Ahmedabad University, GICT campus.

RESEARCH HYPOTHESIS

H0: Positive reinforcement through Nudging does not have more impact on representatives
H1: Positive reinforcement through Nudging has more impact on representatives

SAMPLE FRAME

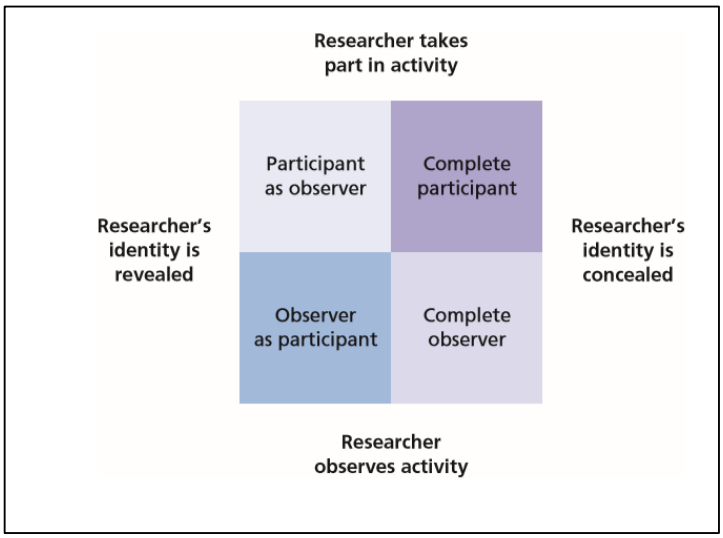
Students of Ahmedabad University visiting the canteen of GICT building were the sample size for the research

DATA COLLECTION

The students of Ahmedabad University, GICT building visiting the canteen have been observed from morning 9 am to evening 6 pm. The emphasis was on discovering the meaning behind the behaviour of leaving the tables dirty and not picking up their dishes after having lunch or snacks. As an observer I had to understand the behaviour and the symbolic meaning associated with their behaviour and getting to the bottom of the construct that an individual has given. These four roles which are performed by the participant-observer (saunder et al., 2009)

- Complete participant the researcher is the part of the group and performs the task but the group is unaware of you being an observer.
- Complete observer researcher does not reveal their identity and is not the part of any activity taking place but they observe the consumer Behaviour of the representatives
- Observer as participant type of researcher are the spectators and assist in the activities taking place like you may assist in the team building activity to observe the representatives and the insights can be jotted down.
- Participants as observer reveal themselves as the observer and the observer tries to gain the trust of the representatives and can ask questions to enhance their understanding of a particular behaviour of the representatives.

Figure 7: Typology of participant observation researcher roles



As the part of this research to be conducted the observer participated as a complete participant, as the observer I was also the user of canteen of GICT building and is a student, and was also being part of the group which have been observed. Thus, the identity of me as an observer and the project was not revealed and the study which took place covered: how many students are following the rules made by the canteen community which consists of removing of plates from the table when the food has been consumed. Certain signage was introduced as part of the intervention and the impact was measured for the same.

RESULTS

Language and Visual

The target audience which was people visiting canteen should feel an impact and the need to perform the task. The language used is on the order tone direct speech to provide with an impact that it is the duty of the person consuming food at the canteen to clear the table and make way for another person to sit comfortably on the table.

Figure 8: Signage

Spellings and caption were double-checked and went through three processes of approval the canteen supervisor, facility manager and research guide to check if the words, language and tone were providing with given impact. The majority of the canteen users were students, faculties and staff members so the notice out up was in the English language. The Notice word is written in Red Arial Black to make it eye-catchy and important on the other hand blue is a soothing colour which will negate with red colour and not provide a danger impact but creating importance.

Deconstructing the Notice:

- A sociocultural practice which is about people leaving their plates, packets and cups on the table and making it inconvenient for others to use the same table. The group behaviour of leaving the plates on the tables lead to a negative impact on dissatisfaction among customers and increase in serving lead time of the order.

- Process Analysis was done when the signage was placed on the table at the eye level with soothing and tinge of bright colour used in the notice format to grab the attention among customers which may also lead to the initiation of the change in the behaviour among few people in the group which may lead to status quo bias and also leading to herd mentality that a person in the group is picking up their plate and if another does not will lead to embarrassment for other.

- Text analysis was created based on **Cohesion** the essentials in parts of interpretation, **coherence** the order of statement which pinpoints to the point that action needs to be taken, **Acceptability** was also observed where no customer in the canteen felt any hurtful message which also remarked the **Situationality** which they were observing and **Intertextuality** which lead to schema and representation of the message which was clear.

DISCUSSION AND CONCLUSIONS

The theoretical framework was developed to analyse the impact of signage as a case of nudging to avoid leaving plates, packets and cups on tables. Thus, will focus on *will reinforcement through Nudging have an impact on representatives?* The customers in the canteen were provided with free will instead of punitive activities leading to a decision to try to make right decisions of making the customers aware about the right habit and leading to influencing of other peers in the group. During a conversation with students about leaving plates, cups and packets on the table it was understood that they believed in not picking up their tableware is not they wanted to do consciously but because of the service of pick up waste from the table was provided by Shetty Hospitability students did not find the need to put their tablewares at the allocated place. Thus, a disconnect was observed between attitude and action.

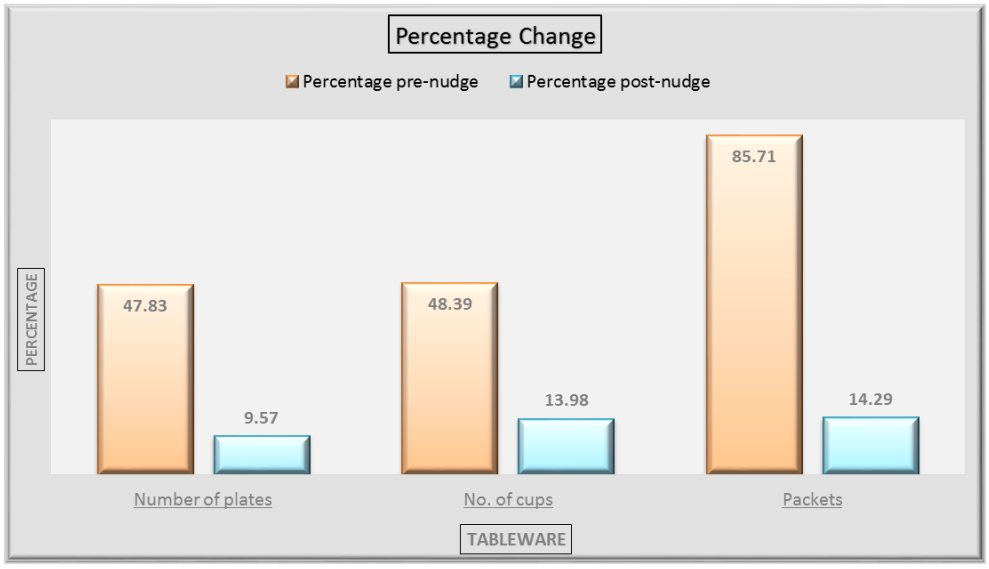
A visual-linguistic reminder has worked as a mode of artefact and help in nudging and provoking people to take the right decision. With interaction with people, it was known that the artefact was "internalized" by customers and bridging the gap for routinizing the behaviour. A student stated that "It is not that I was not willing to pick up my tableware but it never occurred to me that I should put my plate at the allocated window, the signage poked me to remember that I need to perform the activity." Another student, on the contrary, said, " we pay in canteen and employees are appointed for the task why should I pick up my plate but when I am here with my group members some of them picks up their plate so by compulsion I do the same."

The positioning of the artefact-signage on the table was visible at eye level which was helpful for a constant reminder and lead to an external environment and forces which leading towards behavioural change and influencing the social practice. Thus, in the long run, it will impact the change in herd behaviour of keeping the tables clean as pointed and interpreted by system 1 brain.

The nudge design took place based on the 1-week observations which took place based on a few incidences. It was observed that in general before putting up on signage students be it two people sitting and discussing over breakfast, a gang of 6 students who have come to relax and ordered evening snack, or students having a bit of time were watching series or playing mobile games let their plates, tissues, plastic bottles or cups people washed their hand and left and did not bother to clean the table. It was also observed that if group 1 sat on the table had food and left tableware there the next group which sat on the same table with dirty tableware would tend to follow the same behaviour. There were instances which felt that people unintentionally forgot to pick up things from the table. For instance, a student sat alone in the canteen for long hours order hot & cold beverage and a plate of samosa in the afternoon he suddenly rushed out he kept the dirty dish at the right place but forgot to take his beverage cups. Similarly, on the day when the street play took place in canteen 2 regularly visiting girls usually clean their table, on that day they ate, saw the play and left.

On a particular day which accounted of highest sales of 337 units were sold which were consumed inside the cafeteria premises a total of 196 plates, 91 cups of beverages and 50 packets were ordered and almost 124 students in general left tableware where they were sitting.

Figure 10: Percentage change visible pre and post Nudge



When compared to a number of plates almost 50% of people behaved that it was okay to leave cups and packet unhealthy food wrapper on the table even when they were aware that other people will be sitting the same table when they leave.

Post observation the intervention was put to practice when the signage was put on the table, the position was selected after the discussion with concerned people involved in the permission granted, it was also understood that it will be visible is signage was kept on a table and visible to everyone the whole time that they were sitting. The impact recorded was a tremendous decrease in leaving dirty tableware on the table. Initially, almost 52.54% of the tableware was left which reduced to 11.86%, a drastic decrease was observed by putting up signage at the eye level and catchy. On interviewing few students it was known that they were aware that they need to keep tableware at assigned place but many of them did not leave is behind intentionally now with the sign continuously seen they recall at the back of the mind that they need to clear the table of other to sit there comfortably. Another person revealed that at times when they are sitting in a group at least someone would get their eyes on the notice put and would poke others to clear the table.

The *herd behaviour* is being observed where one person of the group prompts others to follow the norm. Though the impact was made by external changes it was developing into a *social norm* where people had a change their attitude and opted to make a right decision mainly because of the *availability* of sign in front of their eyes. Moreover, if one person from the group followed the norm other due to *status quo* effect and peer pressure that they will stand out of the group if they do not keep the plates.

Conclusion

The thesis humans do not always work on the utilitarian theory where people try to make a rational decision of achieving maximum utility. This situation occurs when people are not affected by emotions and their surroundings but the fact lies that human brain defined as system 1 and system 2 brain makes decisions based on peer pressure, religions, differences, gender, legal and past behaviour. This states that people are capable of taking irrational decisions which are supported then by the rational brain as based on the studies from social, cultural, economic, cognitive and psychological disciplines. More than rational decisions being taken people tend to make decisions based on their status quo, herd behaviour, availability or for avoiding losses. Nudge, a positive reinforcement theory suggests people take correct decisions which may be healthier in nature and beneficial in long run but to do so a choice architect of nudge need to be performed to trigger at the appropriate part of the system 1 brain so to make people drive in the particular direction though it may be based on free or libertarian paternalism.

The experiment also supported the nudge theory that people who were leaving their table dirty with their tableware started pursuing the right actions and change of attitude and behaviour due to a shift in social norm which took place. The transparent nudge made people aware and bridge gap in the mind of people which they thought was right but were unable to perform. A transparent nudge has a long-lasting impact because people are aware of the decisions they are taking with their own free will and this impacts more on the behavioural change
The drawback of the study amid coronavirus outbreak the research could not be carried forward to understand the behaviour change if it impacted in the long run by removing the signage and recording the impact. Though with the interviews which took place it was figured that it did trigger in the mind of many and they did recall the notice kept on the table which urged them to make the right decision. So further study can be conducted with the same experiment to understand the in-depth impact.

REFERENCES

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., &Sönmez, T. (2005). The Boston public school match. *American Economic Review*, *95*(2), 368-371.

Ariely, D. (2008). Predictably Irrational: The Hidden Forces that Shape our Decisions. London: Harper Collins.

Ariely, D. (2015). Behavioural Economics Saved my Dog: Life Advice for the Imperfect Human. London: Oneworld.

Asch, S. E. (1952). Social Psychology.Englewood Cliffs, NJ: Prentice-Hall, 1952.

Ashcroft, B. (2012). Bollywood, Postcolonial Transformation and Modernity. *Travels of Bollywood Cinema: From Bombay to LA*, 1-18.

Astin, A. W., Oseguera, L., Sax, L. J., and Korn, W. S. (2002). *The American Freshman: Thirty-Five Year Trends*. Los Angeles: Higher Education Research Institute, UCLA.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, *107*(3), 797-817.

Barska, A. (2005). "Tożsamośćspołeczno-kulturowapłci w kontekścieponowoczesnegoświata," in *TożsamośćSpołeczno-KulturowaPłci*, eds A. Barska and E. Mandal (Opole: UniwersytetOpolski), 15–23.

Behavioral economics: Humans vs. Econs, a history of bringing traditional economics down to earth : Monthly Labor Review. (2016, May 1). Retrieved from https://www.bls.gov/opub/mlr/2016/book-review/behavioral-economics.htm

Better Eating with the 4Ps Framework. (2020, February 12). Retrieved from https://som.yale.edu/blog/better-eating-with-the-4ps-framework

Bordley, Robert, Luisa, Uberti, &Mariacristina. (2013, November 15). A Target-Oriented Approach: A 'One-Size' Model to Suit Humans and EconsBehaviors. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2354058

Bose. Mihir (2006), Bollywood: a History Stroude, UK: Tempu

Camerer, C. F., Ho, T. H., & Chong, J. K. (2002). A cognitive hierarchy theory of one-shot games: Some preliminary results.

Cullen et al., (2006). In their survey of students, 61 per cent said they would take the revolving doors if the person ahead did so.
http://web.mit.edu/~slanou/www/shared_documents/366_06_REVOLVING_DOOR.pdf

Cullum, B. A., Lee, O. L. I. V. I. A., Sukkasi, S. I. T. T. H. A., & WESOLOWSKI, D. (2006). Modifying habits towards sustainability: a study of revolving door usage on the MIT campus. *Massachusetts Institute of Technology*.

Cialdini, Robert B., Reno, Raymond R., and Kallgren, Carl A. (1990), 'A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places', Journal of Personality and Social Psychology, vol. 58, no. 6: 1015–26.

Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827-874.

Camerer, C., Ho, T., & Chong, J. K. (2003). A cognitive hierarchy theory of one-shot games and experimental analysis. *Available at SSRN 411061*.

Dhar, R., Clark, G. R., Dhar, R., Clark, G. R., Yale School of Management, & Yale School of Management. (2017, September 15). Nudging the Consumer: How to Make Desirable Behavior the Path of Least Resistance: Insights Association. Retrieved from https://www.insightsassociation.org/article/nudging-consumer-how-make-desirable-behavior-path-least-resistance

Diekman, A. B., and Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Pers. Soc. Psychol. Bull.* 26, 1171–1188. doi: 10.1177/0146167200262001
Diekman, A. B., and Goodfriend, W. (2006). Rolling with the changes: A role congruity perspective on gender norms. *Psychol. Women Q.* 30, 369–383. doi: 10.1111/j.1471-6402.2006.00312.x
Dolan, P., Hallsworth, M., Halpern, D., King, D. and Vlaev, I. (2010) MINDSPACE Institute for Government, London.
http://www.instituteforgovernment.org.uk/sites/default/files/publications/MINDSPACE.pdf.
Eagly, A. H., Wood, W., and Diekman, A. B. (2000). "Social role theory of sex differences and similarities: a current appraisal," in *The Developmental Social Psychology of Gender*, eds T. Eckes and H. M. Trautner (Mahwah, NJ: Erlbaum), 123–174.
Econ's vs Human's. (2017, January 30). Retrieved from http://www.laeconomiadelcomportamiento.com/en/econs-humans/
Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual review of anthropology*, *19*(1), 283-307.
Harley, A. (2016). An interactive node-link visualization of convolutional neural networks. In *International Symposium on Visual Computing* (pp. 867-877). Springer, Cham.
Ho, T. H., Camerer, C. F., & Chong, J. K. (2002). Functional EWA: a one-parameter theory of learning in games. *University of California Berkeley, Working paper*.
Illeris, K. (2002). The three dimensions of learning.
Kahneman, K. (2013). Thinking Fast vs. Thinking Slow. Retrieved from https://www.youtube.com/watch?v=PirFrDVRBo4
Kosters, M., & Van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. *Evaluation*, *21*(3), 276-291.
Krishnamurti, J., &Bohm, D. (2001). *The limits of thought: discussions*. London: Routledge.
Foley, D. K., (2009). *Understanding capital: Marx's economic theory*. Harvard University Press.
Friedman, M., & Friedman, M. (1970). *Essays in positive economics*. University of Chicago press.
Gigerenzer G. &Selten R. (2001). Bounded Rationality: The Adaptive Toolbox, Cam bridge, The MIT Press
Jan Blommaert; Chris Bucean (2000). "Critical Discourse Analysis". Annual Review of Anthropology. ***29*** *(1): 447–466. doi:10.1146/annurev.anthro.29.1.447.*
Lachover, E., &Brandes, S. B. (2009). A beautiful campaign? Analysis of public discourses in Israel surrounding the Dove Campaign for Real Beauty. *Feminist Media Studies*, *9*(3), 301-316.
Levine, J. M., Resnick, L. B., & Higgins, E. T. (1993). Social foundations of cognition. *Annual review of psychology*, *44*(1), 585-612.
Li, M., & Chapman, G. B. (2013). Nudge to health: Harnessing decision research to promote health behavior. *Social and Personality Psychology Compass*, *7*(3), 187-198.

Livny, E. (2012). On "Humans" vs. "Econs", the mystery of human psyche, and the domain of economic inquiry. Retrieved from https://iset-pi.ge/index.php/en/iset-economist-blog-2/entry/on-e2-80-9chumans-e2-80-9d-vs-e2-80-9cecons-e2-80-9d-the-mystery-of-human-psyche-and-the-domain-of-economic-inquiry
Luo, A. (2019, August 23). Discourse Analysis: A Step-by-Step Guide with Examples. Retrieved from https://www.scribbr.com/methodology/discourse-analysis/

Luthans, F., &Stajkovic, A. D. (1999). Reinforce for performance: The need to go beyond pay and even rewards. *Academy of Management Perspectives*, *13*(2), 49-57.*doi:10.5465/AME.1999.1899548*
Marcel, A. J. (1983), 'Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition', Cognitive Psychology, 15: 197–237.
Moseley, A., & Stoker, G. (2013). Nudging citizens? Prospects and pitfalls confronting a new heuristic. *Resources, Conservation and Recycling*, *79*, 4-10.
Neslin, S. A., and Van Heerde, H. J. (2009), Promotion dynamics, 3: 177–268. Chan, T., Narasimhan, C., and Zhang, Q. (2008), 'Decomposing promotional effects with a dynamic structural model of flexible consumption', J Mark Res, 45: 487–498. Ni Mhurchu, C., Blakely, T., Jiang, Y ., et al. (2010), 'Effects of price discounts and tailored nutrition education on supermarket purchases: a randomized controlled trial', Am J ClinNutr, 91: 736–747.
Ormrod, J. E. (2012). *Human learning*. Pearson Higher Ed.
Pinker, S. (2003). *How the mind works*. Penguin UK.
Roth, A., Barron, G., Erev, I., &Slonim, R. (2000). *Equilibrium and Learning in Economic Environments: the Predictive Value of Approximations,"*. Harvard University working paper.
Pti. (n.d.). COVID-19: Govt holding online awareness sessions with anganwadi workers: India News - Times of India. Retrieved from https://timesofindia.indiatimes.com/india/covid-19-govt-holding-online-awareness-sessions-with-anganwadi-workers/articleshow/74993332.cms
Roberts, C., Moss, B., Wass, V., Sarangi, S., & Jones, R. (2005). Misunderstandings: a qualitative study of primary care consultations in multilingual settings, and educational implications. *Medical education*, *39*(5), 465-475.
Sande, V. (2016). *Nudge fudge or crystal clear? Investigating nudges and how they work* (Master's thesis, Oslo and Akershus University College of Applied Sciences).
Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American economic review*, 465-479.
E Shaw, S., & Bailey, J. (2009). Discourse analysis: what is it and why is it relevant to family practice?.*Family practice*, *26*(5), 413-419.
Shefrin, H. M., &Thaler, R. H. (1988). The behavioral life-cycle hypothesis. *Economic inquiry*, *26*(4), 609-643.
Simmel, G. (2004). *The philosophy of money*. Psychology Press.
Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, *1*(1), 25-39.
Skinner, B. F. (1958). Reinforcement today. *American Psychologist*, *13*(3), 94.
Stevens, P., & Harper, D. J. (2007). Professional accounts of electroconvulsive therapy: A discourse analysis. *Social Science & Medicine*, *64*(7), 1475-1486.
Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making*, *12*(3), 183-206.

Thaler, R. (2015). Misbehaving: The Making of Behavioural Economics. London: Allen Lane.
https://pdfs.semanticscholar.org/6674/a323b0f9b75657224d40e9d73db13c41b7ab.pdf
Weiwei, W. A. N. G., &Weihua, L. (2015). Critical discourse analysis of news reports on China's bullet-train crash. *Studies in Literature and Language*, *10*(2), 42-49.
Wilkinson, S., &Kitzinger, C. (2000). Thinking differently about thinking positive: a discursive approach to cancer patients' talk. *Social science & medicine*, *50*(6), 797-811.

Zelizer, V. A. (2007). Pasts and futures of economic sociology. *American Behavioral Scientist*, *50*(8), 1056-1069.
Żyłko, B. (2001). Culture and semiotics: Notes on Lotman's conception of culture. *New Literary History*, *32*(2), 391-408. Retrieved from https://muse.jhu.edu/article/24570/summary

DECISION SCIENCES INSTITUTE

An Empirical Study: Impact of Performance-based Contracting Features on Quality Investment

Hasan Uvet
Georgia Gwinnett College
Email: huvet@ggc.edu

Hasan Celik
Robert Morris University
Email: celik@rmu.edu

Sedat Cevikparmak
DeSales University
Email: sedat.cevikparmak@desales.edu

Saban Adana
John Carroll University
Email: sadana@jcu.edu

ABSTRACT

This study investigates the impact of performance-based contracting (PBC) features – incentive alignment, goal congruence, joint knowledge generation- on quality investment of firms by collecting data from suppliers, logistics managers, and operation managers. In this study, after analyzing the validity and reliability of instruments through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), we conducted structural equation modeling for inferential analysis. We also examined the effects of information sharing and collaborative communication on joint knowledge generation. We found a positive relationship between PBC features and quality investment.

KEYWORDS: Performance-based contracting, quality investment, joint knowledge generation, goal congruence, incentive alignment

INTRODUCTION

Different from transaction-based contracts, in performance-based contracting (PBC) suppliers, are rewarded when they get the performance objectives in the contract (Berkowitz et al., 2005; Geary & Vitasek, 2005). In PBC, the supplier determines how to achieve the performance objectives defined in the contract by the buyer (Kim, Cohen, & Netessine, 2007). Considering supplier's decisions for making an upfront investment to increase reliability or holding spare parts to increase the readiness of the systems to achieve desired availability rates in PBC, suppliers are motivated to make an upfront investment to increase the reliability of systems (Gansler & Lucyshyn, 2006; Vitasek et al., 2007). This upfront investment by suppliers improves affordability over the remaining life of the supported system which increases financial benefits for suppliers (Kim, Cohen, & Netessine, 2007; Randal, Pohlen, & Hanna, 2010, Uvet, Celik, & Kucuk, 2020).
In literature, many studies are conducted about a trade-off between spare parts and reliability in PBC through mathematical models (Selviaridisa & Wynstrab, 2015). However, there is a significant gap in PBC for experimental studies. The primary purpose of this study is to understand the impact of PBC features – joint knowledge generation, incentive alignment, goal congruence – amongst suppliers on quality investment in PBC. The other goal of this study is to examine the

impacts of collaborative communication and information sharing between suppliers on joint knowledge generation. The primary research question of this study is 'What is the impact of PBC features on quality investment? This paper is organized as follows. The literature review will be presented in Section 2. In Section 3, research methodology and data analysis are represented. The paper will be concluded with a discussion of limitations, future research, and conclusion.

LITERATURE REVIEW

In literature, performance-based contracting also called performance-based logistics contracts, outcome-based contracting, or "power by the hour" contracting. The studies in PBC can be classified under the three major categories. The first category is conceptual papers, which investigates the enablers, barriers, and features of PBC (Berkowitz et al., 2005; Richardson & Jacopino, 2006; Vitasek & Gear, 2007; Sols et al., 2007; Ng et al., 2009; Straub, 2007; Randall et al., 2010; Sols & Johannesen 2013; Randall et al., 2014). The authors in this category highlighted the importance of aligning goals with measurable performance metrics, team environment/partnerships to generate innovative solutions, and effective reward mechanisms with goals and incentives in PBC. The second category in PBC can be classified with case studies and interview-based experimental studies in which the authors investigated the success of PBC (Vitasek et al., 2006; Straub, 2009; Datta & Roy, 2011; Randall et al., 2012; Glas et al., 2013; Selviaridis & Norrman, 2015; Caldwell &Howard, 2014; Randall et al., 2015; Uvet et al., 2020). The authors in this category highlighted the importance of long-term contracts which enables us to translate knowledge into innovation, an early supplier involvement for product improvements, a life-cycle support perspective, transferring the responsibility, and risks to the contractor. The third category in PBC can be categorized with mathematical models that investigate PBC contract features such as length of contracts, the trade-off between reliability investment and spare parts inventory levels (Kim et al., 2007; Kim et al., 2010; Jin & Tian, 2012; Jin et al., 2014; Kim, Cohen & Netessine, 2017; Bakshi et al., 2015; Uvet et al., 2020). In this category, authors found that higher benefits of reliability investment rather than increasing spare parts, the importance of long term of the contract in which suppliers can compensate their investments to increase reliability/quality of products/systems. Although numerous studies were done in PBC, still we lack empirical studies that investigate the impact of PBC features lead to highly reliable products. So, in this study, based on the finding in conceptual papers in literature, we conducted an empirical study to investigate the impact of PBC features on quality investment.

THEORETICAL DEVELOPMENT and RESEARCH MODEL

Due to the importance of collaboration of suppliers to build high-quality products/systems, this study focuses on the impacts of PBC features on quality investment from the perspectives of the resource-based view and relational view. Resource-based view highlights the importance of core competencies and capabilities (Barney, 1991; Prahalad & Hamel, 1990) to get a competitive advantage (Park et al. 2004). The relational view argues that critical resources can attain through collaborative partners (Dyer & Singh, 1998). Based on the relational view sharing and exchanging of knowledge and resources will lead to attaining relational rents to increase the value offerings for customers (Lavie, 2006).

A supply chain consists of partnerships that are the source of knowledge and skills to create value propositions for performance-based outcomes (Vargo & Lusch 2004; Randall et al. 2011). This collaboration of entire supply network entities—involving processes, material, and information around the same goals—is critical for making investments to get a competitive advantage (Cao &d Zhang, 2010; 2011). In this study, we highlighted the major PBC features based on the conceptual papers in literature. These features are joint knowledge generation, goal congruence,

and incentive alignment, which are the major enablers of PBC. Besides, we looked into the information sharing and collaborative communication between suppliers for joint knowledge generation. While defining each dimension, we also emphasize the importance of each component for quality investment.

Goal congruence between supply chain partners is the extent to which each supplier perceives their targets are realized by achieving the supply chain objectives (Cao & Zhang, 2011). Goal congruence is one of the primary PBC features for upfront investments to increase reliability growth of products/systems.

Incentive alignment refers to sharing risks and benefits among suppliers (Cao & Zhang, 2011). Considering the incentive structure of PBC by reward mechanism by getting targeted outcomes, incentive alignment appears major enablers of investment decisions in PBC.

Joint knowledge generation refers to the extent to which a supplier works with supply chain partners to search, explore, and acquire new and relevant knowledge (Cao & Zhang, 2011). Knowledge is the source of value creation that can be attained by inter-firm supply chain relationships (Ng et al., 2009; Randall et al., 2010; Randall et al., 2012). The success of suppliers to achieve desired performance targets by quality investments depends on the knowledge/skills of all supply chain partners.

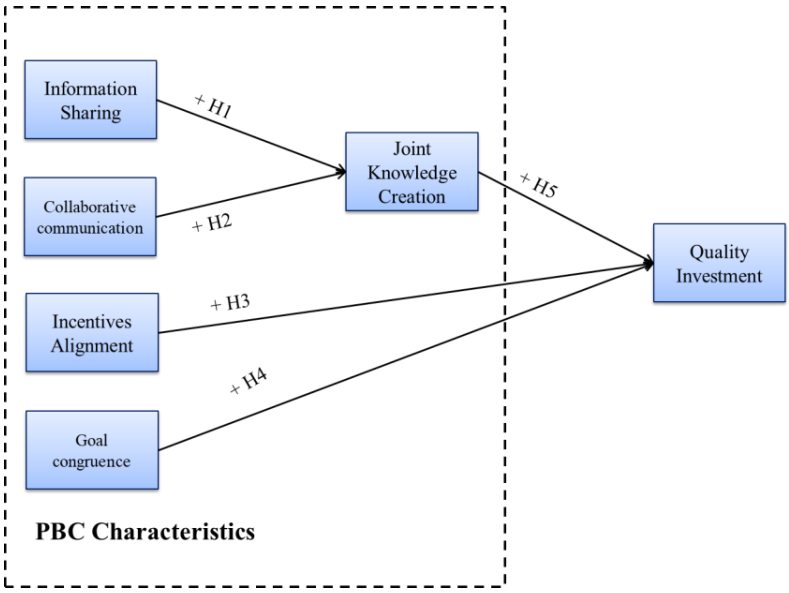
Collaborative communication refers to message transmission between supply chain partners. Frequent contacts of suppliers and the existence of open and two-way communication channels are the main items of collaborative communication in the supply chain. The existence of communication channels within suppliers is an essential element to make knowledge and quality investment to increase the corporate image.

Information sharing refers to the prompt sharing of appropriate, accurate, and relevant information within supply chain partners (Cao et al., 2010). In the supply chain, information sharing is not only critical for improving responsiveness but is also essential for creating new knowledge to existing problems in PBC. Thus, in the supply chain, information sharing appears as an essential requirement for joint knowledge generation (Kleemann & Essig, 2013).

Quality investment refers to the supplier's investment for high-quality products. Highly reliable products are one of the significant findings in literature in PBC. Suppliers can avoid after-sale support costs, such as maintenance and repair, by investing the reliability of products/systems. Additionally, this upfront investment will result in high availability of products/systems which is one of the major performance metrics that is used on PBC. Although there are many studies regarding various aspects of PBC, based on the above arguments, this study hypothesized the impact of these PBC features on quality investment and the impact of information sharing and collaborative communication on joint knowledge generation (see Figure 1).

Hypothesis 1 (H1): Information sharing has a positive effect on joint knowledge generation.
Hypothesis 2 (H2): Collaborative communication has a positive effect on joint knowledge generation.
Hypothesis 3 (H3): Incentive alignment has a positive effect on quality investment.
Hypothesis 4 (H4): Goal congruence has a positive effect on quality investment.
Hypothesis 5 (H5): Joint knowledge generation has a positive effect on quality investment.

Figure 1. Research Model



DATA ANALYSIS

Data Collection

In this study, data were collected in the United States through a survey targeting single informants whose primary job functions are logistics, supply chain management, or operations management. In this research, the Amazon Mechanical Turk (MTurk) platform was used to find a suitable sample frame. At the beginning of the survey, screening questions were asked to limit participation to those who hold the title of manager and upper levels at manufacturing firms and whose job function is supply chain management, operations management, or logistics management. Based on these screening questions (job function and position level) and after cleaning the data, 381 responses are used in the data analysis.

Data Analysis: Results of EFA and CFA

In this study, scale items were adapted based on the existing literature (Cao et al., 2010). After analyzing the validity and reliability of instruments through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), we conducted a structural equation modeling for inferential analysis. Before the analysis, principal component analysis (PCA) with varimax rotation was employed to test the discriminant and convergent validity of the instrument. Inspection of the correlation matrix indicated that all variables had at least one correlation coefficient greater than 0.3. Bartlett's Test of Sphericity was statistically significant ($p < 0.005$), thus indicating that the data was conducive to running PCA. The overall Kaiser-Meyer-Olkin (KMO) measure (0.921) was higher than 0.9, which demonstrates a degree of common variance that is "marvelous" according to the scale developed by Kaiser. After deleting the factor loadings lower than 0.5, in the final PCA, the data exhibited loadings bigger than 0.65 were kept in the analysis. The final component loadings are presented in Table 1.

Table 1. Rotated Component Matrix[a]

	Component					
	1	2	3	4	5	6
JKG_1	.766					
JKG_4	.740					
JKG_3	.730					
JKG_2	.729					
JKG_6	.720					
JKG_5	.698					
QI_3		.771				
QI_5		.756				
QI_2		.755				
QI_4		.741				
QI_1		.672				
CC_3			.822			
CC_4			.705			
CC_2			.693			
CC_1			.663			
GC_2				.761		
GC_3				.745		
GC_1				.666		
IA_3					.831	
IA_2					.771	
IA_1					.740	
IS_1						.761
IS_3						.719
IS_2						.698

IA: Incentive Alignment; JKG: Joint Knowledge Generation; QI: Quality Investment; GC: Goal Congruence; IS: Information Sharing; CC: Collaborative Communication

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

In CFA results, the overall fit indices for the measurement model were acceptable and demonstrated good unidimensionality. (Anderson & Gerbing, 1988; Bagozzi & Yi, 1988). These fit indices were: a Joreskorg and Sorbom (1993) GFI of 0.908, a Bentler (1990) CFI of 0.963, an AGFI of 0.895, a NNFI of 0.955, a RMSEA of 0.045, a normed chi-square (χ^2) of 1.767, and a chi-square (χ^2) of 402.914 with 228 degrees of freedom (see Table 2). Second, as evidence of convergent validity, measurements of factor loadings were all significant at $p < 0.01$, based on t-values, and all standardized item loadings (λs) ranging from 0.667 to 0.852 were above 0.5 (Hair

et al., 2010). Estimates of AVEs for four factors, ranging from 0.545 to 0.596, were greater than the critical value of 0.50 (Fornell & Larcker, 1981). The variance captured by the construct was greater than the variance due to measurement error. Scale reliability measurements were conducted using construct reliability. Construct reliabilities (CR) for four factors, ranging from 0.782 to 0.883, were above the critical value of 0.70 (Hair et al., 2010) (see Table 2).

Table 2. Confirmatory factor analysis results

Construct	Items	Loadings (λ)	t-value	CR	AVE
Goal Congruence				0.805	0.579
	GC1	0.767	14.021		
	GC2	0.761	13.920		
	GC3	0.755	(set to 1.0)		
Incentive Alignment				0.815	0.596
	IA1	0.725	9.067		
	IA2	0.852	10.043		
	IA3	0.733	(set to 1.0)		
Collaborative Communication				0.848	0.584
	CC1	0.667	13.268		
	CC2	0.792	14.809		
	CC3	0.780	(set to 1.0)		
	CC4	0.809	15.183		
Information Sharing				0.782	0.545
	IS1	0.707	12.863		
	IS2	0.778	(set to 1.0)		
	IS3	0.727	13.203		
Joint Knowledge Generation				0.883	0.557
	JKG1	0.734	12.748		
	JKG2	0.772	12.784		
	JKG 3	0.727	15.427		
	JKG 4	0.756	(set to 1.0)		
	JKG 5	0.738	12.887		
	JKG 6	0.751	12.879		
Quality Investment				0.879	0.593
	QI1	0.759	15.198		
	QI2	0.814	16.452		
	QI3	0.783	(set to 1.0)		
	QI4	0.713	15.414		
	QI5	0.777	15.603		

CFA global fit indices: Chi-square=402.914; df=228; Normed Chi-square=1.767; GFI=0.921; AGFI=0.895; NFI=0.918; NNFI=0.955; CFI=0.963; RMSEA=0.045.

The procedure demonstrated by Fornell and Larcker (1981) was used to analyze for discriminant validity. The overall fit indices for the measurement model were also acceptable and proved good unidimensionality. As presented in Table 3, a squared correlation between any two constructs is higher than either of the constructs' AVE. As seen in Table 1 and 3, discriminant validity is supported. These results indicate acceptable levels of internal consistency, convergent validity, discriminant validity, and construct validity (Hair et al., 2010).

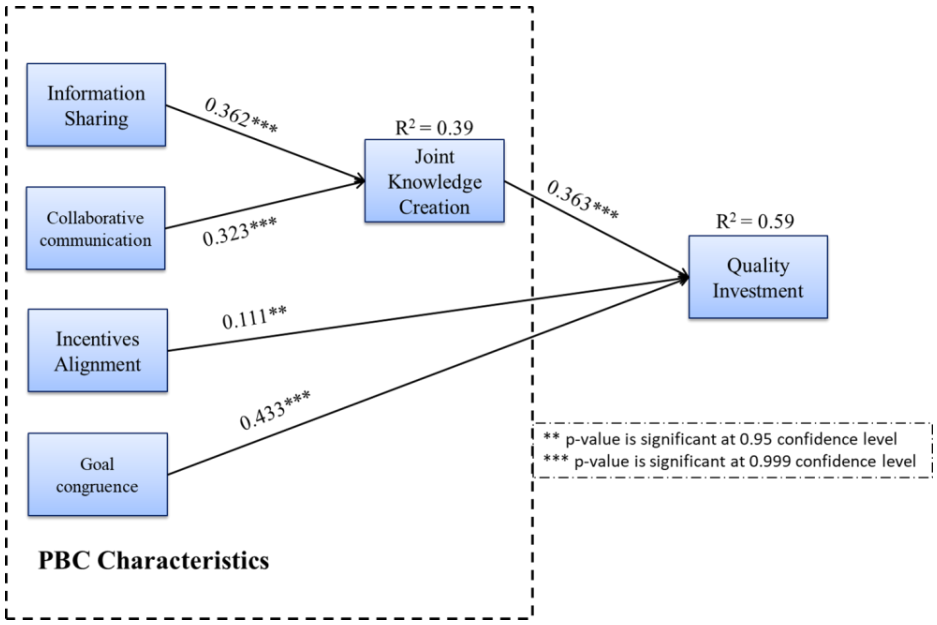
Table 3. Construct Reliability and discriminant validity analysis

	CR	AVE	IA	KI	QI	GC	IS	CC
IA	0.815	0.596	**0.772**					
KI	0.883	0.557	0.518	**0.746**				
QI	0.879	0.593	0.457	0.637	**0.770**			
GC	0.805	0.579	0.436	0.535	0.648	**0.761**		
IS	0.782	0.545	0.457	0.546	0.608	0.761	**0.738**	
CC	0.848	0.584	0.362	0.560	0.617	0.712	0.670	**0.764**
IA: Incentive Alignment; JKG: Joint Knowledge Generation; QI: Quality Investment; GC: Goal Congruence; IS: Information Sharing; CC: Collaborative Communication								

Data Analysis: Results of SEM Analysis:

In this study, structural equation modeling (SEM) used to explain the relationship between variables. SEM coefficients related to the relationship between endogenous and exogenous variables can be seen in Figure 2. Fit indices for the model indicate a satisfactory fit (c2= 437.877, df = 232, GFI = 0.914, AGFI = 0.889, NNFI = 0.948, CFI = 0.956, RMSEA = 0.048). Hypothesis in the research model are supported. The AMOS path coefficient shown in Figure 2 shows that the exogenous variable of joint knowledge generation incentive alignment and goal congruence is significant and positive predictors of quality investment in PBC at the 95% confidence level. Structural modeling revealed an R-square of 0.39 and 0.59 for joint knowledge generation, quality investment, respectively. The results of the SEM analysis are presented in Figure 2. The results suggest that the variables of joint knowledge generation, incentive alignment, goal congruence were significant predictors of a quality investment. Additionally, collaborative communication and information sharing were found as significant predictors of joint knowledge generation.

Figure 2. Results of SEM Analysis



DISCUSSION AND CONCLUSIONS

Considering achievements, such as %25-%40 reliability growth (Guajardo et al., 2012) and a decrease in total life-cycle cost for sustaining a system (Gansler & Lucyshyn, 2006; Mahon, 2007). in PBC arrangements, it is getting critical to explore major drivers of PBC. So, in this paper, after building our research model that shows major drivers of quality investment in PBC, we explored the inferential relationship between variables with SEM analysis. We found that the incentive structure of the PBC reward mechanism creates goal congruence, incentive alignment, and knowledge generation within suppliers. The effects of these variables on quality investment were found as a significant positive indicator of a quality investment. Considering %59 (R2) of variability in quality investment can be explained with these three variables, still we can extend this study by exploring other variables in PBC that lead this quality investment. In this study, we also found that information sharing and collaborative communication are positive significant indicators of knowledge generation within suppliers which seen as a major success of innovative solutions in PBC (Randall et al., 2014). We can extend this study by investigating the impact of quality investment on the financial performance of firms. Also, in future studies, we can use industry types as categorical variables and we can compare the impact of these variables in different sectors.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin, 103(3), 411.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. Journal of the academy of marketing science, 16(1), 74-94.

Bakshi, N., Kim, S. H., & Savva, N. (2015). Signaling new product reliability with after-sales service contracts. Management Science, 61(8), 1812-1829.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological bulletin, 107(2), 238-246.

Berkowitz, D., Gupta, J. N., Simpson, J. T., & McWilliams, J. B. (2005). Defining and implementing performance-based logistics in government. Defense Acquisition Review Journal, 11(3), 255-267.

Caldwell, N., & Howard, M. (2014). Contracting for complex performance in markets of few buyers and sellers. International Journal of Operations & Production Management, 34(2), 270-294.

Cao, M., Vonderembse, M. A., Zhang, Q., & Ragu-Nathan, T. S. (2010). Supply chain collaboration: conceptualisation and instrument development. International Journal of Production Research, 48(22), 6613-6635.

Cao, M., & Zhang, Q., (2011). Supply chain collaboration: impact on collaborative advantage and firm performance, Journal of Operations Management, 29 (3) (2011), 163-180.

Datta, P. P., & Roy, R. (2011). Operations strategy for the effective delivery of integrated industrial product-service offerings: Two exploratory defence industry case studies. *International* Journal of Operations & Production Management, 31(5), 579-603.

Dyer, J.H., & Singh, H. (1998). The relational view: cooperative strategy and sources of interorganizational competitive advantage. Academy of Management Review, 23(4), 660–679.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. Journal of Marketing Research, 382-388.

Gansler, J. S., & Lucyshyn, W. (2006). Evaluation of performance based logistics. Doctoral dissertation, Naval Postgraduate School. Retrieved from http://www.acquistionresearch.org/_files/fy2006/umd-lm-06-040.pdf

Geary, S., & Vitasek, K. (2005). It's a war of ideas. Aviation Week & Space Technology, 163(17), 57.

Glas, A., Hofmann, E., & Essig, M. (2013). Performance-based logistics: A portfolio for contracting military supply. International Journal of Physical Distribution and Logistics Management, 43(2), 97–115.

Guajardo, J. A., Cohen M. A., Kim S.H., & Netessine S. (2012). Impact of performance-based contracting on product reliability: An empirical analysis. Management Science 58 (5), 961-979.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis (7 ed.). Upper Saddle River, NJ: Prentice Hall.

Jin, T., & Wang, P. (2012). Planning performance based contracts considering reliability and uncertain system usage. Journal of the Operational Research Society, 63(10), 1467-1478.

Jin, T., Tian Z., Xie M. (2014). A Game-Theoretical Approach For Optimizing Maintenance, Spares And Service Capacity In Performance Contracting Optimizing Maintenance, Spares And Service Capacity In Performance Contracting. International Journal Production Economics, 161, 31-43.

Joreskog, K.G, & Sorbom, D., (1986). LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods. Scientific Software, Inc., Moorsville, IN.

Kim, S.H., Cohen, M.A., & Netessine, S. (2007). Performance contracting in after-sales service supply chains. Management Science 53 (12), 1843–1858.

Kim, S. H., Cohen, M. A., Netessine, S., & Veeraraghavan, S. (2010). Contracting for infrequent restoration and recovery of mission-critical systems. Management Science, 56(9), 1551-1567.

Kim, S. H., Cohen, M. A., & Netessine, S. (2017). Reliability or inventory? An analysis of performance-based contracts for product support services. In Handbook of Information Exchange In Supply Chain Management, Springer, Cham, 65-88,

Kleemann, F. C., & Essig, M. (2013). A providers' perspective on supplier relationships in performance-based contracting. Journal of Purchasing and Supply Management, 19(3), 185-198.

Lavie, D. (2006). The competitive advantage of interconnected firms: an extension of the resource-based view. Academy of Management Review, 31(3), 638–658.

Mahon, D. (2007). Performance-based logistics: transforming sustainment. Journal of Contract Management, 5(1), 53-71.

Ng, I. C., Maull, R., & Yip, N. (2009). Outcome-based contracts as a driver for systems thinking and service-dominant logic in service science: Evidence from the defence industry. European Management Journal, 27(6), 377-387.

Park, N.K., Mezias, J.M., & Song, J. (2004). A resource-based view of strategic alliances and firm value in the electronic marketplace. Journal of Management, 30(1), 7–27.

Prahalad, C.K., & Hamel, G. (1990). The core competence of the corporation. Harvard Business Review, 68(3), 79–91.

Randall, W. S., Pohlen, T. L., & Hanna, J. B. (2010). Evolving a theory of Performance-based Logistics using insights from service-dominant logic. Journal of Business Logistics, 31(2), 35-61.

Randall, W. S., Nowicki, D. R., & Hawkins, T. G. (2011). Explaining the effectiveness of performance-based logistics: a quantitative examination. The International Journal of Logistics Management, 22(3), 324-348.

Randall, W. S., Brady, S. P., & Nowicki, D. R. (2012). Business case analysis and the confounds of innovation driven by performance-based postproduction support strategies. Transportation Journal, 51(1), 33-58.

Randall, W. S., Nowicki, D. R., Deshpande, G., & Lusch, R. F. (2014). Converting knowledge into value: gaining insights from service dominant logic and neuroeconomics. International Journal of Physical Distribution & Logistics Management, 44(8/9), 655-670.

Randall, W. S., Hawkins, T. G., Haynie, J. J., Nowicki, D. R., Armenakis, A. A., & Geary, S. R. (2015). Performance-Based Logistics and Interfirm Team Processes: An Empirical Investigation. Journal of Business Logistics, 36(2), 212-230.

Richardson, D., & Jacopino, A. (2006, January). Use of R&M measures in Australian defence aerospace performance based contracts. In Reliability and Maintainability Symposium, 2006. RAMS'06. Annual, IEEE, 331-336.

Selviaridis, K., & Wynstra, F. (2015). Performance-based contracting: a literature review and future research directions. International Journal of Production Research, 53(12), 3505-3540.

Sols, A., & Johannesen, L. H. (2013). The role of transition contracts in performance-based logistics initiatives. Systems Engineering, 16(4), 453-463.

Sols, A., Nowicki, D., & Verma, D. (2007). Defining the fundamental framework of an effective Performance-based logistics (PBL) contract. Engineering Management Journal, 19(2), 40-50.

Straub, A. (2007). Performance-based maintenance partnering: a promising concept. Journal of Facilities Management, 5(2), 129-142.

Straub, A. (2009). Cost savings from performance-based maintenance contracting. International Journal of Strategic Property Management, 13(3), 205-217.

Uvet, H., Celik, H., and Kucuk, C.Y. (2019). Understanding the Impacts of Length of the Contract and Fleet Size on Spare Parts Level and Reliability Investments in Performance-based Contracting, 12(4), 225-236.

Uvet, H., Celik, H., Cevikparmak, S., & Adana, S. (2020). Supply chain collaboration in performance-based contracting: an empirical study. International Journal of Productivity and Performance Management, https://doi.org/10.1108/IJPPM-01-2019-0008.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. Journal of marketing, 68(1), 1-17.

Vitasek, K., Geary, S., & Quick, R. (2006). Performance-based logistics: The basics and beyond. Report presented at The International Logistics Society's 41st Annual International Logistics Conference & Exhibition, Dallas, TX.

Vitasek, K., Cothran, J., & Rutner, S. (2007). Best practices of public private partnering: Developing a win-win strategy for working with depots (White paper). Knoxville, TN: The University of Tennessee and Supply Chain Visions.

Vitasek, K., & Geary, S. (2007). Performance-Based Logistics—Next Big Thing. ProLogis Supply Chain Review, Summer, 1-11.

DECISION SCIENCES INSTITUTE
Stakeholders' Impact on Sustainable Suppliers in Factories

Osam Sato
Tokyo Keizai University
Email: osamsato@tku.ac.jp

Yoshiki Matsui
The Open University of Japan and Yokohama National University
Email: ymatsui@ynu.ac.jp

ABSTRACT

Sustainable supply chain management (SSCM) is one of the most important themes for the supply chain management. Factories (plants) request SSCM effort, such as environmental protection, to upper stream suppliers. Top management in a factory encourages suppliers to be sustainable to avoid sustainability risk. However, some other stakeholders affect suppliers to be sustainable either directly or indirectly. Employees affect suppliers directly in their usual business through daily interaction. Government and customers affect the effort to make suppliers sustainable through top management indirectly. We conducted the model as an empirical study with a set of data in hand.

KEYWORDS: Supply chain management, Supplier, Empirical method, Environment, Manufacturing

INTRODUCTION

Sustainability is "the quality of being able to continue over a period of time" (Cambridge University Press, 2005). Sustainability has three dimensions, economic, environmental, and social. Sustainable performance includes "employee satisfaction, quality and inventory performance, environmental performance, and financial and market performance" in a broad sense (Kaynak & Montiel, 2009).

A supply chain is "a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and information from a source to a customer" (Mentzer et al., 2001). Supply Chain Management (SCM) has "three categories: a management philosophy, implementation of a management philosophy, and a set of management processes" (Mentzer et al., 2001). We define SCM as integrating the business processes of managing relationships, information, and materials flow across enterprise borders to deliver enhanced customer service and economic value through synchronized management of physical goods and associated information from sourcing to consumption (La Londe, 1997). We focus on this study on "SCM as implementation of a management philosophy" (Mentzer et al., 2001).

We define sustainable supply chain management (SSCM) as SCM that focus on sustainability. SSCM is one of the most important themes for supply chain management (SCM) today (Tebaldi et al., 2018). Factories (plants) conducted internal SSCM activities themselves. Also, factories request SSCM effort, such as environmental protection, to upper stream suppliers, as a part of the Life Cycle Assessment (LCA) (ISO, 2006). It is essential because if materials and components that the factory used are environmentally unfriendly, it brings a penalty to the factory and the company, and may lose market share. For example, Villena and Gioia (2020)

warned that Multi-National Companies (MNC) must watch and take care of the sustainability level not only of first-tier but also lower-tier suppliers because of sustainability effort of them is weak. EU Commission published the Product Environmental Footprint (PEF) and Organization Environmental Footprint (OEF) methods (EU, 2013) to reject environmentally unfriendly products and companies from the EU market. ISO14040 series formalized the policy and standards (ISO, 2006).

SCM includes the upper stream (supplier), internal (within a factory or plant), and downstream (customer) in the definition (Flynn et al., 2010). SSCM also includes these three parts. The supply side of SSCM is written as sustainable supply management (SSM) (Paulraj et al., 2012; Koster et al., 2017).

Therefore, factories must push suppliers toward sustainability and help their establishment. However, it is not apparent nor straightforward because it needs many efforts and resources on both sides. Our research motivation is finding human drivers in buyers. Formally, top management has authority and responsibility, but some other stakeholders affect the encouragement of sustainable suppliers because they mind environmental pollution problems. We want to find who and how much each stakeholder affects the plant activities to encourage sustainable suppliers. Activities that promote sustainability to suppliers include green purchasing, a blacklist of raw materials, eco-design, green packaging, recyclable materials, parts, etc.

We focus on and limit to suppliers, i.e., the upper stream environmental aspect of SSCM, in this study. Botes et al. (2017) wrote: "Upstream supply disruptions often have a severe impact on a firm's cost base, revenue, and shareholder value."

LITERATURE REVIEW

In a broad sense or from a social point of view, the sustainability movement, especially the environmental protection movement in our research focus, promotes stakeholders in the supply chain to establish SSCM. Institutional theory in organizational studies is the essential sociological background of our research model. "Institutional theory suggests social movement actors can promote shared notions of the kinds of organizational activities that are 'right' through proselytizing and other techniques of moral suasion" (David et al., 2019). Therefore, the environmental protection movement (homo sociologicus) lets stakeholders move to sustainability. As a result, environmental protection movement grows as a concern for every stakeholders (government, customers, employees and managements) even in a factory (as a homo economicus entity) (David et al., 2019).

Based on the above theoretical framework, we extend our study from an operational point of view. There are many academic researches about SSCM (Carter & Rogers, 2008; Kaynak & Montiel, 2009; Carter & Easton, 2011; Gualandris & Kalchschmidt, 2014; Gupta & Palsule-Desai, 2011; Campos et al, 2017; Koster et al, 2017; Tebaldi et al, 2018; Xu & Cong, 2011; Dubey et al, 2019).

Dubey et al. (2019) conducted an empirical study that explains supplier relationship management (SRM) with external pressure to the circular economy (CE), i.e., environment or sustainability. Dubey et al. (2019) insisted that top management commitment (TMC) as a mediating factor between external pressure and SRM. Certainly, TMC is an essential driver of the introduction of SSCM practices. "Nothing is possible without leadership from the top of the organization" (Christopher & Peck, 2004). However, more stakeholders may affect the introduction of SSCM in a factory (plant). We believe TMC is just one of some stakeholders in the introduction of upper stream SSCM. According to the literature, the primary drivers are top management, employees, governmental regulation, and customers' pressure. Therefore, we need to include these significant stakeholders in our empirical study.

However, so far as our literature survey, we experience a paucity of the empirical study of influencers who encourage suppliers to introduce SSCM in a factory (Melnyk et al., 2014). This study tries to fill the gap with the application of our data in hand.
The introduction of SSCM is not easy. It needs a lot of effort from many stakeholders and resources, not only top management but also employees and customers. Customers request a certain level of green product with quality to which they pay. Customers want environmentally friendly material or parts because it is a part of quality. Governmental regulation and pressure from NPO (Non-Profit Organizations) and NGO (Non-Government Organizations) are essential. "Firms need to consider other non-market stakeholders such as the community, non-governmental organizations, and the media" (Kaynak & Montiel, 2009). Tebaldi et al. (2018) defined stakeholders' theory as "a conceptual framework of business ethics and organizational management which addresses moral and ethical values in the management of a business or other organization." We conduct an influencer analysis based on institutional theory and stakeholder theory concept. Stakeholder theory is "a conceptual framework of business ethics and organizational management which addresses moral and ethical values in the management of a business or other organization" (Tebaldi et al., 2018).
We included NPO and NGO in governmental regulators in this study.

RESEARCH HYPOTHESES

The relative importance of stakeholders and the pattern of influence are different because of the relationship with suppliers and differences of interest. Top management and employees affect the introduction of SSCM by suppliers directly through business contracts and negotiation. However, government and customers affect the institutional activities of suppliers indirectly. A typical example of indirect governmental intervention is Product Environmental Footprint (PEF) and ISO 14040 by the EC Commission, as introduced above. They affect the introduction of SSCM through recognition by top management. Employees interact with suppliers in everyday work and collaboration. However, both government and customers have no direct connection to suppliers. Therefore, we categorize stakeholders as direct (top management and employees) and indirect (government and customers) who affect top management.
Social constructionism is the base of consideration of recognition by stakeholders (i.e., top management, customers, and employees) theoretically (Galbin, 2014). Social pressure to sustainability consists of government regulation and pressure from customers and other stakeholders. Institutional theory suggests that top management feels pressure from both government and customers, which motivates top management to act to encourage suppliers to introduce SSCM.
We set five hypotheses based on the above inference.
National and local government bodies request or regulate businesses and factories to participate and cooperate with environmental adaptation initiatives of local governments (e.g., The Committee on Approaches to Climate Change Adaptation, 2010). If the pressure from the government is felt so strong and has an impact on business, top management decides to pour more resources, both human and financial, to solve the problem and evade the risk. If top management feels that environmental regulation by the government becomes intense, the factory must adapt to it.
H1: The more a plant feels pressure from governments (including NGO and NPO), the more top management encourage suppliers to SSCM practices.
As environmental pollution problems become more extensive, the voice of Eco-Conscious Consumers and Green Consumers become larger. "Eco-Conscious Consumers are defined as internet users aged 16-64 who say that concern for the environment affects their day-to-day buying decisions" (Young, 2018). Green Consumers are "a customer who wants to buy things

that have been produced in a way that protects the natural environment" (Cambridge Business English Dictionary, 2011). The larger the voice of them, the more top management needs to consider the request.

H2: The more a plant feels pressure from customers, the more top management encourages suppliers to SSCM practices.

If top management persuades or order employees to extend SSCM, employees' valuation to environmental protection become more prominent and more active.

H3: The more top management encourages suppliers to SSCM practices, the stronger the employees' value.

The SCM activities include (1) integrated behavior, (2) mutually sharing information, (3) mutually sharing risks and rewards, (4) cooperation, (5) the same goal and the same focus on serving customers, (6) integration of processes, and (7) partners to build and maintain a long-term relationship (Mentzer et al., 2001). We can apply these activities to SSCM too. Employees in a factory proceed internal SSCM activities such as lean manufacturing, eco-design, green packaging, recyclability, based on the order from their boss and top management. Empowered employees extend SSCM to suppliers on the base of their value or belief.

Factories usually extend internal SCM to both upper stream suppliers and downstream customers based on the above SCM activities. Some prior studies concluded that internal SCM is a prerequisite of external SCM (Flynn et al., 2010). The impact of employees to encourage sustainable suppliers is more direct than that of government because of daily interaction with suppliers. Social constructionism suggests that the strength of effects depends on the value and belief of employees about sustainability and contribution to environmental protection. "Individuals' perceptions and interpretations of reality are influenced by the social construction of their environment, based on the actions and words of their co-workers" (Boh & Won, 2013).

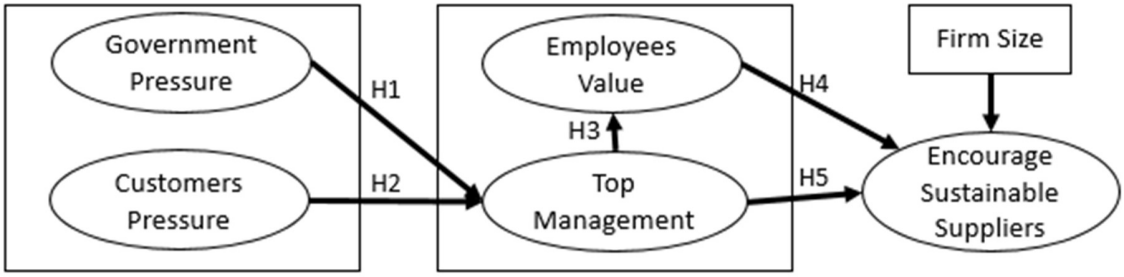
H4: The more employees value sustainability, the more employees encourage sustainable suppliers.

The top management has the responsibility and is the decision-maker who represents the factory externally. Therefore, the level of impact of the factory to encourage suppliers to sustainability depend on how top management thinks of and acts to promote sustainable suppliers.

H5: The more top management pushes the introduction of SSCM, the more top management encourages sustainable suppliers.

The above five hypotheses imply positive relationships between constructs, but we have no idea whether the relationships are linear or higher-order. We assume linear relationships for simplicity. We include firm size as a control variable by following Dubey et al. (2019). Figure 1 shows our research model.

Figure 1: Research Model



RESEARCH DESIGN

Operationalization of Constructs

We employ data from High-Performance Manufacturing (HPM) round 4. Table 1 shows the questions that we hired for each construct. Except for the control variables examining firm size (FS), all scales were five-point anchored with 1= No extent whatsoever, 3=Moderate extent, and 5= Very great extent. We assess FS by the logarithmic values of the number of employees in a plant (Dubey et al., 2019). Environmental Affairs staff in a plant responded to these questions.

Table 1: Constructs and Items (Questions)	
Encourage sustainable suppliers	
SSCM	Please indicate the degree to which your plant is engaged in the following initiatives/practices:
SSCM1	Encouraging suppliers to improve the environmental performance of their processes
SSCM2	Requesting that your suppliers sign a code of environmental conduct
SSCM3	Incorporating environmental considerations in evaluating and selecting suppliers
SSCM4	Providing design specification to suppliers in line with environmental requirements (e.g., green purchasing, a blacklist of raw materials)
SSCM5	Co-development with suppliers to reduce the environmental impact of the product (e.g., eco-design, green packaging, recyclability)
Motivation by Stakeholders	
Stakeholders	My plant's involvement in environmental initiatives has been motivated by:
Top Management leadership	
Top Management1	The examples top management provides
Top Management2	Requirements made by senior management
Top Management3	Top-down initiatives
Top Management4	Top management's commitment to environmental responsibility
Customers' preference	
Customers1	Programs that our customers have in place
Customers2	Customers who seek environmentally responsible suppliers
Customers3	Increased awareness of environmental issues among our customers
Customers4	Customers who believe that environmental protection is important
Employees value	
Employees1	The morals of individual employees
Employees2	The personal desires of employees to do what is right
Employees3	A personal sense of obligation among employees
Employees4	The underlying values of employees
Government regulation	
Government1	Current government legislation
Government2	The threat of future government legislation
Government3	Industry or government regulation
Government4	Regulations dealing with the environment

DATA ANALYSES AND RESULTS

Data

We employ the fourth round dataset of the High-Performance Manufacturing (HPM) project (Schroeder & Flynn, 2001; Phan, 2014) because it is the latest version available. Each team in our international project visit representative plants and ask/leave a set of the standardized questionnaire in each country. Multiple staff in the factory responded to the questionnaire and returned them to our teams. These teams collected the data from factories in electric & electronics, machinery, and transportation equipment. Transportation equipment factory includes automobile assemblers and their parts manufacturers. The fourth round of the HPM project started in 2012-2013, and we collected data in 2014-2015. Table 1 shows the distribution of sample size by country. The total sample size amounts to 306.

Table 2: Data distribution by country

Country	Count	%	Country	Count	%
Brazil	24	7.8	Japan	22	7.2
China	30	9.8	Korea	26	8.5
Spain	25	8.2	Sweden	9	2.9
Finland	17	5.6	Taiwan	30	9.8
Germany	28	9.2	United Kingdom	13	4.2
Israel	26	8.5	Vietnam	25	8.2
Italy	29	9.5	missing	2	0.7
			Total	306	100

Table 3 shows the sample size by industry.

Table 3: Data distribution by industry

Industry	Count	%
Electronics	114	37.3
Machinery	110	35.9
Transportation	80	26.1
Missing	2	0.7
Total	306	100.0

Preliminary Data Analysis

First, we conducted exploratory factor analysis (EFA) with all items in Table 1. EFA is a statistical technique that reduces data to a smaller set of factors and explores the underlying theoretical structure of the phenomena without prior specifications of the number of factors and their loadings (Venkatraman, 1989; Statistics Solutions, 2020). We employed a maximum likelihood method with the ProMax rotation in SPSS version 25. We extracted six factors with Eigenvalue criteria, as assumed. Table 4 shows the pattern matrix.

Table 4: Result of Exploratory Factor Analysis

	Employee	Customer	SSCM	Government	Top Management

SSCM1	0.09	-0.046	0.894	0.031	-0.125
SSCM2	-0.09	0.179	0.546	-0.033	0.103
SSCM3	-0.108	0.063	0.634	-0.025	0.204
SSCM4	-0.072	-0.079	0.752	0.035	0.069
SSCM5	0.196	0.029	0.688	-0.038	-0.105
Top Management1	0.104	-0.116	-0.02	0.044	0.887
Top Management2	0.029	0.139	-0.043	-0.064	0.814
Top Management3	0.027	0.02	0.082	-0.066	0.757
Top Management4	-0.012	-0.037	0.062	0.189	0.673
Customers1	-0.06	0.691	0.077	-0.018	0.12
Customers2	-0.053	0.91	0.031	-0.011	-0.02
Customers3	0.06	0.92	-0.014	0.059	-0.043
Customers4	0.077	0.92	-0.053	0.023	-0.037
Employees1	0.82	0.008	-0.047	-0.081	0.138
Employees2	0.841	0.084	0.007	-0.034	0.008
Employees3	0.925	-0.057	0.022	0.041	0.036
Employees4	0.898	-0.006	0.02	0.062	-0.055
Government1	0.031	0.026	-0.002	0.775	-0.032
Government2	0.032	0.038	0.128	0.503	0.047
Government3	-0.022	0	-0.025	0.85	0.029
Government4	-0.031	0.006	-0.039	0.895	-0.003

Next, we conducted a Confirmatory Factor Analysis (CFA) with Amos 25 (Table 5). CFA "simultaneously tests the unidimensionality of a set of correlated latent constructs, which helps to establish both unidimensionality and convergent validity" (Wu et al., 2010). We computed Cronbach alpha, the average variance extracted (AVE), composite reliability (CR), and the square root of AVE at the bottom of the table to assess convergent and discriminant validity. All AVEs exceeded 0.5 (Chin, 1988; Chung et al., 2015). All CR and Cronbach alpha values exceeded 0.7 (Gefen, Straub, & Boudreau, 2000; Nunnally, 1978). They endorse the measurement reliability and internal consistency of our constructs (Cronbach, 1951; Forza & Salvador, 2000).

Table 5: Confirmatory Factor Analysis (CFA)

	SSCM	Top Management	Customer	Employee	Government
SSCM1	0.747				
SSCM2	0.722				
SSCM3	0.757				
SSCM4	0.673				
SSCM5	0.844				
Top Management1		0.872			
Top Management2		0.84			
Top Management3		0.822			

Top Management4		0.801			
Customers1			0.761		
Customers2			0.881		
Customers3			0.959		
Customers4			0.928		
Employees1				0.847	
Employees2				0.884	
Employees3				0.939	
Employees4				0.894	
Government1					0.803
Government2					0.641
Government3					0.868
Government4					0.853
Cronbach Alpha	0.857	0.904	0.934	0.939	0.863
AVE	0.692	0.807	0.869	0.879	0.752
CR	0.918	0.944	0.964	0.967	0.923
Sqrt (AVE)	0.832	0.898	0.932	0.938	0.867

Table 6 shows correlation coefficients between latent variables and the square root of AVE. The square root of AVE (the value of diagonal elements) should be greater than the inter-correlations of constructs (off-diagonal components) (Chung et al., 2015; Dubey et al., 2019). Table 6 satisfy it in all columns. Therefore, discriminant validity is proved (Fornell & Larcker, 1981), and we proceeded to the next step, simultaneous equations modeling (SEM).

Table 6: Correlation between Latent Variables and Square Root of AVE

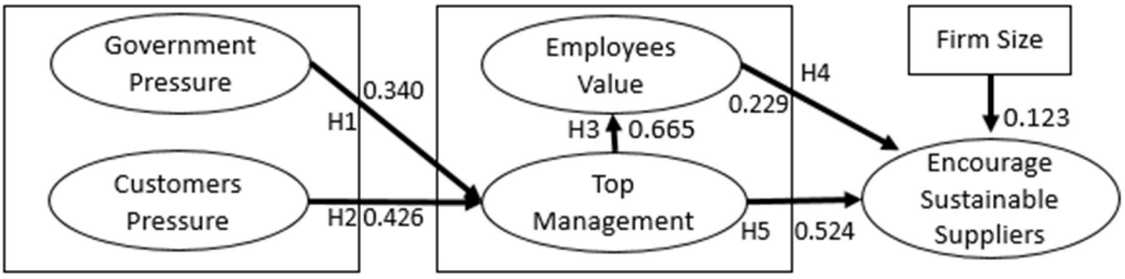
	SSCM	Customer	Employee	Government	Top Management
SSCM	0.832				
Customer	0.637	0.898			
Employee	0.579	0.548	0.932		
Government	0.431	0.546	0.398	0.938	
Top Management	0.666	0.570	0.649	0.564	0.867

Simultaneous Equations Modeling Result

We apply SEM to data with AMOS version 25. Figure 2 shows the result. All standardized coefficients are significant. The effect of firm size is not strong but still significant (p=0.029). The positive coefficient means that the larger the plant, the more the plant solicits suppliers to SSCM. All other coefficients, except H2 (from customers to top management, p = 0.002) are strongly significant (p < 0.001). Model-fit is good enough. Chi-square = 396.059 (D.F. = 201, p = 0.000), CMIN/DF = 1.970, NFI = .918, IFI = .958, TLI = .946, CFI = .957, FMIN = 1.299, and RMSEA = .056. Chi-square rejects the null hypothesis because of the over 300 sample size

(Bagozzi et al., 1991, p.436). "With a large N the statistical power of the Chi-square test may reject good models" (Dion, 2008). "The Chi-Square statistic nearly always rejects the model when large samples are used" (Hooper et al., 2008).

Figure 2: SEM result



DISCUSSION AND CONCLUSIONS

The Conclusion from SEM Result

SSCM is a critical subject to the manufacturing industry and a vital decision theme for each factory (plant). Stakeholders, such as top management and employee, encourage suppliers to sustainability or SSCM.
Companies must obey government regulations. Government regulation pushes top management (H1). Unless a plant follows government regulation, the government usually forces a hefty penalty to the plant. For example, United States Environmental Protection Agency (EPA) operates a series of environmental protection acts such as the Clean Air Act, Clean Water Act, and Toxic Substances Control Act and regulations (Environmental Protection Agency, 2020).
Customers also lift their voices to improve sustainability to top management (H2).
Our result (Figure 2) supported that top management is the prime decision-maker of SSCM (H5), and it is a natural result because of the responsibility of top management. Also, top management orders to support the establishment of sustainability to employees (H3).
The importance of employee value is also endorsed. Our results endorsed that motivated employees play a significant role in supporting suppliers to be sustainable (H4).
Our result reveals that the impact of customers and the government to encourage sustainable suppliers is indirect through top management, but both are strong enough. Figure 2 shows that the effect of customers is a little more substantial than the government to top management.

Academic Contribution

This study is an influencer study of SSCM within factories. Sociologically, we illustrate a typical influencer pattern of environmental protection movement among stakeholders in factories quantitatively. We reveal that both top management and employees are strong influencers to encourage upper stream suppliers to SSCM in a plant. We also demonstrate that firm size is an essential factor for SSCM decision. Generally speaking, the larger plant is more active and resourceful for SSCM than smaller plants. We believe that there are some explanations such as financial power, low fixed cost per product because of the size effect, and human resource and knowledge level of engineers in a plant.
Our research is an empirical study about influencers to SSCM. Melnyk et al. (2014) wrote: "while much supply chain research has delved into the relevance of design decisions and

building blocks, there is a paucity of robust empirical and analytical research that explicates the role of influencers in the supply chain design process." Our research fills the gap.

Practical Contribution

This study is descriptive (empirical) rather than normative. First, our result suggests that the leadership of top management is critical. Second, employees affect the encouragement of sustainable suppliers strongly. If you want to encourage suppliers to SSCM, you can persuade related people. Motivated employees play a significant role in supporting suppliers to be sustainable. Today, leadership and empowerment of employees to sustainability is becoming more critical than ever. More leadership development and human resource management are requested (University of Cambridge Institute for Sustainability Leadership (CISL), 2018). Third, government agencies and customers can impact firms through the top management of the firm indirectly. Government regulation has a strong impact. Although these are already known wisdom for intelligent CEO and government agencies, this study result underscores the effectiveness.

Research Limitation

Some research limitations go to future studies. First, although enough sample size for CFA and SEM analysis, we need to repeat the same research with different and hopefully with more data and other methods to confirm the results. A single method error may affect our results. Multi Trait and Multi-Method (MTMM) may be requested (Bagozzi et al., 1991). Second, because the data are not new, we need to repeat the study with the latest dataset. Because of IT innovation and expansion of environmental pollution, the importance of the SSCM approach, subject and technology has changed in a short period. We need to repeat this study to confirm the ubiquity of the conclusion.

ACKNOWLEDGMENT

This study was supported by Grants-in-Aid for Scientific Research (B), Grant Number 19H01520, sponsored by the Japan Society for the Promotion of Science (JSPS).
The authors would like to gratefully acknowledge the comments provided by two anonymous reviewers who allowed us to revise and helpful suggestions.

REFERENCES

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421-458.

Boh, W. F., & Won, S. S. (2013). Organizational Climate and Perceived Manager Effectiveness: Influencing Perceived Usefulness of Knowledge Sharing Mechanisms. *Journal of the Association for Information Systems*, 14(3), 122-152.

Botes, A., Niemann, W., & Kotzé, T. (2017). Buyer-supplier collaboration and supply chain resilience: A case study in the petrochemical industry. *South African Journal of Industrial Engineering*, 28(4), 183-199.

Cambridge University Press. (Ed.) (2011) *Cambridge Business English Dictionary*. Cambridge, UK: Cambridge University Press.

Cambridge University Press. (Ed.) (2005) *Cambridge Advanced Learner's Dictionary*. Cambridge, UK: Cambridge University Press.

Campos, J. K., Straube, F., & Wutke, S. (2017). Creating value by sustainable manufacturing and supply chain management practices – a cross-country comparison. Paper presented at the *14th Global Conference on Sustainable Manufacturing*, Stellenbosch, South Africa.

Carter, C. R., & Easton, P. L. (2011). Sustainable supply chain management: evolution and future directions. *International Journal of Physical Distribution & Logistics Management*, 41(1), 46-62.

Carter, C. R., & Rogers, D. S. (2008). A framework of sustainable supply chain management: Moving toward new theory. *International Journal of Physical Distribution and Logistics Management*, 38(5), 360-387.

Christopher, M., & Peck, H. (2004). Building the Resilient Supply Chain. *International Journal of Logistics Management*, 15(2), 1-13.

Chung, K., Oh, J., Kim, W., & Park, G. (2015). The Effects of Perceived Value of Mobile Phones on User Satisfaction, Brand Trust, and Loyalty. *Advanced Science and Technology Letters*, 114, 10-14.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests *Psychometrika* (Vol. 16, pp. 297-334).
David, R. J., Tolbert, P. S., & Boghossian, J. (Eds.). (2019) *Oxford Research Encyclopedia of Business and Management*. Cambridge, UK: Oxford University Press.

Dion, P. (2008). Interpreting Structural Equation Modeling Results: A Reply to Martin and Cullen. *Journal of Business Ethics*, 83(3), 365-368. doi:10.1007/s10551-007-9634-7

Dubey, R., Gunasekaran, A., Childe, S., Papadopoulos, T., & Helo, P. (2019). Supplier relationship management for circular economy. *Management Decision*, 57(4), 767-790. doi:https://doi.org/10.1108/MD-04-2018-0396

Environmental Protection Agency. (2020). *Laws & Regulations*. Retrieved from https://www.epa.gov/laws-regulations

EU. (2013). Commission Recommendation of 9 April 2013 on the use of common methods to measure and communicate the life cycle environmental performance of products and organisations. *Official Journal of the European Union*, 56.

Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of Operations Management,* 28(1), 58-71. doi:10.1016/j.jom.2009.06.001

Fornell, C. R., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error, *Journal of Marketing Research* (Vol. 18, pp. 39-50).

Forza, C., & Salvador, F. (2000). Assessing some distinctive dimensions of performance feedback information in high performing plants. *International Journal of Operations & Production Management*, 20(3), 359-385.

Galbin, A. (2014). An Introduction to Social Constructionism. *Social Research Reports*, 26, 82-92.

Gualandris, J., & Kalchschmidt, M. G. M. (2014). Customer pressure and innovativeness: Their role in sustainable supply chain management. *Journal of Purchasing and Supply Management*, 20(2), 92-103.

Gupta, S., & Palsule-Desai, O. D. (2011). Sustainable supply chain management: Review and research opportunities. *IIMB Management Review*, 23, 234-245. doi:doi:10.1016/j.iimb.2011.09.002

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-59.

ISO. (2006). *ISO 14040:2006 Environmental management — Life cycle assessment — Principles and framework* (Vol. 14040): ISO.

Kaynak, H., & Montiel, I. (2009, 7-11 August). The rtelationship between sustainable supply chain management and sustainable performance: An integrated framework. *Paper presented at the Academy of Management*, Chicago, IL.
Koster, M., Vos, B., & Schroeder, R. (2017). Management innovation driving sustainable supply management Process studies in exemplar MNEs. *Business Research Quarterly*, 20, 240-257.

La Londe, B. J. (1997). Supply Chain Management: Myth or Reality? *Supply Chain Management Review*, 1, 6-7.

Melnyk, S. A., Narasimhan, R., & DeCampos, H. A. (2014). Supply chain design: issues, challenges, frameworks and solutions. *International Journal of Production Research*, 52(7), 1887-1896.
Mentzer, J. T., DeWitt, W., Keebler, J. S., Soonhoong, M., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining Supply Chain Management. *Journal of Business Logistics*, 22(2), 1-25.

Paulraj, A., Chen, I. J., & Lado, A. A. (2012). An Empirical Taxonomy of Supply Chain Management Practices. *Journal of Business Logistics*, 33(3), 227-244. doi:10.1111/j.0000-0000.2012.01046.x

Phan, A. P. (2014). High Performance Manufacturing (HPM) Project Summary. Retrieved from http://dl.ueb.edu.vn/bitstream/1247/7579/2/Mr%20Chi%20Anh_En.pdf

Schroeder, R. G., & Flynn, B. B. Eds. (2001). *High Performance Manufacturing: Global Perspectives* : Wiley.

Statistics Solutions. (2020). Exploratory Factor Analysis. Retrieved from https://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/

Tebaldi, L., Bigliardi, B., & Bottani, E. (2018). Sustainable supply chain and innovation: A review of the recent literature. *Sustainability* (Switzerland), 10(11). doi:10.3390/su10113946

The Committee on Approaches to Climate Change Adaptation. (2010). *Approaches to Climate Change Adaptation Tokyo, Japan*: Ministry of Environment, Japan Retrieved from https://www.env.go.jp/en/earth/cc/adapt_guide/pdf/approaches_to_adaptation_en.pdf.

University of Cambridge Institute for Sustainability Leadership (CISL). (2018). *Rewiring the Economy: The future we want, the leadership we need*. Retrieved from Cambridge, UK: www.cisl.cam.ac.uk/publications

Venkatraman, N. (1989). Strategic orientation of business enterprises: The construct, dimensionality, and measurement. *Management Science*, 35(8), 942-962. doi:10.1287/mnsc.35.8.942

Villena, V. H., & Gioia, D. A. (2020). A more sustainable supply chain. *Harvard Business Review*, 98(2), 84-93.

Wu, S. J., Melnyk, S. A., & Flynn, B. B. (2010). Operational Capabilities: The Secret Ingredient. *Decision Sciences*, 41(4), 721-754.

Xu, K., & Cong, H. (2011). A Framework of Sustainable Supply Chain Management in Beijing Environmental Logistics. *Paper presented at the International Conference on Computational Sciences and Optimization*, Singapore.

Young, K. (2018). What Drives Eco-Conscious Consumer Behavior? Retrieved from https://blog.globalwebindex.com/chart-of-the-day/eco-conscious-consumer/#:~:text=Eco%2DConscious%20Consumers%20are%20defined,%2Dto%2Dday%20buying%20decisions.

DECISION SCIENCES INSTITUTE

Are corporate social responsibility initiatives worth having?
Empirical evidences from SA 8000 decertification

Matteo Podrecca
University of Udine
Email: podrecca.matteo@spes.uniud.it

Guido Orzes
Free University of Bozen
Email: guido.orzes@unibz.it

Marco Sartor
University of Udine
Email: marco.sartor@uniud.it

Guido Nassimbeni
University of Udine
Email: guido.nassimbeni@uniud.it

ABSTRACT

The purpose of this paper is to shed light on the topic of decertification from one of the most important corporate social responsibility initiatives, the SA 8000 ethical standard. Starting from a literature analysis, we develop six research hypotheses, three of them relating decertification to firm performance and three which shed light on the differences between certified and decertified firms. The results demonstrate that decertified firms exhibit performance benefits from decertification and that certified and decertified firms differ in terms of home country, industry and SA 8000 adoption timing.

KEYWORDS: Decertification, CSR, SA 8000, Event-study

DECISION SCIENCES INSTITUTE

Electric vehicles in India – To buy or not? Mining opinions of consumers

Shameem Shagirbasha
Great Lakes institute of Management
Email: shawaja@gmail.com

ABSTRACT

Electric vehicles (EV) are one of the emerging solutions to curb the pollution problem and move towards green development. This paper analyses the opinion, reviews and sentiments of consumers towards adopting electric vehicles in Indian market using textual analysis of tweets and posts on social media. Using sentiment analysis of around 25000 tweets and pattern identification, the factors associated with consumers' willingness to buy were categorized. The positive factors influencing consumers' acceptance towards electric vehicles are affordability, economic benefits from long term fuel savings, high energy efficiency, pollution reduction, tax redemption by government. The negative factors emerged are cost, charging stations for electric vehicles, production of surplus electricity, speed and distance constraints. The study also highlights the implications of the findings to the various stakeholders involved in the process.

KEYWORDS: electric vehicles, sentiment analysis, twitter, willingness to buy, India

INTRODUCTION

Currently India's transportation system is majorly dependent on fossil fuels such as petrol and diesel. More and more usage led to the depletion of these resources as well as increase in the pollution in the ecosystem. Electric vehicles (EV) can be considered as one of the emerging solutions to curb the problem of pollution and move towards green development. Efforts are taken by the Indian government to encourage the use of electric vehicles, however, the success depends on consumer's adoption of electric vehicles (Jena, 2020).This paper aims to summarise findings about the opinion, reviews and sentiments of consumers towards adopting electric vehicles in Indian market using textual analysis of tweets and posts on social media. Electric Vehicles have batteries that can offer a solution of the storage to India's clean energy push. Efficiency of electrical mobility is more than mobility by petrol or diesel. EVs are an area of huge interest to India since it holds the potential of reducing the demand for liquid fuel to a greater extent.

India's National Democratic Alliance government is promoting the EVs as the achievable strategic goal, which is keen to cut the country's oil imports. Adoption of EVs in India will lead to cleaner air and the noise level as well as the oil import bills will be down. If the EVs penetration goes up to 30% in India by 2030 it will generate an estimated saving of 474 Millions of tonnes of oil equivalent and 846 million tonnes of net CO2 emissions over their lifetime. Through sentimental analysis, the insights are obtained if the people are ready to buy electrical vehicles in India, which helps the automobile companies to have a head start in the E- vehicle segment and gives an early mover advantage in the industry. It will also help the government to get a fair idea of the sentiments of Indian people towards E-vehicle so that they can plan the subsidies and tax benefits for more penetration of EVs in Indian market.
India, the world's third-largest energy consumer after US and Republic of China, is working towards building a green and sustainable economy and plans to achieve 175 gigawatt (GW) of renewable energy capacity by the year 2022 as part of the mission and commitments under the global climate change accord. Of this, 100GW is from solar. Though EV industry occupies a small percentage of overall market, it is positively seen as an uplift for it to reach and impact the international market strongly in the coming years. With such shift to renewable energy, it

makes good sense for India which paid Rs4.16 trillion to buy 202.85 million tonnes of crude oil .Electric vehicles' impact on the environment has been considered a lot of times without any serious steps taken in that direction. The economic model is original because it examines rational consumers' choice between switching to electric vehicles (EVs) and purchasing all other goods (Verma, Verma and Khan, 2020). The paper also explored the environmental impact of adapting to electric vehicle, which will supplement our argument for whether or not pushing the consumers for electric vehicle is a good choice or not.

LITERATURE REVIEW

"All electric vehicles on road by 2030" - government's ambitious plan. Opportunities are there with these transformative changes (Utpal Bhaskar, 2017; Motwani and Patil, 2019). Electric mobility can be seen as the way to improve air quality and meet environment necessities, but it is highly integrated for smarter cities. Electric vehicles are associated with traditional purchase and use models, and are generally considered as just cars still: innovative uses and services related to batteries, or integration with smart buildings, either are ignored or not given proper thought. For positive outcome, impeccable and profound planning and patience in execution strategy are the needs of electric vehicle market while introducing. Governments should give emphasis on creating an atmosphere positive toward EV acceptance in India. International collaboration lowers the cost for government to introduce EVs and demonstrates to the world its environmental concern and commitment towards this new way of transportation. The ultimate beneficiaries from EVs are people and environment.
Electric vehicles are producing 97% less pollution compared to gasoline powered vehicles It does not produce any tailpipe emissions that place particulate matter into the air. The particulate matter, (carcinogens) released into the atmosphere by gasoline powered vehicles, aggravate health condition of asthma patients, and also irritate respiratory systems. In a world where environmental protection and energy conservation are two most growing concerns, the development of electric vehicle is taking place at higher rate. Concerned with the environment, EVs can provide emission-free urban transportation. From the energy aspect, electric vehicle can offer as a secured, comprehensive and balanced energy option that is more efficient and environment friendly. Furthermore, EVs, potentially have great impact on energy, environment and transportation as well as promotion, new industry creation and also on development of economy.

Currently there are two types of EVs available in the market:
1. All-electric vehicles (AEVs)
2. Plug-in hybrid electric vehicles (PHEVs).
In future electric cars will most likely carry lithium-ion phosphate (LiFePO4) batteries that are now becoming popular in other geographies. The LiFePO4 batteries are rechargeable and powerful and can be used in electric bikes and scooters. Electric vehicle transportation is well suited for developing nations like India which already has a potential EV market. The purpose of this research is to analyse consumers' sentiments for electric vehicle future in India and its acceptance among them. There is absolutely no doubt that auto industry can build Electric Vehicles, but the key question is, whether consumers buy them. Electric vehicle costs similar to the same model of a conventional vehicle. If an electric vehicle price is high, what value will the consumer obtain for the extra cost at which electric vehicles penetrate the market? Electric vehicle sales should be evenly distributed across the country, or located in certain urban areas in different proportions. Consumers want to purchase new vehicles which provide better attributes as compared to their existing vehicles. For electric vehicles to be competitive, they must be able to perform at the same level, should have the same attributes as the conventional vehicles that they are replacing. China, at present is the largest market for electric vehicles, about 650,000 electric vehicles on road, representing about a third of the world's total. India is joining China and is setting aggressive EV targets. India is thinking to have even more radical action, with plans to electrify all present vehicles by 2030. Which means India needs

to sell more than 10 million Electric Vehicles by 2030 (Kumar and Padmanaban, 2019). Experts predicted that demand for Electric Vehicles are going to go even higher in future. Electric vehicle may constitute to almost one third of new vehicle sales. Electric mobility is widely seen as a way to improve air quality and meet climate goals today.

Effect of price on consumer willingness to buy Electric vehicles.
One of the drawbacks of Electric vehicles is their higher cost than ICE vehicles. Technological developments as well as economies of scale help with solving these roadblocks. Battery costs have been cut by a factor four since 2008 and the battery range of EVs has increased 13 tremendously. Tesla seems to get closer and closer to the very competitive $35,000 price-tag it is targeting for its Model 3. Tesla's gigafactory strongly contributes to this, as it enables them to reach even more economies of scale when it comes to producing the costliest part of the car: the battery. Price is an important factor when choosing an EV. The research has shown that (Hidrue, Parsons, Kempton and Gardner, 2011) purchase price is one of the main concerns leading to purchase of EVs.

H1: There is a negative relationship between price and consumer willingness to buy EV.

Effect of mobility on consumer willingness to buy Electric vehicles.
India is moving towards increased mobility. Government has launched programs and aims to expand shared mobility, electric mobility and EVs in India. Like many developing nations, India is shifting towards EVs and increased mobility to reduce air pollution, carbon emissions and fuel imports. Lane and Potter (2007) found that the oil price policy, purchase subsidies and the charging infrastructure will affect the market penetration of cleaner vehicles.

H2: There is positive relationship between mobility and Consumer willingness to buy EV.

Effect of recharging on consumer willingness to buy Electric vehicles.
Charging stations are very much essential to an EV. It is necessary to build charging infrastructure to remove the consumers "range anxiety". It is found that the inadequate charging infrastructure is one of the inhibitors. The charging infrastructure readiness is rather important to influence the consumers' willingness to purchase EV (Browne, O'Mahony and Caulfield, 2012). There are several barriers for public charging stations such as regulatory uncertainty, a lack of standardization and most of all high cost (Dagsvik, Wennemo, Wetterwald and Aaberge, 2002). This is where there is an important role for the government. There will be approximate increase of 25% of electric vehicles as estimated by the government by 2030 by providing adequate charging infrastructure across the country. As per the guidelines, there must be at least one station at every 25 Km on each side of the road.

H3: There is a positive relationship between Recharging and Consumer willingness to buy EV.

Effect of government subsidy on consumer willingness to buy Electric vehicles.
The government is the main driving force in the early stage of the EV industry development. The policies include monetary incentives and non-monetary incentives. The monetary incentives consist of the purchase incentives, charging infrastructure incentives, purchase tax exemptions and electricity cost subsidies. The non-monetary incentives consist of exemptions of road tolls and free public charging. Gallagher and Muehlegger (2011) found that the tax subsidies are more effective to encourage the consumers' purchase than other supporting incentives. The government asserted the subsidies for promotion of electric mobility in India.

H4: There is a positive relationship between government subsidies and consumer willingness to buy EV

Effect of environmental concerns on consumer willingness to buy Electric vehicles.

Studies have shown that environmental concerns have a positive impact on the intention of buying Electric Vehicles. If the high percentage of electricity mix will come from renewable source then this may reduce the pollution to a great extent. The government of India is committed to solve New Delhi pollution problem by hundred percent of light duty consumer vehicles to electric vehicles by 2030. But human behaviour changes are must to achieve this result. (Dhar, Pathak and Shukla, 2017). The government should adopt policies to increase the percentage of renewable source of energy in the electricity mix. Customers who are environmentally sensitive and identified themselves as an environmentally-friendly person would be more likely to adopt EVs.

H5: There is a positive relationship between environmental concerns and consumer willingness to buy EV

Effect of fossil fuel extinction on consumer willingness to buy Electric vehicles.
We know that currently the world is heading towards the extinction of fossil fuel reserves. Currently transport is one of the vital aspects of many people's lives. This means that world is quite dependent on oil. We have seen in the past there are several oil crises happened. Much of it is influenced by politics. There is always a risk of sudden increase in oil prices. By using EV, we can eliminate that risk to a great extent. For this oil exporting countries could also get the benefit if all reduce production simultaneously. This will reduce the risk of extinction of fossil fuels. It can be used by the consumers where it is necessity (Blok and Nieuwlaar, 2020). Altogether, it will have the positive impact on the consumers.

H6: There is a positive relationship between fuel extinction and Consumer willingness to buy EV.

Effect of consumer belief on consumer willingness to buy Electric vehicles.
Belief in EV technology is an important criterion for consumers to accept it. According to study, they referred to EV as future of transportation. Consumers believe it as the vehicle for the coming years and it can be extensively used for travel and transport purposes. These were found to be positive associations (Egbue and Long, 2012).

H7: There is a positive relationship between consumer belief and consumer willingness to buy EV.

Effect of range anxiety on consumer willingness to buy Electric vehicles.
It means that consumers may have the fear that the battery of their vehicle will be depleted at the middle of their trip which could make them helpless. When drivers are unsure about the trip length or energy consumption for heating or cooling the car, it is hard for them to guess how long their battery will last. Increasing the battery size or recharging while on the road using charging stations are solutions of this. From the literature it is known that limited range is a barrier for EV adoption. Research confirms that range anxiety is a persisting problem, as perception of range remains a great barrier 15 for acceptance even with experienced EV drivers (Bühler, Cocron, Neumann, Franke and Krems, 2014). The study by Hidrue et al. (2011) named range anxiety to be one of the main concerns of consumers.

H8: There is a negative relationship between range anxiety and Consumer willingness to buy EV.

Effect of social influence on consumer willingness to buy Electric vehicles.
Individual's behaviour tends to influence by others which ultimately affects the choice. Our buying decisions are affected by social externalities such as social norms, peer pressure. A study focusing at purchasing behaviour of consumer electric vehicles found that there is a significant impact of neighbour on willingness to buy electric vehicles (Mau, Eyzaguirre,

Jaccard, Collins-Dodd and Tiedemann, 2008). For example, if there are many EV drivers in the neighbourhood, then it is most likely that it will influences us to buy EV.

H9: There is a positive relationship between social influence and consumer willingness to buy EV.

METHODOLOGY

This study involves sentiment analysis on twitter for data collection and analysis. It is a generative probabilistic model for collections of discrete data such as text corpora (Blei, Ng, and Jordan, 2003). When consumers make a decision or a choice regarding the selection of a product, the reputation of the product plays an important, which is derived from the opinion of others. Sentiment analysis can reveal what other people think about any product. When we choose a product, we are generally attracted towards the certain aspects of it. Sentiment analysis can regroup the opinions of the reviewers and estimate ratings on various aspects of the product. Sentiment Analysis can be defined as a systematic analysis of online expressions. The field is evolving with various techniques and approaches to examine the emotional behaviour of users (Huh, Yetisgen-Yildiz, and Pratt, 2013; Tuarob, Tucker, Salathe, and Ram, 2014). Twitter has become an integral part of our life to share the experiences of life. People tweet on twitter show their sentiments. These tweets are of concise length of about 280 words. (Pak and Paroubek, 2010; Singh, Dwivedi, Rana, Kumar and Kapoor, 2019). This study collects the tweets from twitter for the period of 2 months. The polarity is determined, meaning categorised sentiments into negative, positive and neutral, which presents the relationship between the ratings of different aspects of the products (Cao, Thompson and Yu, 2013). The tweets are categorized as positive or negative polarities with the extent of various emotion classifications. Emotions can be captured by counting the words based on positive and negative sentiments (Fredrickson and Losada, 2005). Counts based on only positive/negative sentiments capture users' behaviour in the context of new product development (Jiang, King and Prinyawiwatkul, 2014).

In the forthcoming sections, the author presents how various data mining methods were executed to get the insights from Twitter and explicit techniques used to visualize data. There are three phases: data pre-preparing, data analysis, and data visualization/understanding. This methodology is implemented through R application utilizing various contents.

Data Collection

The objective of this paper is to examine the patterns of consumer reaction towards electric vehicles in India. The author analysed a sample of around 25-30 thousand tweets in addition to other data. R software was used to extract the tweets. An API (Application Programming Interface) allows users to access (real-time) twitter data. It establishes a connection to the streaming APIs meaning to make a very long lived HTTP request, and parsing the response incrementally. Then we authenticate when making requests to the Twitter API which includes 4 steps:

1. creating a Twitter account
2. logging in to your Twitter account via https://developer.twitter.com/
3. Creating an app.
4. Creating keys, access token and secret.

Authorized tweets are public tweets as well as those tweets which are protected and are available to the user after authenticating using registerTwitterOAuth. Then the author used searchTwitter function to extract the tweets using search String and then filter the data using geocode for specific locations and set the limit for data using max no of tweets.

Data Pre-Processing

Data pre-processing includes different steps, including keyword selection, tweet extraction, and tweet preparation. Twitter, a micro blogging platform, is used as a data source. To perform the search, an R script was used to download the tweets over weekly time periods. The posts/comments were collected for the appropriate time period; for example we collected data over a period of 3 months in weekly manner for EV. As a first step, it is necessary to determine the most appropriate keywords related to the EV. Effective keyword selection (e.g., the popular key terms related to a specific EV's) helped in achieving the most fitting tweets to create the analytic corpus for the specific event. The hash tags used are described in the below table:

Insert Table 1

Twitter data contain a variety of languages, unstructured abbreviation and grammar, and unofficial format. Tweets are filtered to be in English language. The tweets are extracted based on the search query using "@" annotation followed by the Electric vehicle (cars, bikes etc.). After that, the author cleaned the datasets from any unnecessary characters such as re-tweets and usernames' symbols, hash tags, numbers, punctuations, stop words, whitespaces and html links. Datasets are performed through an automated text mining R script and the use of syntactic/semantic libraries (Natural Language Packages) through the following techniques: Tokenization, Filtering, stemming, and Lemmatization as below:
Tokenization: Tokenizing means splitting your text into minimal meaningful units. It is a mandatory step before any kind of processing of the data. Reads the text that will be mined and removes all tabs and punctuations between words and replaces them with a white space.
Filtering: It will remove words such as: stop words, repeated words and rarely repeated words. For example a, and, the, but, is, are, they, them etc.
Lemmatization: converting words to dictionary form which will be used to transform all the verbs to the infinite tense and all the nouns to the singular form.
Stemming: Reducing words to their root form.Some more example of stemming for root word "like" include: "likes","liked","likely", "liking" etc.

FINDINGS

To understand such behavioural intentions, it is important to observe significant differences between positive and negative sentiments

H1: There is negative relationship between price and consumer willingness to buy EV.

```
> table(price)
price
Negative  Neutral Positive
      62       48      155
> |
```

Hypothesis: Does not hold TRUE, since the positive sentiments are higher than negative. Even if the prices go up, the customers are willing to pay for the EVs

H2: There is positive relationship between mobility and consumer willingness to buy EV.

```
> table(mobility)
mobility
Negative  Neutral Positive
      26       98      278
> |
```

Hypothesis: holds TRUE, since the positive sentiments is higher than negative. Consumers are willing to adopt and pay for higher mobility of EVs.

H3: There is a positive relationship between recharging and Consumer willingness to buy EV.

```
> table(charging_Station)
charging_Station
Negative  Neutral Positive
      44      150      206
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. Having good charging stations or increasing its number influence the number of people who are willing to buy EVs.

H4: There is a positive relationship between government subsidies and Consumer willingness to buy EV

```
> table(subsidy)
subsidy
Negative  Neutral Positive
      28       21       59
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. It seems as more subsidies or incentives or tax exemptions are done by the Government then more customers can be inclined towards buying EVs.

H5: There is a positive relationship between environmental concerns and Consumer willingness to buy EV

```
> table(Environment_concern)
Environment_concern
Negative  Neutral Positive
      58       65      170
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. Consumers who are more environmentally concerned are more inclined towards buying EVs and reduce the carbon footprints.

H6: There is a positive relationship between fuel extinction and Consumer willingness to buy EV.

```
> table(energy_resource)
energy_resource
Negative  Neutral Positive
      93      177      453
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. With increasing concern of non-renewable fossil fuels, the EVs can be an alternative option at hand which is preferred in long term and customers are willing to buy EVs for this reason as well.

H7: There is a positive relationship between consumer belief and Consumer willingness to buy EV.

```
> table(Belief)
Belief
Negative  Neutral Positive
       8       17       62
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. The belief has a positive relationship with the willingness to buy the EVs. Having faith in the EVs by the customers is making them to shell out money to adopt it and get rid of IC's vehicles.

H8: There is a negative relationship between range anxiety and Consumer willingness to buy EV.

```
> table(Range_anxiety)
Range_anxiety
Negative  Neutral Positive
      31        1        6
> |
```

Hypothesis: Holds TRUE, since the negative sentiments is higher than positive. It may be people's fear about the battery life will reduce their purchase intention to buy EV.

H9: There is a positive relationship between social influence and consumer willingness to buy EV.

```
> table(social_influence)
social_influence
Negative  Neutral Positive
      11        7       25
> |
```

Hypothesis: Holds TRUE, since the positive sentiments is higher than negative. It can infer that social interactions with neighbours, friends and family members have positive influence on buying EV.

DISCUSSION

Sentimental Analysis Model

The author used NLP (Natural Language processing) and advanced ML algorithms to extract the polarity from the tweets as positive, negative and neutral. That led to find out different kinds of emotions related to the sentiments of the potential customers around the globe. The defined emotions are joy, trust, anticipation, surprise, fear, anger, sadness, disgust. On the basis of the emotions selected we formulated a word cloud which shows the intensity of various keywords chosen for the tweets which helps in the analysis of the sentiments in broader terms.

Emotion classification highlights the high counts for joy, anticipation, and trust, showing users' bending towards positive emotions. In emotion classification, trust has been increased with a significant decrease in anger and sadness. Joy and surprise are also considerably decent for EVs usage. Users are feeling good about EVs performance and appearance. The design part of EV is more impressive for them. With joy, users are also surprised with marketing and advertising side during product promotion. Categorizing words in emotions helps to determine product attributes reflecting users' preferences. Inferences for product aspects are not essentially made deliberately about mental judgements. Word categorization shows that joy emotions are associated with "good," "great," "happiness," "important," "love," "suspension," "support," "marketing," "kit," etc. It shows users' positive attitudes towards technical and managerial aspects. Users are sad about the lack of luxurious experiences and heavy with fray looking along with bad dealer. Moreover, users are feeling disgusted because of the rude dealership .Users got surprised with amazing and fantastic "EVs hybrid", which is better than other conventional vehicles (Chan, 2007). With all emotions, users are expecting new brake concepts and space in anticipation emotions.

Insert Table 2

Insert Figure 1

The analysis was performed using approximately 45000 tweets. After cleaning the data, around 18000 tweets were obtained. The result showed the majority of positive sentiments with the count of 15800 in support to launch of EVs. After that about 6000 emotions is observed in the trust category which means people are slowly putting faith in the new technology. In the anticipation category there were more than 5900 sentiments. It shows that people still have more expectations and more queries about the electric vehicles. Nevertheless, the study also gathered quite a few negative sentiments by the consumers. The counts are about half of the positive sentiments which means that there are some people who have difficulty in accepting this technology reason being the high price or range anxiety. Sentiments regarding anger and disgust are quite less but still present. It can be reduced if the manufacturing companies can infuse supercharger batteries in vehicles and simultaneously try to reduce price.

Insert figure 2

The 18000 tweets are further divided into 3 polarities: positive, negative and neutral.
We got 10800 positive, 2700 negative and neutral 4500. This shows that majority of consumers are showing their willingness to buy EV.

Insert figure 3

From the word cloud it can be inferred that certain words hold a major weightage in the categories of emotions.

Positive Sentiments: - In the sentiment analysis majority of people are inclined towards positive sentiments due to the technology and growth opportunities that they see. People believe that EVs are the future technology. It gives them the sense of advancement and contribution towards clean environment.

Negative Sentiments: - The prominent words in this category are noise and sales. As we know that electric vehicles produce quite less noise that other vehicles. It possesses danger to the on-road travellers especially those who are blind or partially blind.

Anger: - In the anger category the important words are battery and money. Due to the inadequate charging stations people are facing issues like range anxiety. Moreover, in India affordability is the main concern. It is believed that for the next two or three years there will be no real buying.

Fear: - Change and Government are the two words that highlight in this category. This is probably because there are still many people who are opposed to change. They are still uncertain about the EV tech. Moreover, manufacturers have growing concerns whether government will provide subsidy or not for the large projects. As this is still the growing technology, manufacturers are not sure whether consumers will buy in future or not.

Joy: - In this category the word with the highest weightage is love. Many people are in favour of this change as currently environment threats are the prominent threats and adoption of these vehicles will significantly reduce the carbon emission in the environment. Due to this reason many people are showing significant interest in Electric vehicles.

Sad: - In this category problem and cyber truck are the two highlighted words. It seems that some people are not happy with the design of the truck that tesla launched. People are anticipating more from the manufacturers. They need visually appealing designs of vehicles probably because it will heighten their social stature.

Surprise: - The prominent words in this category are Excited and SUV Cars. People are surprised and excited to see the hybrid designs of the new SUVs that falls under this category.

Anticipation: - Production and Delivery are the highlighted words in this category. These are the most significant word because people are anticipating that in future the production of electric vehicles will increase and subsequently with the increase in demand the delivery of the product will be streamlined.

Disgust: - Bad and ugly are the most disgust words shown by the people. It may be due to the charging capacity of the EV and their range anxieties.

Insert figure 4

Insert figure 5

In India (New Delhi), the author collected around 2800 tweets on the newly launched brands of EV's. After cleaning the data and removing re-tweets 1100 tweets were obtained. The majority polarity shows positive which is around 670, neutral around 220 and negative around 110.The higher positive sentiments shows that people has widely accepted the newly launched EV in India.

Insert figure 6

CONCLUSION

Electric vehicles (EV) proves as one of the emerging solutions to curb the pollution problem and move towards green development (Nimesh, Sharma, Reddy and Goswami, 2020). The paper address an important issue and has highlighted the opinions, reviews and sentiments of consumers towards adopting electric vehicles in Indian market using textual analysis of tweets and posts on social media. The above study highlights that factors such as cost, charging stations emerged as negative factors rather as factors of concern for the consumers to adopt using electric vehicles. However, majority of consumers showed their willingness to buy the electric vehicles provided if they get some government subsidy or tax benefits. Since electric vehicles run in batteries without fossil fuels, they can offer a solution of the storage to India's clean energy push. Efficiency of electrical mobility is more than mobility by petrol or diesel. EVs are an area of huge interest to India since it holds the potential of reducing the demand for liquid fuel to a greater extent. To reiterate, India's National Democratic Alliance government is promoting the EVs as the achievable strategic goal, which is keen to cut the country's oil imports. Adoption of EVs in India will lead to cleaner air and the noise level as well as the oil import bills will be down. If the EVs penetration goes up to 30% in India by 2030 it will generate an estimated saving of 474 Millions of tonnes of oil equivalent and 846 million tonnes of net CO2 emissions over their lifetime. Through sentimental analysis, the paper provided valuable insights on the consumers' willingness to buy electrical vehicles in India, which helps the automobile companies to have a head start in the E- vehicle segment and gives an early mover advantage in the industry. The paper also help the government and policy makers to get a fair idea of the sentiments of Indian consumers towards E-vehicle so that they can plan the subsidies and tax benefits for more penetration of EVs in Indian market.

REFERENCES

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Blok, K., and Nieuwlaar, E. (2020). Introduction to energy analysis. Routledge.

Browne, D., O'Mahony, M., and Caulfield, B. (2012). How should barriers to alternative fuels and vehicles be classified and potential policies to promote innovative technologies be evaluated?. Journal of Cleaner Production, 35, 140-151.

Bühler, F., Cocron, P., Neumann, I., Franke, T., and Krems, J. F. (2014). Is EV experience related to EV acceptance? Results from a German field study. Transportation Research Part F: traffic psychology and behaviour, 25, 34-49.

Cao, Q., Thompson, M. A., and Yu, Y. (2013). Retraction notice to" Sentiment Analysis in Decision Sciences Research. Decision Support Systems, 54(2), R1.

Chan, C. C. (2007). The state of the art of electric, hybrid, and fuel cell vehicles. Proceedings of the IEEE, 95(4), 704-718.

Dagsvik, J. K., Wennemo, T., Wetterwald, D. G., and Aaberge, R. (2002). Potential demand for alternative fuel vehicles. Transportation Research Part B: Methodological, 36(4), 361-384.

Dhar, S., Pathak, M., and Shukla, P. R. (2017). Electric vehicles and India's low carbon passenger transport: a long-term co-benefits assessment. Journal of Cleaner Production, 146, 139-148.

Egbue, O., and Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. Energy policy, 48, 717-729.

Fredrickson, B. L., and Losada, M. F. (2005). Positive affect and the complex dynamics of human flourishing. American psychologist, 60(7), 678.

Gallagher, K. S., and Muehlegger, E. (2011). Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology. Journal of Environmental Economics and management, 61(1), 1-15.

Hidrue, M. K., Parsons, G. R., Kempton, W., and Gardner, M. P. (2011). Willingness to pay for electric vehicles and their attributes. Resource and energy economics, 33(3), 686-705.

Huh, J., Yetisgen-Yildiz, M., and Pratt, W. (2013). Text classification for assisting moderators in online health communities. Journal of biomedical informatics, 46(6), 998-1005.

Jena, R. (2020). An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. Industrial Marketing Management. 90, 605-616. https://doi.org/10.1016/j.indmarman.2019.12.012

Jiang, Y., King, J. M., and Prinyawiwatkul, W. (2014). A review of measurement and relationships between food, eating behavior and emotion. Trends in Food Science and Technology, 36(1), 15-28.

Kumar, M. A. R., and Padmanaban, D. S. (2019). Electric Vehicles for India: Overview and Challenges. IEEE India Info, 14(2).

Lane, B., and Potter, S. (2007). The adoption of cleaner vehicles in the UK: exploring the consumer attitude–action gap. Journal of cleaner production, 15(11-12), 1085-1092.

Mau, P., Eyzaguirre, J., Jaccard, M., Collins-Dodd, C., and Tiedemann, K. (2008). The 'neighbour effect': Simulating dynamics in consumer preferences for new vehicle technologies. Ecological Economics, 68(1-2), 504-516.

Motwani, D. B., and Patil, A. (2019). Customer Buying Intention towards Electric Vehicle in India. International Journal of Mechanical Engineering and Technology, 10(5), 391-398.

Nimesh, V., Sharma, D., Reddy, V. M., and Goswami, A. K. (2020). Implication viability assessment of shift to electric vehicles for present power generation scenario of India. Energy, 195, 116976.

Pak, A., and Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

Singh, J. P., Dwivedi, Y. K., Rana, N. P., Kumar, A., and Kapoor, K. K. (2019). Event classification and location prediction from tweets during disasters. Annals of Operations Research, 283(1), 737-757.

Tuarob, S., Tucker, C. S., Salathe, M., and Ram, N. (2014). An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. Journal of biomedical informatics, 49, 255-268.

Utpal Bhaskar. (Nov 16, 2017). India's electric vehicle drive: Challenges and opportunities retrieved from https://www.livemint.com/Industry/ji96zXi5dZz3L1XUSkiZxM/Indias-electric-vehicle- drive-Challenges-and-opportunities.html, accessed on October 1, 2020.

Verma, M., Verma, A., and Khan, M. (2020). Factors Influencing the Adoption of Electric Vehicles in Bengaluru. Transportation in Developing Economies, 6(2), 1-10.

Table 1: Hashtags used for collecting tweets for the product – Electric vehicles.

Product	Hashtags (Combination of hashtags)
Electrical vehicle	#tigor , #kona , #mghector , #glanza , #hyundaikona #tatatigor , #toyotaglanza , #mg , #nexonev , #mgzsev #tycan , #bmwi3 , #etron , #ev , #electriccar , #ebike , #electricvehicle , #zeroemissions , #tesla , #zeroemission

Table 2: Count of discrete emotions

	count	emotion
anger	2330	anger
anticipation	6162	anticipation
disgust	1082	disgust
fear	2428	fear
joy	6710	joy
sadness	1746	sadness
surprise	4949	surprise
trust	6376	trust
negative	4836	negative
positive	15963	positive

Figure 1: Emotion classification Electric vehicles

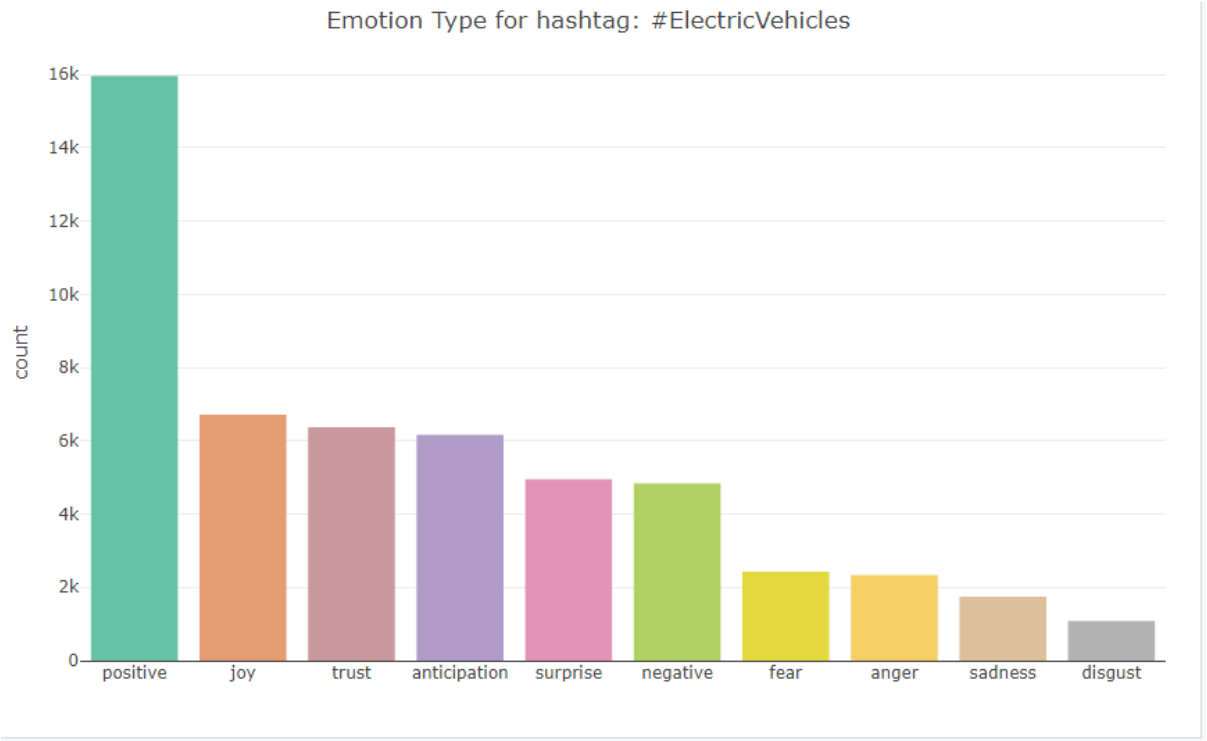


Figure 2: Sentiments of electric vehicles adoption

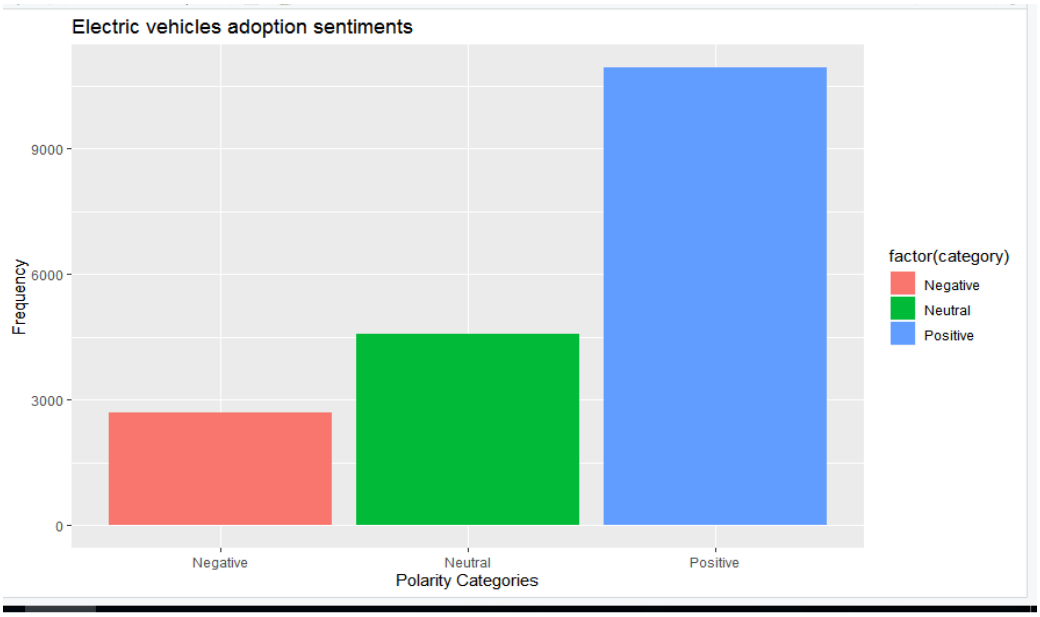


Figure 3: Word categorization based on emotion scores

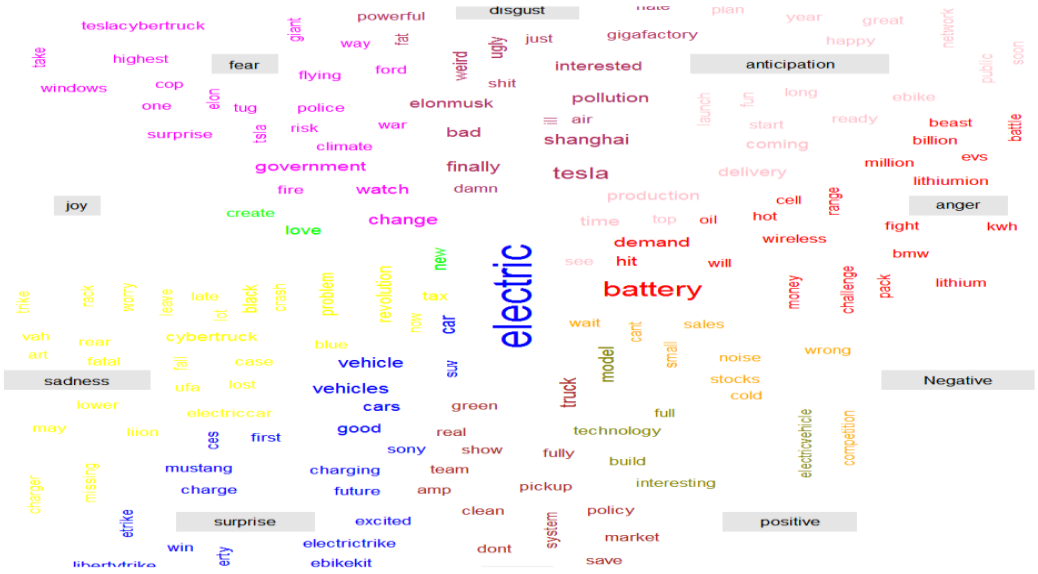


Figure 4: Electric vehicles sentiments in India

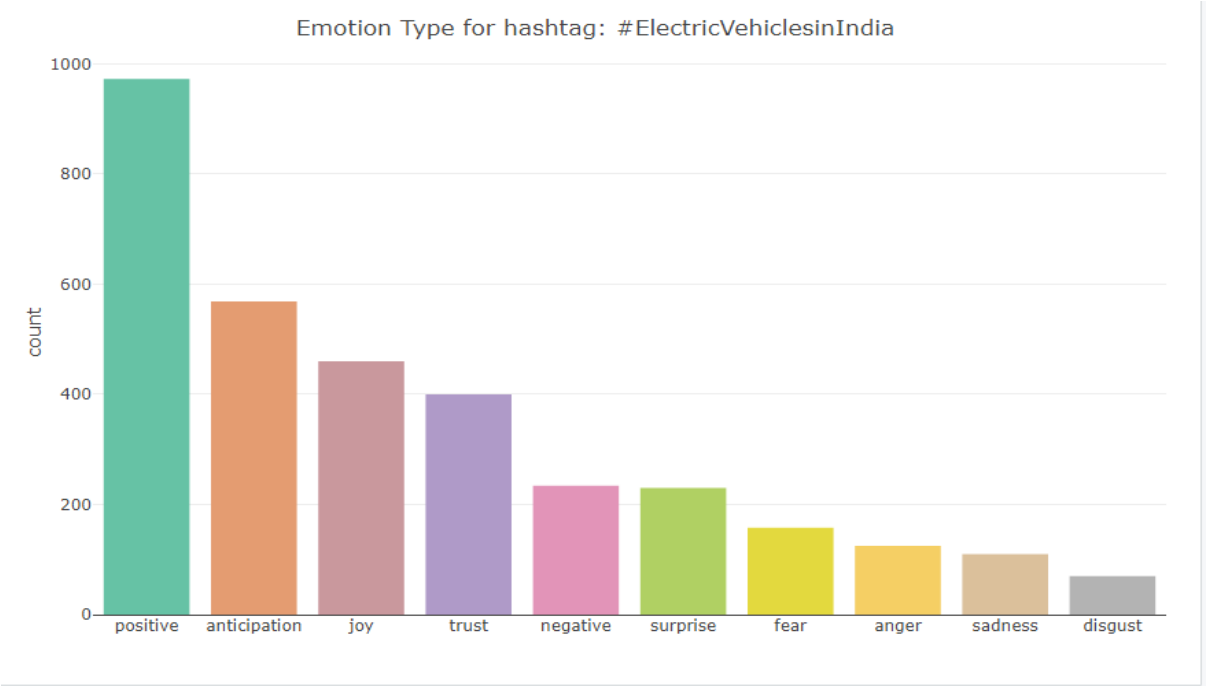


Figure 5: Sentiments of EV adoption in India

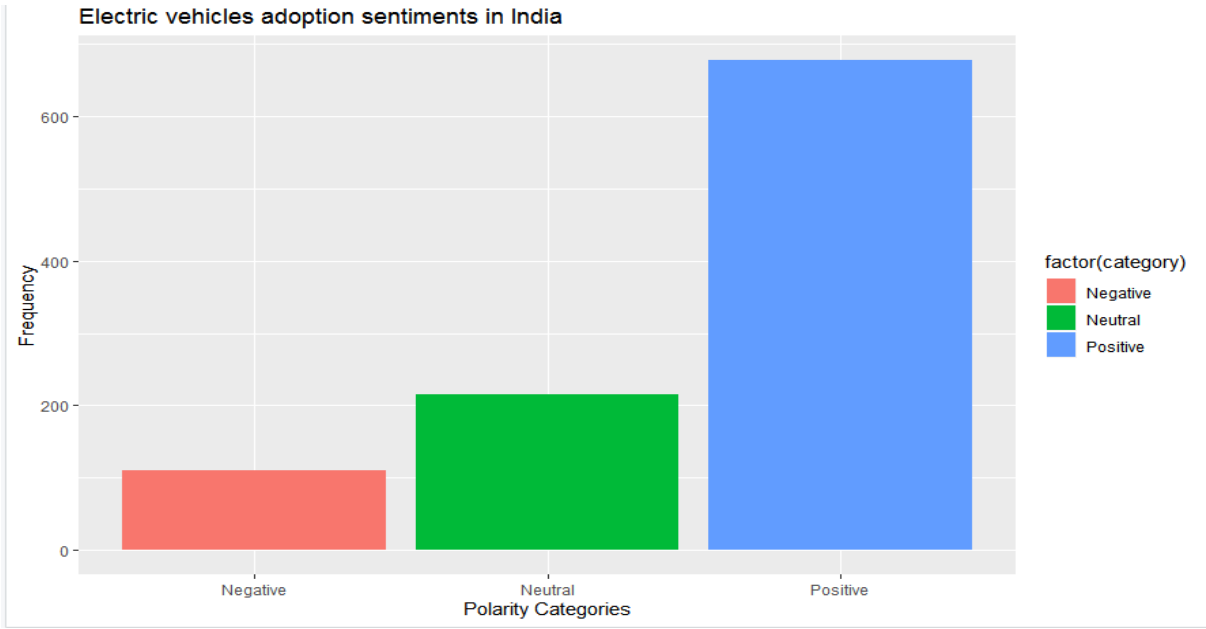
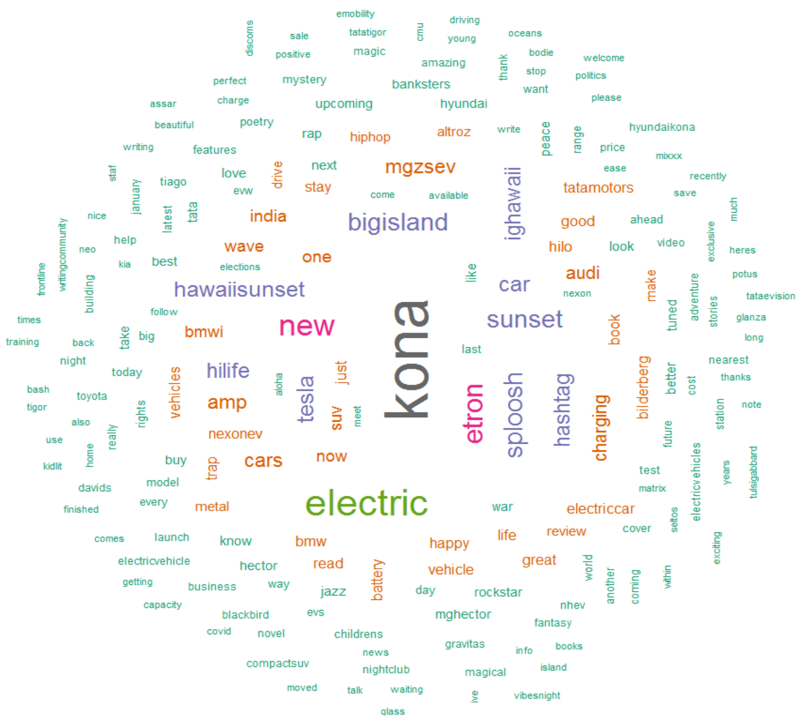


Figure 6: Word Cloud for EV tweets in India


kona
electric
new
bigisland
hawaiisunset
sunset
etron
splooosh
hashtag
ighawaii
tesla
hilife
mgzsev
charging
amp
cars
car
audi
india
wave
one
bmwi
now
nexonev
electriccar
tatamotors
bilderberg
hyundai
upcoming
banksters
altroz
hiphop
rap
stay
next
love
tiago
good
hilo
look
ahead
video
vehicle
happy
bmw
read
great
review
life
war
metal
know
hector
jazz
battery
rockstar
mghector
day
vehicles
suv
just
tuned
better
book
make
nearest
test
buy
model
today
toyota
big
take
back
night
best
help
neo
kia
nice
tiago
latest
tata
drive
poetry
features
beautiful
writing
assar
perfect
charge
mystery
magic
amazing
tatatigor
emobility
sale
positive
discoms
driving
young
oceans
bodie
stop
want
thank
welcome
politics
please
hyundaikona
price
range
peace
write
mixxx
ease
recently
save
much
heres
exclusive
potus
tataevision
glanza
long
stories
adventure
thanks
note
station
cost
future
electricvehicles
years
tulsigabbard
exciting
within
coming
another
world
cover
seltos
matrix
nhev
fantasy
books
info
island
gravitas
magical
vibesnight
live
waiting
talk
glass
nightclub
news
childrens
moved
compactsuv
novel
covid
blackbird
evs
way
business
capacity
getting
electricvehicle
launch
comes
finished
every
davids
kidlit
home
really
rights
trap
use
also
bash
tigor
training
times
frontline
writingcommunity
building
follow
elections
available
come
evw
january
staf
aloha
meet
last
like
nexon

Electric vehicles in India

DECISION SCIENCES INSTITUTE

Channel Choice and Incentive for Non-transplant Hospitals in Cadaveric Organ Supply Chain in India

Akansha Misra
Indian Institute of Management Bangalore
Email: akansha.misra14@iimb.ac.in

Haritha Saranga
Indian Institute of Management Bangalore
Email: haritha.saranga@iimb.ac.in

Rajeev R. Tripathi
Indian Institute of Management Bangalore
Email: rajeev.tripathi@iimb.ac.in

ABSTRACT

The shortage of transplantable organs is a concern in India. Non-transplant hospitals are potential sources of deceased organ donors (also known as cadaveric donors) but such hospitals have to be adequately incentivized to become organ retrieval hospitals. Additional challenge surfaces as organs are diverted from such hospitals to preferred recipients. We develop a mathematical model to determine the optimal fee that the government should use to incentivize such hospitals to join the authorized channel instead of an uncertain unauthorized channel. We find the choice between decentralized and cooperative decision-making is dependent on the cost-sharing factor and is not completely straightforward.

KEYWORDS: Health Care, Supply Chain Incentives, Public Sector, Cadaveric Organ Supply Chain

INTRODUCTION

The enormous demand for transplantable organs has consistently outstripped the supply across the world. The problem is more acute in some of the developing nations like India where approximately half a million patients succumb to organ failure while waiting for transplant due to lack of organs. A major concern in India is not lack of potential donors but the lack of mechanisms to convert the potential donors to actual donors. For example, almost 1.5 lakh brain-deaths are reported annually in India due to road accidents. A significant proportion of these victims could be potential organ donors (Srivastava and Mani, 2018). However, as the number of transplant centers is low and mostly concentrated in the urban areas, a sizeable volume of the road fatalities is admitted to hospitals that do not conduct transplants (ORGAN India report 2013). Since 2014, the legislation allows hospitals that do not conduct transplants but are equipped with some pre-requisite infrastructure to register as organ retrieval hospitals. Despite being assured monetary compensation for each donor, hospitals that are eligible for organ retrieval are not forthcoming. Consequently, a large pool of potential deceased donors, also known as cadavers, in these hospitals remains out of reach and unutilized. A pertinent question for the policy makers here is how to engage the hospitals eligible for organ retrieval in the cadaver organ donation system.

The organ donation rate in India stands at an abysmal 0.65 pmp. A well-functioning cadaveric organ system requires multiple entities to actively participate and coordinate. And to optimize cadaveric organ donation, it is important to remove disincentives in the system (Declaration of Istanbul 2008). It is imperative to evaluate if the eligible hospitals in India are being disincentivized in the current setting and if they could be incentivized to join the authorized channel for cadaveric organ donation. This problem forms the base of our study.

Extent OM literature has identified several mechanisms to align incentives of multiple parties to secure economic and social benefits at various levels, i.e. individual and system, in different contexts (Avci et al, 2014; Porteous et al, 2015; Arya and Mittendorf 2015; Guo et al. 2019). Organ sharing has broader implications on social welfare (Su and Zenios, 2006; Schummer 2016; Dai et al, 2019). Few studies have recommended contracts or mechanisms to align the decisions of the stakeholders with the social optimum (Arikan et al, 2018; Arora and Subramaniam, 2019). However, there is scant literature on the supply side of a cadaveric donor organ donation systems. With the exception of Arikan et al. (2018) and Arora and Subramaniam (2019) there are no other OM papers to the best of our knowledge that focus on the supply side entities.

A compelling motivation for the promotion of cadaveric donation program across the world was the rampant organ commerce and exploitation of living donors. A recent press report exposed a cartel formed between hospitals and transplant centers in an Indian city. The current rules also create a scenario of conflicting interests for transplant centers because the law permits a private transplant center, which declares a deceased patient to be a potential donor, to reserve some of the organs during the allocation process for its own patients. Thus, the transplant centers have an incentive to make declaration of deceased donor in their own facility (Srinivasan 2013, Nagral and Amalorpavanathan, 2014). In this paper, we refer to such tie-ups between hospital and transplant centers as the unauthorized channel.

Extent research corroborates that in India, religious and spiritual values take precedence in decision making regarding organ donation (Suresh and Mariappan, 2017; Srivastava and Mani, 2018). In addition to physical capacity, the approach to convert a potential donor to an actual donor requires effective communication, sensitivity, and patience (Adanir et al, 2014). As a result, the entire process appears costly both economically and emotionally (Jendrisak et al, 2002; Nagral and Amalorpavanathan, 2014). In this study, we include a cost-sharing parameter to allow the involvement of the coordinating organization (CO), the government body that oversees the cadaveric donor program, to participate and share this cost to ease the load on the hospital.

The specific questions we address in this study are: (i) From the hospital's perspective when is it optimal to choose the authorized channel over the unauthorized channel? (ii) How could the CO design its reimbursement scheme to divert the hospital from the unauthorized channel to the authorized channel? (iii) In the authorized channel, what is the impact of the hospital's participation on the decisions and the payoffs at the individual and the system level? (iv) If the hospital and the CO were to cooperatively decide on the reimbursement scheme, how would the decisions and the result change in comparison to the prior set up, which is essentially a decentralized one. The mathematical modeling approach is better suited to the analysis of the questions we have raised. Hence, we build a stylized model of the supply side entities of a cadaveric organ sharing system. We characterize the equilibrium decisions of the parties in our model.

Our main contribution to the ongoing research on healthcare operations is to model the reimbursement scheme of a CO, which can ensure that unauthorized practices are discouraged on the supply side of a cadaveric donor program. We highlight the role of reimbursement fee and the cost sharing factor on the system level and individual level payoffs. Specifically, we show that cooperative decision making is in the interest of the system and the hospital.

However, this result is dependent on the cost sharing factor. In fact, the cooperative system's total payoff fares worse than a decentralized set up when the cost sharing factor exceeds a threshold. The cross-channel analysis reveals interesting insight into how the coordinating organization could leverage the demand uncertainty in the unauthorized channel and induce hospitals to join the authorized channel by offering a fee lower than the price they get in the parallel channel as long as the fee is greater than a minimum amount. The formulations allow us to get an understanding of the factors that the CO could use as levers to stimulate the cadaver organ supply in the authorized channel.

Research Context

We discuss the cadaveric donor organ donation process in India. Figure 1 presents the three main entities and the interaction between them in a typical cadaveric organ donation process which involves organ retrieval at a hospital. The main entities include the hospital that performs the organ retrieval, the coordinating organization/authority (CO), and the transplant center. In India, the CO is constituted by the department of health and family welfare of the state government. We refer the department of health and family welfare of the state government as the planner. A description of the entities and the common terms associated with the cadaveric donor program is provided in Table 1.

Table 1: Description of the key terms in a cadaveric-donor organ supply chain

Transplant center	A hospital that is licensed to conduct organ transplants.
Hospital	Refers to a hospital that is not licensed to perform organ transplants but may be equipped to perform organ retrieval.
Coordinating Organization (CO)	State level organization which coordinates activities between the donor hospital/transplant center and the recipient transplant centers. It also conducts awareness programs to increase knowledge on organ donation among common public.
Deceased/Cadaveric organ	Organ retrieved from a donor who is brain dead
Brain death	It is an irreversible condition caused due to cessation of brain-stem activity. It is confirmed by neurological tests.
Transplant coordinator	A person appointed by a hospital/transplant center to coordinate all matters related to removal or transplantation of human organ with knowledge in the field of medicine, social work and public health.
Potential donor	A patient who is brain dead and whose organs could be donated for transplant purpose but the consent for donation is yet to be received.
Actual/Converted donor	A patient who is brain dead and whose organs could be donated for transplant and the consent for donation has been received.
Authorized channel	The channel where the potential donors are brought to the notice of CO and the organ allocation is done according to the central waiting list.
Unauthorized channel	The channel which is not legally recognized by the CO. Generally, it is used by transplant centers and hospitals to circumvent the transplant waitlist. This reduces the waiting time for the organ

	recipient and the transplant center accrues economic benefit in return.

A transplant center is a hospital that is licensed to conduct both organ retrieval and transplantation process. However, the organ retrieval process, which precedes the organ transplantation process, may be performed at other hospitals that are not licensed for transplant, but have ICU and operating room (OR) facilities. Such hospitals may get themselves registered to perform organ retrieval. For the ease of exposition, we simply use the term hospital to refer to the hospital that does not conduct transplants but may serve as an organ retrieval hospital.

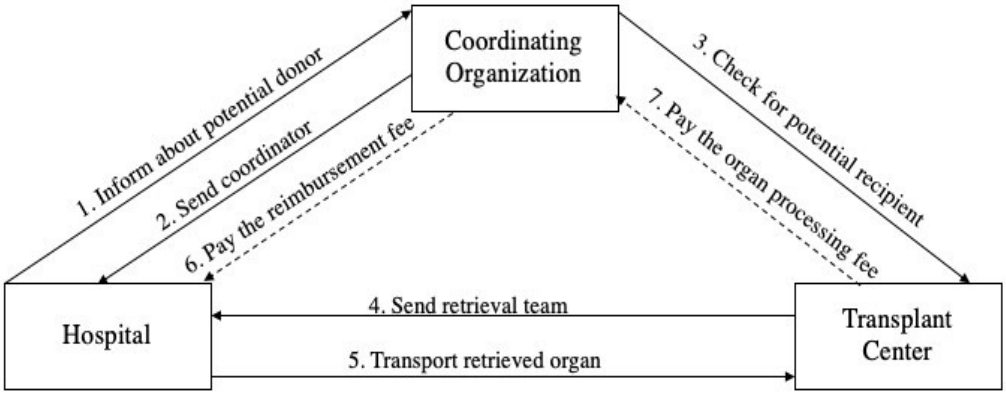


Figure 1. Description of the cadaveric organ donation process

Potential organ donors can be identified either at a transplant center or at a hospital. When the donor is identified at a transplant center, that transplant center gets priority during the organ allocation process. The transplant center can lay claim over some of the retrieved organs; generally, one kidney and some other organ if multiple organs are retrieved. The transplant center can use those organs for the patients waiting for transplant on its internal waitlist. The rest of the organs are allocated by the CO to other transplant centers. Alternatively, a potential donor could be detected in a hospital as depicted in Figure 1. In this case, the hospital informs the CO once it identifies that the patient is indeed a potential donor. The CO sends one of its coordinators to the hospital. Once consent from the donor's family is received, the CO allocates organs between the transplant centers according to a central waitlist. A team from the recipient transplant center generally travels to the hospital to retrieve the organs and returns with the retrieved organ. The CO reimburses the hospital by a fixed maintenance fee on a per donor basis. The CO further charges each receiving transplant center for each organ.

For the purpose of this study, we restrict the scope to the case where the hospital identifies the potential donor and focus on the supply side entities: the hospital and the CO. We also

consider a planner from the government side, who is responsible for regulating the entire system. The CO works under the ambit of the planner.

LITERATURE REVIEW

We draw from three different streams of literature and position our work and its contributions in the intersection of these three streams. Below, we first present a brief review of relevant publications in the Cadaveric Organ Sharing literature, which has major contributions from both scholars working in healthcare operations as well as in the field of Economics. Next we examine the Gray Market Operations related studies in the OM literature to compare and contrast the unauthorized organ donation market, which is an important channel in our context. Finally, we review the studies in Public Policy and Government Interventions to understand the role of policy in incentivizing hospitals to function in authorized markets and in increasing the social welfare through increased pool of organ supply.

Cadaveric Organ Sharing

Scholars from both healthcare operations and economics field have contributed to our understanding of the organ sharing system. Almost all of the papers reviewed in this area focus on kidney or liver allocation as these are the organs for which the demand exceeds the supply. Predominantly, the extant literature in the cadaveric organ sharing area has focused on improving the efficiency of organ allocation system or by affecting the individual behaviour (see Bertsimas et al, 2013; Zenios et al, 2000; Su and Zenios, 2006; Kong et al, 2010; Akan et al, 2012 for work in the first category and Kessler and Roth, 2012; Kessler and Roth, 2014; Dai et al, 2019 for work related to the second category). Besides efficiency, equity in organ allocation is also a concern. Zenios et al. (2000) and Su and Zenios (2006) study the dynamics between efficiency and equity in organ sharing. The effect of geographical inequity on the system's efficiency is studied by Kong et al. (2010) and Ata et al. (2017).

On the operations side, the existing literature has relatively fewer studies that examine the supply of cadaveric organs or the procurement rate from the hospital's end. Arora and Subramaniam (2019) and Arikan et al. (2018) belong to this category. For instance, Arora and Subramaniam (2019) study the impact of the decisions of the supply side entities of the organ donation value chain on their own payoffs and on societal outcome. Their paper considers a social planner whose objective is to maximize the quality-adjusted life-years after transplant. They design pareto optimal contracts to align the objective of the supply side entities with that of the social planner to achieve the socially optimal performance. The second example is of Arikan et al. (2018), who address the challenge of increasing procurement rate through slight change in the kidney allocation policy in the US.

Since organ sharing has broader implications on social welfare, Dai et al. (2019) examine the social welfare implications of the donor-priority rule. They propose a freeze period in the rule which could guarantee an increased social welfare. Schummer (2016) studies the effect of influencing the deferral behavior of agents in an organ waitlist. The author shows that the effect of the influence on welfare varies according to the risk behavior of the agents in the waiting queue. Su and Zenios (2006) consider aggregate utility and minimum utility social welfare functions. They study allocations under both objective functions.

Gray Market Operations

The general discussion around gray market in OM literature has focused on settings where the gray market serves as an alternate channel, which is unauthorized. For instance, a large chunk of the studies on gray market consider a retailer who diverts unsold goods to the gray markets at a price usually lower than the authorized channel prices (Dasu et al, 2012; Ahmadi et al, 2014; Shao et al, 2016). This is in contrast to the gray market for organs, where organs are sold at a higher price in the gray market. This contrast has to do with the nature of the goods in question. While it is attractive to reduce the unsold inventory of consumer products in alternate channels, in case of scarce goods like organs, a functioning gray market results in inequity and raises concern over the fairness of organ retrieval and allocation system. Similar to Altug and Sahin (2017), in our study, we consider cooperation between hospital and coordinating organization, for deciding the optimal reimbursement fee, which will increase the overall social welfare.

Public Policy and Government interventions

Incentives and penalties are among the most common instruments that policy makers employ to urge firms and consumers to adopt certain practices that could be beneficial either at individual consumer level, or societal level or environmental level. For instance, government could provide subsidies or price discount to boost electric vehicle adoption (Luo et al, 2014; Huang et al, 2013; Shao et al. 2017), to encourage corporate social responsibility (Arya and Mittendorf, 2015), to induce firms to move towards socially optimal remanufacturing (Zhang and Zhang, 2018), or to enhance the quality of life of the economically backward sections (Yu et al, 2018). According to Dai (2015), although operation scholars have applied the operations research techniques to specific problem within the healthcare set up, the matter of incentives has received rather scant attention. Since then, there have been studies that investigated monetary tools such as incentives or penalties in healthcare set up, that influence crucial measures such as waiting time (Qian et al, 2017), adoption of technology Health Information Exchange (Adjerid, Acquisti, Telang, & Padman, 2016) and excess readmissions (Zhang et al, 2016). Incentives could also be offered in a non-monetary form. A good example of an incentive offered to a patient can be found in Dai et al. (2019)'s study mentioned in the previous section, which is *priority on the waitlist*.

MODEL DESCRIPTION

We develop a model comprising of a private hospital and a coordinating organization (CO). We model only private hospital for two reasons. First, the monetary incentive scheme that we discuss is directed at private hospitals. Second, at present more than 90% of the cadaver organ donations take place in the private sector (Times of India report, 2017). First, we provide a timeline of the sequence of events in a cadaveric organ recovery at a hospital with a more detailed description than provided using Figure 1 in the Introduction section. Thereafter, we model the payoffs of the hospital in the unauthorized and the authorized channel and derive the minimum maintenance fee (also referred as reimbursement fee) that the CO should offer to the hospital to make the latter participate in the authorized channel. Subsequently, we analyze the individual and system level payoffs in a decentralized setting in the authorized channel. We follow it up with an analysis of the individual and system level payoffs in a cooperative setting in the authorized channel. This helps us in verifying if there are any advantages forthcoming from setting the fee in a cooperative manner.
Sequence of events in a cadaver organ recovery process as follows:

i) At the beginning of a period the CO decides a fixed maintenance fee per cadaver. The CO announces the maintenance fee to the candidate hospitals i.e. hospitals that meet the prerequisite to be an organ retrieval hospital.

ii) The hospital decides whether to become an organ retrieval hospital or not. If it does not accept the offer, it may operate in the unauthorized channel.

iii) Depending on the hospital's decision in the previous step, there can be two possible scenarios:

A) If the hospital operates in the authorized channel, the sequence of events that follows is given below:

a. In the event that a brain-dead patient is admitted to the hospital or a patient suffers brain-death while undergoing treatment in the hospital, the hospital identifies that the patient is brain dead and is a potential donor. The hospital follows the set protocol to certify the patient as brain dead and notifies the CO about the potential donor.
b. The transplant coordinator at the hospital informs the family of the deceased about the concept of brain death and organ donation. The CO also sends its coordinator to the hospital to convince and counsel the family members of the potential donor and tries to convert him/her to an actual donor.
c. If the family's consent is received, billing for the donor stops. The CO notifies the transplant centres about the available organs from the cadaver. If the transplant centre(s) confirm(s) that recipient(s) is(are) available, the CO allocates the organs to the patient(s) on the waitlist of those transplant centre(s).
d. In most cases, the transplant team from the transplant centre visits the hospital for the retrieval. The CO pays the hospital a maintenance fee for the cadaver.

B) If the hospital operates in the unauthorized channel [There is no fixed protocol of unauthorized channel operations. The sequence mentioned is based on reports cited in the media.], a possible sequence of steps could be:

a. In the event that a brain-dead patient is admitted to the hospital or a patient suffers brain-death while undergoing treatment in the hospital, the hospital identifies that the patient is brain dead and is a potential donor.
b. The hospital informs the family of the deceased about the concept of brain death and organ donation.
c. The hospital notifies the transplant centres about the cadaver. If a transplant centre confirms a match with one of their existing patients; the hospital transfers the cadaver to the transplant centre.
d. The transplant centre reimburses the hospital with a fee for the cadaver.

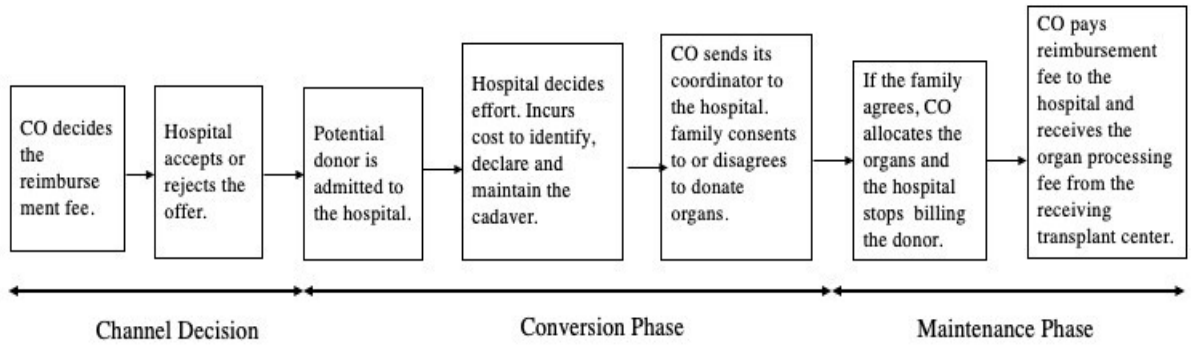


Figure 2. Timeline of events when Hospital is registered as an organ retrieval hospital

The sequence of events that follows the admission of a potential donor to a hospital is given in Figure 2. The timeline is given for a single episode of donor conversion. However, our model is at the aggregate level in a single time period and not at the individual donor level. We consider a single period where the CO offers a maintenance fee per cadaver to the hospital to become an organ retrieval hospital. The hospital's choice at this stage is referred to as the channel decision. The hospital can either choose to accept the offer and start operating as an organ retrieval hospital in the authorized channel (referred to as A) or reject the offer. If the hospital does not accept the offer, it may opt to operate in the unauthorized channel (referred to as S). Once the channel decision is made, the hospital selects the effort level in identifying a donor and convincing the donor's family. Hospital's choice of effort level dictates the level of manpower and resources it is willing to dedicate to identify and declare potential donors followed by the counselling and persuasion to get consent of the potential donor's family for organ donation. We refer to this stage as the conversion phase. The conversion phase stretches from the time hospital starts working towards identification and declaration of the brain-dead patient till the time counselling and convincing of the patient's family results in either a successful or a failed conversion. The outcome of the conversion phase is either a success or failure in gaining consent from the donor family. Only if the family agrees for donation, the potential donor becomes an actual donor and the maintenance phase begins. For identification and declaration of brain death, doctors with certain expertise (neurologist/neurosurgeon) are required. The procedure is time consuming as the law mandates that a panel of doctors confirm brain death twice with a gap of six hours between the two examinations. We represent hospital's effort as the fraction of potential donors towards which it can dedicate its resources. Let the total number of potential donors admitted to the hospital in a given time period be N. The effort level of the hospital be denoted by e_H^i, where $i \in \{A, S\}$, A and S represent authorized and unauthorized channels respectively. We follow the common practice in modelling effort decision by assuming that the effort level is bounded by a maximum and a minimum value, which we normalize to 1 and 0 respectively. From the discussion it follows that the total number of potential donors that the hospital follows up with the respective families would be further is Ne_H^i. We introduce another parameter viz. the effectiveness of the hospital in convincing the family of the donor, denoted by w. The total supply of actual donors from the hospital in a single period is thus given by $Ne_H^i w$.

For each potential donor who is approached for donation, effort is expensed, which involves a cost. This implies that even if the deceased donor's family declines the proposal, the cost of effort is incurred. Following the literature, we model the cost of effort to be convex in the

effort level. Thus, for a chosen effort level e_H^i, the cost is $c^i\left(e_H^i\right)^2$, where $i \in \{A, S\}$. From the discussion in the previous paragraph, we recall that the hospital follows up with a total of Ne_H^i potential donors. The total cost of effort in a single period is given by, $Ne_H^i * c^i\left(e_H^i\right)^2 = c^i N\left(e_H^i\right)^3$. The hospital receives a maintenance fee from the CO for each actual donor. Thus, the hospital receives maintenance fee for $Ne_H^i w$ converted donors.

We model the CO's objective as the benefit from the volume of transplants conducted (measured in monetary terms) minus the cost incurred in the period. The planner is concerned about the total welfare encompassing the hospital's as well as the CO's objective. The key decision variables, parameters and other variables used in the model are given in Table 2.

Table 2: Notations used in the model

Decision variable		
Hospital	e_H^i	Effort level of hospital in channel $i, i = \{S, A\}$; S: unauthorized channel, A: authorized channel; H: hospital
Coordinating organization	p^A	Maintenance fee in the authorized channel
Parameters		
	ξ	Fraction of cost of effort shared by the Coordinating organization, $\xi < 1$
	m	Probability of finding a match in the unauthorized channel
	w	Effectiveness of the Hospital in convincing the family to consent for organ donation
Other variables		
	N	Number of potential donors at the hospital
	c^i	$i = \{S, A\}$; S: unauthorized channel, A: authorized channel;
	p^S	Price reimbursed by the transplant center to the hospital in the unauthorized channel
	r	Sum of utility from a transplant (measured in monetary units) and organ processing fee charged by CO for an organ

The modeling assumptions are critical to the result and the interpretation of the results. We list the main assumptions used in the model in Table 3.

Table 3: Assumptions

A1	In the authorised channel, the identification of a cadaveric donor is analogous to conversion of the donor as the probability of finding a donor through the authorised channel is very high and is assumed to be equal to one.
A2	Price in the unauthorized channel is given
A3	The CO doesn't have any decision to make in the unauthorized channel

At first, we conduct our analysis at the cross-channel level where we analyze the optimal channel decision of the hospital in response to the coordinating organization's reimbursement scheme. Thereafter, we consider the case wherein the hospital has registered as an organ retrieval hospital and study the structure under decentralized and cooperative set up.

Unauthorized Channel

Hospital's Problem

In the unauthorized channel, hospital deals with the transplant centers directly. The possible steps in the unauthorized channel is described in part (iii)B of the sequence. We do not discuss the price setting mechanism between a transplant center and the hospital and assume that the price is given in the unauthorized channel (assumption A2). In case of a matched demand, the transplant center offers price, p^S, which the hospital accepts. The hospital has access to a limited number of transplant centers with which it can deal directly in the unauthorized channel. This implies that there is a chance that an exact match may not be obtained. Hence, upon identification of a potential donor, it is possible that there is no matched demand. We assume that the probability of finding a matched demand in the unauthorized channel is m. Therefore, the actual supply of converted donors in the unauthorized channel is given by: $Ne_H^S wm$. Therefore, in a single period, the total reimbursement that the hospital receives is: $p^S Ne_H^S wm$. The total cost in a single period for all potential donors is $c^S N\left(e_H^S\right)^3$. Encapsulating the features discussed above, we construct the payoff function of the hospital in the unauthorized channel as:

$$\Pi_H^S = p^S Ne_H^S wm - c^S N\left(e_H^S\right)^3 \quad (1)$$

As per assumption 3, when the hospital operates in the unauthorized channel, the CO does not have any decision to make. Hence, the CO's problem is relevant only in the authorized channel.

Authorized Channel

Hospital's Problem

In the authorized channel, the hospital receives a fixed maintenance fee, p^A, for each actual donor. The CO allocates the organs to recipients from a large pool of patients on the waitlist awaiting transplant. We assume that the list is sufficiently large so that all the donors identified by the hospital find a match in this channel (Assumption A1). In the authorized channel, the CO also gets involved in some intermediate steps in the conversion phase thereby sharing some of the responsibilities in the process. The CO may also assist the hospital in imparting training to the coordinators and preparing the hospital to handle the identification of the potential donors in a better way. For instance, the CO in the state of Karnataka sends its own coordinator to the hospital as soon as it is intimated about a potential cadaver. Due to the participation of the CO, the cost of effort of the hospital reduces to some extent. Therefore, we model CO's participation in the process in the form of sharing of effort cost. The cost component in the hospital's problem includes a factor $(1-\xi)$. Here, ξ denotes the fraction of the cost of effort borne by the CO, and $\xi \in (0,1]$. Equivalently, hospital's cost of effort reduces by a factor, ξ. We restrict the value of the factor ξ between 0 and 1.
The payoff function of the hospital is given by:

$$\Pi_H^A = p^A N e_H^A w - (1-\xi) c^A N\left(e_H^A\right)^3 \quad (2)$$

We note that the objective function given in (2) is concave in the effort level. Thus, we use the first order conditions to derive hospital's optimal effort as a function of the maintenance fee in the authorized channel:

$$e_H^A(p^A) = \sqrt{\frac{p^A w}{3c^A(1-\xi)}} \quad (3)$$

From expression (3), we note that the hospital's best effort level chosen in response to a given maintenance fee, is *increasing in the maintenance fee ,* p^{A} *and decreasing in the cost of effort,* c^A*.* This implies that the hospital has a higher incentive to put in more effort to convert potential donors when the hospital is assured a high maintenance fee, and the corresponding cost of effort is low.

Coordinating Organization's Problem

The CO associates a utility with each organ transplant which we represent in monetary terms. The CO allocates the organs to transplant centers and in return charges an organ processing fee per organ to the receiving transplant centers. We use a single term, r, to represent the sum of the utility from each transplant represented in monetary terms and the organ processing fee per organ. Thus, for each transplant that is conducted in the system, the CO gains an amount r. The CO also reimburses the hospital for the maintenance of the donor with a maintenance fee, denoted by p^A, per converted donor. The CO may participate in the conversion phase of the donor family at the hospital. The participation of the CO brings down hospital's cost of effort by ξ. In the CO's payoff function, we represent this by subtracting ξ part of hospital's effort cost from CO's total payoff. The payoff function of the CO is represented as:

$$\Pi_G^A = rNe_H^A w - p^A N e_H^A w - \xi c^A N\left(e_H^A\right)^3 \quad (4)$$

MODEL ANALYSIS AND RESULTS

Decentralized Case

Hospital's Channel Decision

Here, we find the settings under which the CO could induce the hospital to choose the authorized channel over the unauthorized channel. We find the lower bound on the authorized channel maintenance fee as the smallest p^{A} for which under the same choice of effort the hospital receives equal payoff from both channels i.e. $\Pi_{\mathrm{H}}^{\mathrm{A}} = \Pi_{\mathrm{H}}^{\mathrm{S}}$ holds. On comparing the hospital's payoff in the two channels, we find that for a given effort level, CO's fee decision is contingent on the level of uncertainty in the channel. We present the result below:

Proposition 1. If the uncertainty in the unauthorized channel is higher than a threshold level of uncertainty, m_t, then there exists a lower bound, p_{lb}, on the authorized channel's maintenance fee such that as long as $p^A > p^A_{lb}$, the following is true:
For a given effort level, hospital's payoff in the authorized channel is greater than the hospital's payoff in the unauthorized channel. The result holds even if the maintenance fee is lower than the unauthorized channel price.

If $(1-m) > m_t = \frac{(c^A(1-\xi)-c^S)e^2}{p^S w}$ holds,

And, if in addition, $p^A > p^A_{lb} = mp^S + \frac{(c^A(1-\xi)-c^S)e^2}{w}$ holds, then even for $p^A < p^S$

$$\Pi^S_H < \Pi^A_H$$

Else, only if $p^A > p^S$

$$\Pi^S_H < \Pi^A_H$$

The lower bound on the fee offered by the CO to the hospital for it to choose authorized channel over unauthorized one is given by:

$$p^A_{lb} = \begin{cases} mp^S + \frac{(c^A(1-\xi)-c^S)e^2}{w} & \text{if } (1-m) > m_t = \frac{(c^A(1-\xi)-c^S)e^2}{p^S w} \\ p^S & \text{otherwise} \end{cases} \quad (5)$$

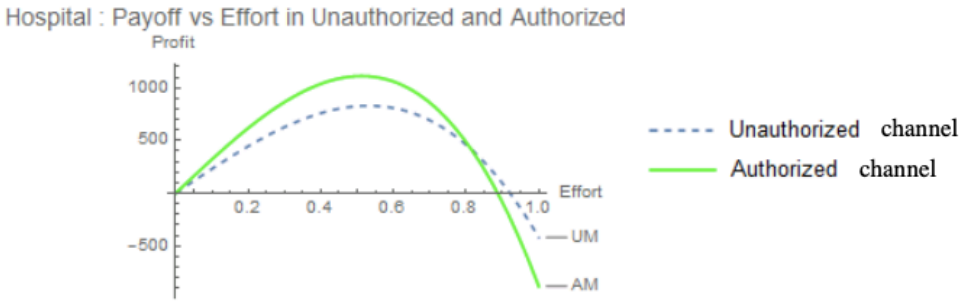


Figure 3. Hospital payoff in the authorized and the unauthorized channel

An instance of the hospital's payoff under both channels is depicted in Figure 3. From the figure, we find that under certain settings authorized channel yields higher payoff to the hospital in comparison to the unauthorized channel.

Proposition 1 shows that, while it is intuitive that to incentivize a hospital, the CO should set the maintenance fee more than the unauthorized channel price, under certain conditions, the CO could induce the hospital to choose the authorized channel even if the maintenance fee is lesser than the unauthorized channel price. The latter happens when two conditions hold: first, the uncertainty in the unauthorized channel is higher than a threshold level of uncertainty, m_t and second, the maintenance fee is greater than a minimum threshold fee. The result helps to

establish the lower bounds on the CO's fee decision. The lower bound depends on the cost structure in the unauthorized and the authorized channel. We note that the expression $\mathrm{p^S w(1-m)}$ can be interpreted as the cost of uncertainty in the unauthorized channel because the hospital loses out on an amount equal to the expression $\mathrm{p^S w(1-m)}$ whenever there is no match in the unauthorized channel. This loss is attributed to the presence of uncertainty in the unauthorized channel as with probability $(1-\mathrm{m})$ no match is found in the unauthorized channel. However, note that hospital still incurs a cost in identifying the donor.
The result indicates that if the sum of the cost of uncertainty and the cost of effort in the unauthorized channel is greater than the cost of effort in the authorized channel, then the hospital should prefer the authorized channel for a maintenance fee less than $\mathrm{p^S}$ but greater than the minimum amount. The lower bound denotes that the CO's fee for each donor, $\mathrm{p^A}$, should be high enough so that the net payoff in the authorized channel exceeds the payoff that the hospital receives from the unauthorized channel. However, if the combined cost of uncertainty and the cost of effort in the unauthorized channel is lower than the cost of effort in the authorized channel, then the fee that CO should offer to the hospital has to be greater than or equal to $\mathrm{p^S}$.

CO's Maintenance Fee Decision

The CO solves a constrained optimization problem subject to the lower bound on the maintenance fee.

$$max_{p^A}\ \Pi_G^A = rNe_H^A w\ - p^A N e_H^A w - \xi c^A N\left(e_H^A\right)^3$$

$$subject\ to\colon \begin{cases} mp^S + \dfrac{(c^A(1-\xi)-c^S){e_H^A}^2}{w} \le\ p^A & if\ \ p^S > \dfrac{(c^A(1-\xi)-c^S)e^2}{(1-m)w} \\ p^S \le\ p^A & otherwise \end{cases} \tag{6}$$

The CO's problem is a constrained maximization problem with inequality constraints. Hence, we consider the Karush-Kuhn-Tucker (KKT) conditions, which provide the necessary and sufficient conditions for the optimal fee decision.

Lemma 1.

(i) When$(1-\mathrm{m}) \ge \mathrm{m_t}$, there exists a critical value of the unauthorized channel price, denoted by $\mathrm{p_{C1}^S} = \frac{2\mathrm{r}(1-\xi)+\frac{\mathrm{rc^S}}{\mathrm{c^A}}}{3\mathrm{m}(3-2\xi)}$, such that the following holds:

a. If $\mathrm{p^S} > \mathrm{p_{C1}^S}$, then the optimal maintenance fee in the authorized channel, $\mathrm{p^A}$ is $\frac{3\mathrm{c^A}(1-\xi)\mathrm{mp^S}}{2\mathrm{c^A}(1-\xi)+\mathrm{c^S}}$.

b. Otherwise, the optimal value of $\mathrm{p^A}$ is given by $\frac{\mathrm{r}(1-\xi)}{3-2\xi}$.

(ii) When$(1-\mathrm{m}) < \mathrm{m_t}$, there exists a critical value of the unauthorized channel price, denoted by $\mathrm{p_{C2}^S} = \frac{\mathrm{r}(1-\xi)}{3-2\xi}$, such that the following holds:

a. If $p^S > p^S_{C2}$, then the optimal maintenance fee in the authorized channel, p^A is same as p^S.
b. Otherwise, the optimal value of p^A is given by $\frac{r(1-\xi)}{3-2\xi}$.

In Lemma 1, case (i) arises when the uncertainty in the unauthorized channel is higher than a threshold value. In the presence of high uncertainty, if the prevailing price in the unauthorized channel is lower than a critical price, the CO can ignore the unauthorized channel and set the maintenance fee independent of the unauthorized channel characteristics such as the price, cost of effort and the level of uncertainty. The optimal value of the maintenance fee is given in (i)b. However, in the occasion when along with high uncertainty, the unauthorized channel price is also high enough, the CO cannot ignore the unauthorized channel features. The optimal decision of the CO is dependent on cost of effort, price and uncertainty in the unauthorized channel and the optimal maintenance fee is given by (i)a.
Scenario (ii) occurs when the uncertainty in the unauthorized channel is low. In this case, the CO has to offer a maintenance fee which is at least $\frac{r(1-\xi)}{3-2\xi}$. However, if the unauthorized channel price is above a critical value, the CO has to offer a maintenance fee equal to p^S. This scenario, with a high probability of finding a matched demand coupled with a high reimbursement price in the unauthorized channel presents a strong case for the unauthorized channel. It is the ideal scenario where unauthorized channel can flourish. Hence, in this setting, the CO is left with no other choice other than to match its fee with the unauthorized channel price.

The result stated in Lemma 1 (i)a and (ii)a elaborates the result in proposition 1 for different situations. While, cases (i)b and (ii)b represent the scenario where the unauthorized channel price is less than the critical price and hence it does not impact the CO's maintenance fee decision. The result from the two cases also correspond to the situation when there is no functional unauthorized channel. In this case, only the authorized channel features, such as the positive payoff from transplants and CO's degree of participation in the conversion phase of the process, guide the CO's maintenance fee decision.

Proposition 2.

a. When both uncertainty level and price in the unauthorized channel are high, the optimal maintenance fee increases with an increase in the cost of effort in the authorized channel i.e. c^A. The optimal effort level decreases with an increase in the cost of effort in the authorized channel.
b. The maintenance fee, p^A, decreases in the coordinating organization's degree of participation in conversion phase, ξ, whereas, the effort of the hospital, e^A_H, increases in the coordinating organization's degree of participation in conversion phase.
The impact of coordinating organization's degree of participation is higher on the maintenance fee than on the hospital's effort i.e. for a given increase in ξ, the decrease in optimal maintenance fee is greater in magnitude than the corresponding increase in hospital's effort.

High uncertainty to find a match will deter the hospital from engaging in the unauthorized operations. Whereas, a high price in the unauthorized channel will attract the hospital to the unauthorized channel. Hence, these two opposing forces will guide hospital's decision. From Lemma 1 we know that in scenarios when the unauthorized channel price is low, the CO is in a position to offer a maintenance fee based solely on its individual parameters such as the payoff

from each transplant and its level of participation in the conversion phase. Also, when the uncertainty is low and unauthorized channel price is high, the CO is compelled to match the maintenance fee with the price. Only under the conditions given in Proposition 2a will the CO be forced to set a fee, which increases with an increase in the cost of effort so that the hospital stays in the authorized channel. The second part of Proposition 2a is intuitive as with any increase in the cost of effort, hospital will be inclined to exert lesser effort. While the cost of effort matters to the CO only under specific conditions, it is a deciding factor in hospital's effort decision always.

Proposition 2b underscores that both maintenance fee and effort level are impacted by the CO's participation level in the conversion phase but in different ways. It is understood that if the CO is contributing significantly to the conversion phase by facilitating the process either through taking upfront initiatives to upskill the hospital staff or through active participation in the conversion process, the CO will tend to offer a lower maintenance fee. At the same time, the hospital will find it easier to exert higher effort when it is assured that the CO will also assume some part of the cost of effort. The CO's maintenance fee is more sensitive than hospital's effort level to the change in CO's participation. Hence, if the CO increases its participation in the conversion phase, it would lead to an increase in effort by the hospital and at the same time a decrease in the maintenance fee. The corresponding decrease in the maintenance fee would be more than the increase in the cost of effort resulting in an overall gain for the CO.

The result indicates that the CO should strive to make it easier for the hospitals to operate in the authorized channel by facilitating the conversion process. To some extent the COs already practice this. However, COs can contribute in a broader way such as providing training to the transplant coordinators at the hospitals.

Cooperative Case

In the cooperative scenario, we consider a set up where the CO and the hospital decide unanimously on the maintenance fee using the Nash bargaining approach. We use the results from this setting to contrast against the existing scenario, which is the decentralized setup. Note that only the maintenance fee decision is made cooperatively. The hospital decides an effort level based on its own objective unilaterally and is the same as given in (3). The Nash bargaining approach arrives at a unique solution that decides the payoff allocation between the hospital and the CO.

The hospital can operate in the authorized channel only if it operates as per the planner's guidelines. Thus, the hospital cannot operate independently in the authorized set up. In the absence of the hospital, the CO too cannot create value with the potential donors admitted to the hospital in the channel independently. Hence, we assume that independently, the CO and the hospital, cannot create any value. The Nash bargaining approach maximizes the utility of both entities. The optimal maintenance fee is given as a solution to the problem:

$$\arg\max_{p} \Pi_{GH}^{A} = \Pi_{G}^{A} * \Pi_{H}^{A}$$

$$= \left(rNe_{H}^{A}w - p^{A}Ne_{H}^{A}w - \xi c^{A}N\left(e_{H}^{A}\right)^{3}\right)\left(p^{A}Ne_{H}^{A}w - (1-\xi)c^{A}N\left(e_{H}^{A}\right)^{3}\right) \quad (7)$$

We substitute the value of e_{H}^{A} from equation (3) into the objective function of the CO.

The optimal maintenance fee obtained using the Nash bargaining approach between the hospital and the CO is given by

$$p_C^{A^*} = \frac{2r(1-\xi)}{3-2\xi} \quad (8)$$

The optimal effort, the best response to the maintenance fee is given by:

$$e_C^{A^*} = \sqrt{\frac{2rw}{3c_A(3-2\xi)}} \quad (9)$$

And the resultant payoffs of the hospital and the CO in the cooperative set up are given by:

$$\Pi_{HC}^{A} = \frac{4\sqrt{2}(1-\xi)}{3\sqrt{3}(3-2\xi)^{\frac{3}{2}}} \frac{Nw^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c_A}} \quad (10)$$

$$\Pi_{GC}^{A} = \frac{\sqrt{2}}{3\sqrt{3(3-2\xi)}} \frac{Nw^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c_A}} \quad (11)$$

Lemma 2.
In the cooperative set up, the total system payoff is maximized at $\xi = 0.5$. Total system payoff is non-monotonic in ξ.
If $\xi < 0.5$, hospital's payoff is greater than the CO's payoff. Else the CO's payoff is greater than the hospital's payoff.

The Nash bargaining framework allocates the total payoff equally between the hospital and the CO when both the parties contribute equally to the conversion phase. Figure 4 illustrates Lemma 2.

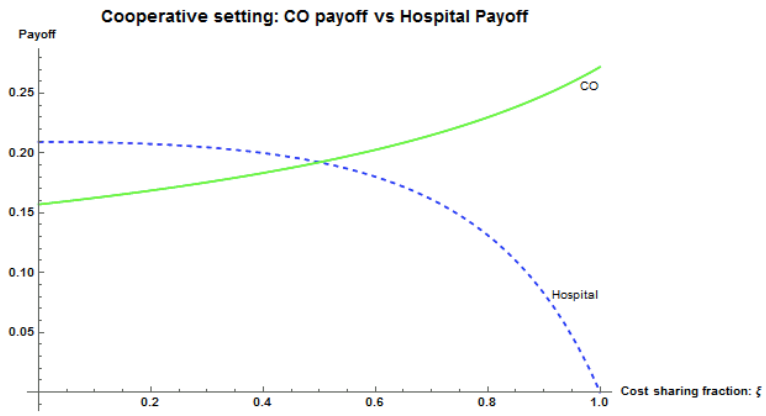


Figure 4. Hospital payoff vs Coordinating Organization payoff in the Cooperative scenario for different values of ξ

The CO faces a trade-off between paying the maintenance fee and incurring the cost of participation in the conversion phase. The impact of the CO's choice has repercussions on the optimal effort level of the hospital.

Specifically, when the CO is less participative in the effort to convince the donor's family i.e. ξ is lower than 0.5, the CO saves cost of participation. However, it has to pay a higher maintenance fee to the hospital. From Proposition 2 we already know that as the CO's level of participation decreases, the effort exerted by the hospital also decreases. The lack of support from the CO's end makes the entire exercise costly for the hospital. The hospital itself is not inclined to exert a higher effort. The lower effort from hospital results in lesser number of converted donors, thereby lowering the aggregate payoff from transplants for the CO. From Proposition 2 we also know that for a given decrease in CO's participation level the increase is maintenance fee is greater than the decrease of participation cost. Thus, for the CO, the cost savings from reduced participation is offset by the decrease in payoff due to a lesser number of transplants and a higher maintenance fee per donor. In this scenario, hospital exerts lesser effort resulting in lower number of transplants while receiving a higher maintenance fee for each converted donor. Hence, the hospital ends up capturing a higher share of the pie.

On the contrary, when the CO has higher participation in the conversion phase (when $\xi > 0.5$), the hospital also exerts a higher effort level. This subsequently leads to more converted donors. Even though the maintenance fee is lower than the previous case (when $\xi <$ 0.5), the CO manages to achieve a higher payoff in comparison to the previous case through more number of transplants and a lower maintenance fee. However, the CO incurs a higher cost of participation compared to the previous case. But, from Proposition 2 we know that for the CO, the increase in the positive payoff is higher than the corresponding increase in costs from participation. Hence, the CO is better off in this scenario. Arguing in a similar fashion, we find that for the hospital, a high ξ promises greater cost sharing but the advantage from cost sharing is superseded by the decrease in the maintenance fee and increase in effort. Thus, the hospital receives a smaller share of the pie in this scenario.

Hence, ξ = 0.5 marks the boundary between the two regions where the payoffs of the hospital and the Planner are always contradictory. This is evident from Figure 4. The point of agreement will arise when ξ is close to 0.5 to maximize the system level payoff. Therefore, in the cooperative set up, it is in the interest of the CO to participate to a certain extent in the conversion phase of the entire episode to encourage hospitals to exert higher effort in the conversion phase.

Comparison of the Decentralized and the Cooperative Cases

We compare the payoffs under the cooperative set up and the decentralized set up at the individual level and the system level. For the ease of comparison, we use only one of the optimal values of the maintenance fee of the CO from Proposition 2. We choose the maintenance fee for the scenario when the prevailing price in the unauthorized channel is low, irrespective of the uncertainty levels in the unauthorized channel. This is because in this setting, the response of the CO is based on the authorized channel characteristics and this facilitates the comparison with the cooperative case.

A comparison of the decision variables and the payoffs of each party for the decentralized case and the cooperative case is given in Table 4.

Lemma 3. *The optimal maintenance fee in the cooperative setting is twice that of the optimal maintenance fee in the decentralized set up. The optimal effort in the cooperative case is* $\sqrt{2}$ *times of the optimal effort in the decentralized set up.*

Lemma 3 implies that in the cooperative set up the hospital always receives double the amount of maintenance fee that it receives in the decentralized set up. The result is not surprising because in the cooperative scenario the maintenance fee is set by considering not only the CO's objective but also accommodating the hospital's payoff. Setting the maintenance fee in this way favors the hospital.

While maintenance fee goes up by two times, the corresponding increase in effort is not the same, the latter increases by $\sqrt{2}$ times only. This is because effort is still decided unilaterally by the hospital. Although the higher maintenance fee does inspire the hospital to put in more effort, yet the hospital has the same cost of effort; thus, the optimal response allows the hospital to increase the effort by a factor less than the corresponding increase in the maintenance fee.

Further, it is straightforward to see the impact of the increased maintenance fee and effort on the payoffs. The CO's payoff is always less in the cooperative scenario by a factor of $\sqrt{2}$ times than the payoff it receives in the decentralized scenario. On the contrary, the hospital's payoff is always greater in the cooperative scenario by a factor of $2\sqrt{2}$ compared to its payoff in the decentralized scenario. Hospital will always prefer that the CO sets the maintenance fee such that the hospital receives a higher share of benefit accruing from the increased donations.

Proposition 3. *If $\xi < 0.76$, the total payoff in the cooperative case is more than the payoff in the decentralized case. Else, the payoff in the decentralized case is higher than the payoff in the cooperative case.*

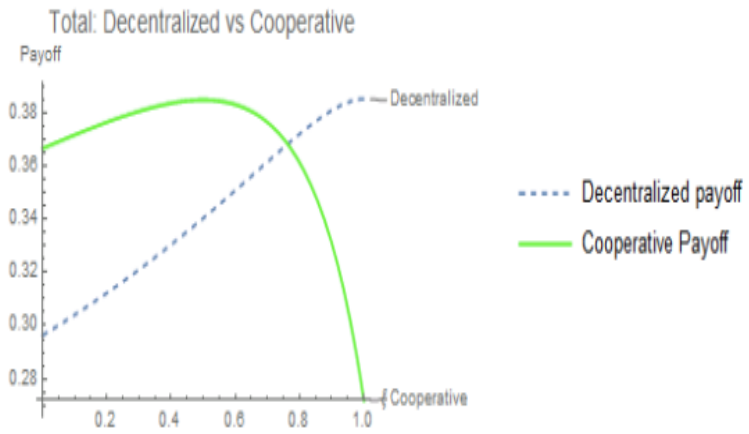


Figure 5a. Total payoff in the Decentralized and Cooperative scenario for different values of ξ

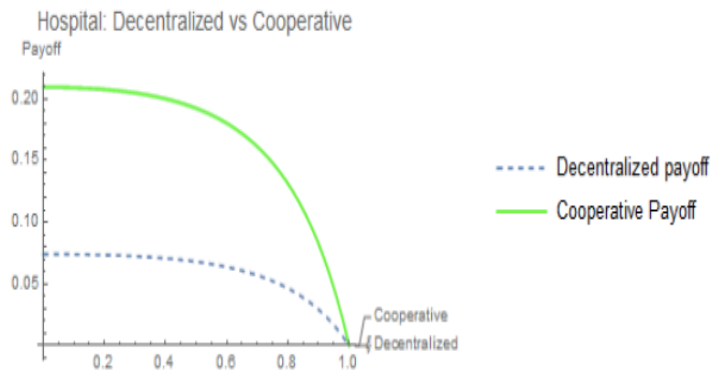


Figure 5b. Hospital's payoff in the Decentralized and Cooperative scenario for different values of ξ

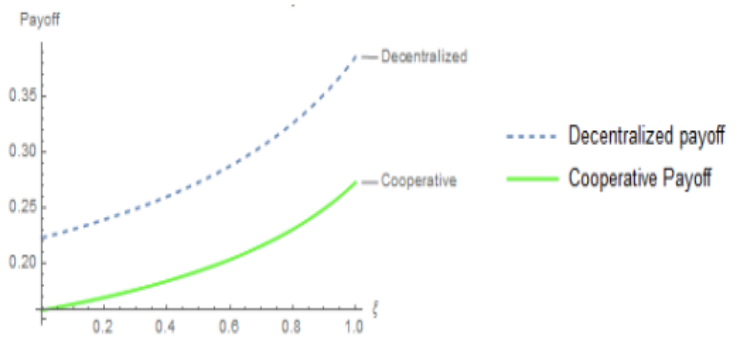


Figure 5c. CO's payoff in the Decentralized and Cooperative scenario for different values of ξ

Figure 5 a, b and c depict how the payoffs vary with ξ for at the system and individual player level. In general, it is expected that cooperative setting will lead to better payoff at the system level than decentralized set up. However, proposition 3 shows that this may not always be the case. The contradictory finding arises when the CO's participation is very high ($\xi > .76$). The role of the Planner is significant here. If most of the work in the conversion phase is done by the CO itself such as convincing the family and providing the counselling, the maintenance fee decreases rapidly. The Planner sees no benefit in letting the CO and the hospital set the maintenance fee cooperatively. In which case, it is optimal at the system level to not set the maintenance fee cooperatively. Hence, the Planner would consider the cooperative set up only when the CO's participation is not very high. For the Planner, the benefits from indulging in cooperative set up are forthcoming only when the hospital and CO put comparable effort.

Combining the insights from Lemma 2 and Proposition 3, we can say that the Planner would prefer the system optimal level of participation from the CO i.e. almost equal share of cost of effort between the hospital and the CO.

Table 4: A comparison of the decision variables and payoffs under the decentralized and cooperative set up

	Decentralized case analysis	Cooperative case analysis
Hospital	$\Pi_H^A = p^A N w e_H^A - (1-\xi) c^A N\left(e_H^A\right)^3$	
Coordinating organization/ Government	$\Pi_G^A = r N w e_H^A - p^A N w e_H^A - \xi c^A N\left(e_H^A\right)^3$	
		$\Pi_{GH}^A = \Pi_G^A * \Pi_H^A = \left(r N w e_H^A - p^A N w e_H^A - \xi c^A N\left(e_H^A\right)^3\right)\left(p^A N w e_H^A - (1-\xi) c^A N\left(e_H^A\right)^3\right)$
$e_H^A(p^A)$	$\sqrt{\frac{p^A w}{3c_A(1-\xi)}}$	$\sqrt{\frac{p^A w}{3c_A(1-\xi)}}$
$\mu = 0$: p^{A^*}	$\frac{r(1-\xi)}{3-2\xi}$	$\frac{2r(1-\xi)}{3-2\xi}$
$e_H^{A^*}$	$\sqrt{\frac{rw}{3c_A(3-2\xi)}}$	$\sqrt{\frac{2rw}{3c_A(3-2\xi)}}$
Π_H^A	$\frac{2(1-\xi)}{3\sqrt{3}(3-2\xi)^{\frac{3}{2}}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c^A}}$	$\frac{4\sqrt{2}(1-\xi)}{3\sqrt{3}(3-2\xi)^{\frac{3}{2}}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c_A}}$
Π_G^A	$\frac{2}{3\sqrt{3(3-2\xi)}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c^A}}$	$\frac{\sqrt{2}}{3\sqrt{3(3-2\xi)}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c_A}}$
Total payoff	$\frac{2(4-3\xi)}{3\sqrt{3}(3-2\xi)^{\frac{3}{2}}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c^A}}$	$\frac{\sqrt{2}(7-6\xi)}{3\sqrt{3}(3-2\xi)^{\frac{3}{2}}} \frac{N w^{\frac{3}{2}}(r)^{\frac{3}{2}}}{\sqrt{c_A}}$

We observe that the cost sharing parameter plays an important role in the optimal values of the decision variable and the payoffs thereof. Table 5 tabulates how the decision variables

and the payoffs change with change in ξ.

Table 5. Impact of ξ on the decision variables and payoffs in different scenarios

			Hospital		Coordinating organization		Total Social Welfare	
ξ	Maintenance Fee	Effort	Decentralized	Cooperative	Decentralized	Cooperative	Decentralized	Cooperative
0	0.333	0.577	0.074	0.210	0.222	0.157	0.296	0.367
0.1	0.321	0.598	0.074	0.209	0.230	0.163	0.304	0.372
0.2	0.308	0.620	0.073	0.208	0.239	0.169	0.312	0.377
0.3	0.292	0.645	0.072	0.205	0.248	0.176	0.321	0.381
0.4	0.273	0.674	0.071	0.200	0.259	0.183	0.330	0.384
0.5	0.250	0.707	0.068	0.192	0.272	0.192	0.340	0.385
0.6	0.222	0.745	0.064	0.180	0.287	0.203	0.351	0.383
0.7	0.188	0.791	0.057	0.161	0.304	0.215	0.361	0.377
0.8	0.143	0.845	0.046	0.131	0.325	0.230	0.372	0.361
0.9	0.083	0.913	0.029	0.083	0.351	0.248	0.381	0.331

From Table 5 we note that:

- Maintenance Fee is decreasing in ξ while effort is increasing in ξ
- For the Hospital: payoff in cooperative is always greater than the payoff in decentralized case
- For the Coordinating organization: payoff in cooperative is always lower than the payoff in decentralized case
- For the overall system level

  If $\xi < .76$

  payoff in cooperative is greater than the payoff in decentralized system

  Else

  payoff in cooperative is lower than the payoff in decentralized system

DISCUSSION AND CONCLUSION

On the basis of our model and analysis we conclude that the CO could exploit the unauthorized channel's uncertainty of finding a matched demand to incentivize a hospital to join the authorized channel. The uncertainty could be leveraged to influence hospital's decision even if the maintenance fee offered by the CO is lower than the price offered in the unauthorized channel. This result helps us to obtain a minimum amount to be set as maintenance fee. There are several other factors that we identify in our analysis and explain how they impact the maintenance fee setting decision of the CO. The analysis could help COs to discourage the unauthorized practice by hospitals and at the same time boost the volume of cadaver organ transplants in the authorized system.

Prior studies in pursuit of finding the reasons for failure to convert potential donors have consistently pointed out the reluctance to approach families of the cadaver and unidentified donors as some of the root causes. Our study findings indirectly corroborate this factor. It is

therefore critical for COs to understand how the cost of effort of hospitals is driving the decisions related to the fee and the effort of CO and the hospital respectively. Further, it is also important to understand how COs may contribute to reduce the cost of effort of hospitals in the cadaver organ recovery.

Our findings show that it is in fact in the interest of the CO to participate in or contribute to the effort during the conversion phase. The contribution from the CO in sharing the cost of effort could come in various forms. For instance, training the coordinators in a hospital is vital as lack of trained coordinators is often regarded as a hurdle in effective persuasion of the donor family. COs could provide training to the staff thereby reducing hospital's cost.

For the hospital, while maintenance phase of the donor is demanding, in terms of resource and infrastructure requirement, the conversion phase is equally taxing. The maintenance fee reimburses them after the donor is converted but the upfront effort required to convert a donor is vital and inevitable. The CO needs to take cognizance of the importance of the conversion phase from not only the viewpoint of the direct output i.e. converting a donor but also from the viewpoint of the input in terms of effort and the indirect fallout such as emotional stress that hospital staff experiences. The analysis suggests that a high cost sharing could in fact be beneficial to the CO as it not only brings down the optimal maintenance fee but at the same time pushes up the effort of the hospital. The latter translates to a higher volume of conversions. The decentralized system always favors the CO. This is because the CO fixes the reimbursement rate keeping only its own best interest in mind.

As an extension to the existing base model, we use a cooperative scenario to see what reimbursement rate is in the best interest of the system when the fee decision is made to improve both parties' payoffs. In US, such negotiations are permissible between the organ retrieving hospital and procurement organization, which is equivalent to the CO in our case (Medicare, 2016). As expected, once the fee decision is made keeping in the interest of the hospital also in consideration, the reimbursement rate goes up. In fact, the optimal fee in a cooperative setting is exactly double of what the CO offers in a decentralized set up. The hospital always finds the cooperative scenario more beneficial, whereas the CO finds the decentralized scenario beneficial at an individual level. The gap between the optimal fee in the cooperative and the decentralized set up is likely to be a reason for discouraging active participation of hospitals in decision making by the planner.

However, a different dynamic unfolds at the total system level. The cost sharing parameter has implications which are not straight forward but at the same time it is not far-fetched either. The system level payoff is what the planner is concerned about. In the cooperative set up, the total payoff depends on the cost sharing factor and is higher than the decentralized set up till a critical value of the cost sharing parameter is achieved. Beyond the critical value of the parameter, the system starts performing poorly in comparison to the decentralized system.

The study provides insight to coordinating organizations on how they should set the reimbursement scheme for hospitals to keep them in the authorized channel. However, our study has a few limitations. For example, there are a few other components of hospital operations that we have not managed to include in our study. For instance, there are fixed costs for a hospital to register as an organ retrieval hospital such as administrative level hurdles. In fact, anecdotal evidence suggests that too much paperwork, lengthy processes, ambiguity in the steps involved in the process from CO's side discourages hospitals to come forward to register as an organ retrieval hospital (Times of India report, 2015). Additionally, there are best practices which can be adopted from high performance countries, such as Spain. For instance, based on the advice of medical experts on the need for trained intensivists, Spain, which is currently a leader in donation rate, has mandated the presence of trained intensivists in the ICUs. However, we currently do not differentiate such costs and therefore cannot capture the impact the intensivists make on conversion rate Also, we use a monetary reimbursement scheme and

acknowledge that there could be other means to achieve a similar objective. For example, monitoring of hospitals through audits and penalty on non-conformation of hospitals. Both of the listed options have been explored to a limited extent in the cadaver organ supply chain literature and provide scope for further work.

APPENDIX

Technical proofs of results available upon request.

REFERENCES

Adanir, T., Erdogan, I., Hunerli, G., Unveren, G., Dasci, H., Cetin, H. Y., ... & Aydin, U. (2014, December). The effect of psychological support for the relatives of intensive care unit patients on cadaveric organ donation rate. In *Transplantation proceedings* (Vol. 46, No. 10, pp. 3249-3252). Elsevier.

Adjerid, I., Acquisti, A., Telang, R., & Padman, R. (2016). The Impact of Privacy Regulation and Technology Incentives: The Case of Health Information Exchanges. *Julia Adler-Milstein*, *62*(4), 1042–1063.

Akan, M., Alagoz, O., Ata, B., Erenay, F. S., & Said, A. (2012). A broader view of designing the liver allocation system. *Operations research*, *60*(4), 757-770.

Altug, M. S., & Sahin, O. (2019). Impact of Parallel Imports on Pricing and Product Launch Decisions in Pharmaceutical Industry. *Production and Operations Management*, *28*(2), 258–275.

Andritsos, D. A., & Tang, C. S. (2018). Incentive Programs for Reducing Readmissions when Patient Care is Co-Produced.

Arıkan, M., Ata, B., Friedewald, J. J. and Parker, R. P. (2018), Enhancing Kidney Supply Through Geographic Sharing in the United States. *Production and Operations Management*, 27: 2103-2121

Arora, P., & Subramanian, R. (2019). Improving Societal Outcomes in the Organ Donation Value Chain. *Production and Operations Management*, *0*(0), 1–22.

Arya, A., & Mittendorf, B. (2015). Supply chain consequences of subsidies for corporate social responsibility. *Production and Operations Management*, *24*(8), 1346-1357.

Ata, B., Skaro, A., & Tayur, S. (2017). OrganJet: Overcoming Geographical Disparities in Access to Deceased Donor Kidneys in the United States. *Management Science*, *63*(9), 2776–2794.

Avcı, B., Loutfi, Z., Mihm, J., Belavina, E., & Keck, S. (2014). Comparison as incentive: Newsvendor decisions in a social context. *Production and operations management*, *23*(2), 303-313.

Bertsimas, D., Farias, V. F., & Trichakis, N. (2013). Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. Operations Research, 61(1), 73-87.

Dai, T. (2015). Incentives in U.S. Healthcare Operations. *Decision Sciences*, *46*(2), 455–463.

Dai, T., & Tayur, S. (2019). Healthcare operations management: A snapshot of emerging research. Manufacturing & Service Operations Management.

Dai, T., Zheng, R., & Sycara, K. (2019). Jumping the line, charitably: Analysis and remedy of donor-priority rule. *Management Science*.

Dasu, S., Ahmadi, R., & Carr, S. M. (2012). Gray markets, A product of demand uncertainty and excess inventory. *Production and Operations Management*, *21*(6), 1102–1113.

Guo, P., Tang, C. S., Wang, Y., & Zhao, M. (2019). The Impact of Reimbursement Policy on Social Welfare, Revisit Rate, and Waiting Time in a Public Healthcare System: Fee-for-Service Versus Bundled Payment. *Manufacturing & Service Operations Management*, *21*(1), 154–170.

Huang, J., Leng, M., Liang, L., & Liu, J. (2013). Promoting electric automobiles: Supply chain analysis under a government's subsidy incentive scheme. *IIE transactions*, *45*(8), 826-844.

Jendrisak, M. D., Hruska, K., Wagner, J., Chandler, D., & Kappel, D. (2002). Cadaveric-donor organ recovery at a hospital-independent facility. *Transplantation*, *74*(7), 978-982.

Karunakaran, C. E., & Amalorpavanathan, J. (2018). Hospital awareness rather than public awareness is key to promote organ donation. *The National medical journal of India*, *31*(4), 193.

Kessler, J. B., & Roth, A. E. (2012). Organ allocation policy and the decision to donate. *American Economic Review*, *102*(5), 2018-47.

Kessler, J. B., & Roth, A. E. (2014). Getting more organs for transplantation. *American Economic Review*, *104*(5), 425-30.

Kessler, J. B., & Roth, A. E. (2014). Loopholes undermine donation: An experiment motivated by an organ donation priority loophole in Israel. *Journal of Public Economics*, *114*, 19-28.

Kong, N., Schaefer, A. J., Hunsaker, B., & Roberts, M. S. (2010). Maximizing the efficiency of the US liver allocation system through region design. Management Science, 56(12), 2111-2122.

Luo, C., Leng, M., Huang, J., & Liang, L. (2014). Supply chain analysis under a price-discount incentive scheme for electric vehicles. *European Journal of Operational Research*, *235*(1), 329–333.

Medicare (2016). Retrieved from: https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R471pr1.pdf

Nagral, S., & Amalorpavanathan, J. (2014). Deceased organ donation in India: Where do we go from here?. *Indian journal of medical ethics*, *11*(3), 162-166.

Nathan, H. M., Conrad, S. L., Held, P. J., McCullough, K. P., Pietroski, R. E., Siminoff, L. A., & Ojo, A. O. (2003). Organ donation in the United States. *American Journal of Transplantation*, *3*, 29-40.

ORGAN India report 2013. Retrieved from https://www.organindia.org/wp-content/uploads/2014/11/ORGAN-Research-Report.pdf

Panwar, R., Pal, S., Dash, N. R., Sahni, P., Vij, A., & Misra, M. C. (2016). Why are we poor organ donors: a survey focusing on attitudes of the lay public from northern India. *Journal of clinical and experimental hepatology*, *6*(2), 81-86.

Porteous, A. H., Rammohan, S. V., & Lee, H. L. (2015). Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties. *Production and Operations Management*, *24*(9), 1402-1413.

Qian, Q., Guo, P., & Lindsey, R. (2017). Comparison of Subsidy Schemes for Reducing Waiting Times in Healthcare Systems. *Production and Operations Management*, *26*(11), 2033–2049.

Sahay, M. (2018). Transplantation of human organs and tissues Act-"Simplified". *Indian journal of transplantation*, *12*(2), 84.

Schummer, J. (2016). *Influencing waiting lists*. Technical report, Working paper, Kellogg School of Management, 2015. 41.

Shao, J., Krishnan, H., & McCormick, S. T. (2016). Gray Markets and Supply Chain Incentives. *Production and Operations Management*, *25*(11), 1807–1819.

Shao, L., Yang, J., & Zhang, M. (2017). Subsidy scheme or price discount scheme? Mass adoption of electric vehicles under different market structures. *European Journal of Operational Research*, *262*(3), 1181–1195.

Shroff, S. (2019). Twenty-five years of transplantation law in India–Progress and the way forward. *Indian Journal of Transplantation*, *13*(3), 151.

Srinivasan, S. (2013). Has Tamil Nadu turned the tide on the transplant trade?. *BMJ*, *346*, f2155.

Srivastava, A., & Mani, A. (2018). Deceased organ donation and transplantation in India: Promises and challenges. *Neurology India*, *66*(2), 316.

Su, X., & Zenios, S. A. (2006). Recipient Choice Can Address the Efficiency-Equity Trade-off in Kidney Transplantation: A Mechanism Design Model. *Management Science*, *52*(11), 1647–1660.

Suresh, Adithyan & Mariappan, M. (2017). Factors that Determine Deceased Organ Transplantation in India. Indian journal of transplantation. 11. 10.4103/ijot.ijot_13_17.

Times of India report (2017). Retrieved from: https://timesofindia.indiatimes.com/city/mumbai/private-hospitals-make-94-of-cadaver-organ-donations/articleshow/61843904.cms

Times of India report (2015). Retrieved from: https://timesofindia.indiatimes.com/city/nagpur/States-adoption-of-organ-donation-rules-welcomed/articleshow/48367303.cms

Thomas, P. G., Aswathy, C., Joshy, G., & Mathew, J. (2018). Elements of a successful hospital-based deceased donation programme in India: Zero to eighty-five in two years. *The National medical journal of India*, *31*(4), 201.

Yu, J. J., Tang, C. S., & Shen, Z. J. M. (2018). Improving consumer welfare and manufacturer profit via government subsidy programs: Subsidizing consumers or manufacturers?. *Manufacturing & Service Operations Management*, *20*(4), 752-766.

Zenios, S. A., Chertow, G. M., & Wein, L. M. (2000). Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research*, *48*(4), 549-569.

Zhang, D. J., Gurvich, I., Mieghem, J. A. Van, Park, E., Young, R. S., & Williams, M. V. (2016). Hospital Readmissions Reduction Program: An Economic and Operational Analysis. *Management Science*, *62*(11), 3351–3371.

Zhang, F., & Zhang, R. (2018). Customer Purchasing Behavior, and Government Policy. *Manufacturing & Service Operations Management*, *20*(4), 601–616.

DECISION SCIENCES INSTITUTE

The Role of Information Systems (IS) Factors and the IS Investment on the Financial Performance of a Hospital

Anjee Gorkhali
Susquehanna University
Email: gorkhali@susqu.edu

Fatou Diouf
George Mason University
Email: fdiouf@gmu.edu

Yifang Ma
Old Dominion University
Email: yma004@odu.edu

ABSTRACT

Reviewing the existing theory, the authors find a direct positive relationship between IT investment and healthcare service efficiency, which in turn generate positive financial returns. However, there is little evidence that supports expected efficiencies and quality improvements in practice. Therefore, the purpose of this study is to examine the relationship between health information technology expenditures, IS organizational factors such as IS staff, IS plans deployed by management, IS components and initiatives related to IS, hospital financial performance and productivity. This paper developed two multiple regression models to test and analyze the impact of IT investment on hospital performance. The findings show that IS staff in PC Support and Security have the most statistically significant impact on hospital performance. The result implied that to have a proper benefit from the organization strategies for achieving a performance level, staffing fit is a key factor. In addition, the authors investigated research limitations and future opportunities.

KEYWORDS: Information Systems (IS) Factors, IS Investment, Electronic Health Records (EHR), Hospital Financial Performance

INTRODUCTION

Every organization is trying to find a balance between its expenses and revenue and hospitals are not different in this pursuit. Hospitals have implemented various mechanisms to find a balance and one of the major criteria has been to increase their revenue. One such mechanism implemented is the use of information systems particularly Health Information Exchange through Electronic Health Records (EHR) systems. The Centers for Medicare & Medicaid Services (CMS) have been providing incentive payments to hospitals that can successfully demonstrate meaningful use of EHR technology. The meaningful use initiative includes several health care goals such as: improvement of quality, safety, and efficiency; the reduction of health disparities; the engagement of patients and families in care decisions; the improvement of public health; and the ascertainment of privacy and security of patient health information (HealthIT.gov, 2015, Holland, 2015). The objective of measuring meaningful use is to achieve better clinical outcomes, improved population health outcomes, increased transparency and efficiency, empowered

individuals, and more robust research data on health systems (HealthIT.gov, 2015). According to the classification provided by Holland, 2015, meaningful use initiative comprises of 3 stages; where the Stage 1 ranged from 2011-2014 mainly focused on data capture and sharing; Stage 2 started from 2015-2017, and focused on advance clinical processes; and Stage 3 starting after 2017 focused on improved outcomes. Furthermore, "Meaningful Use" also sets some specific objectives that hospitals need to achieve to qualify for the CMS incentive payments. For example: in Stage 2, hospitals must have a EHR systems that can protect electronic health information and use clinical decision support systems to improve performance (CMS, 2016). In addition to CMS Meaningful Use standards for EHR systems, HITECH also has requirements for the proper use of EHR systems. Such as, it must be interconnected and contribute to improve the quality of health care. Hence, Accounting Information Systems researchers assert that an expansion of healthcare research is required to include business process influences (Fichman et al., 2011).

In theory, there is a direct positive relationship between IS investment and healthcare service efficiency which in turn generate positive financial returns (Menon et al., 2000). However, in practice, there is little evidence that supports expected efficiencies and quality improvements. Therefore, in addition to the direct relationship between IS investment and hospital performance, there needs to be a study of supporting IS environment as well. Therefore, the purpose of this study is to examine the relationship between health information technology expenditures, IS organizational factors such as IS staff, IS plans deployed by management, IS components and initiatives related to IS, hospital financial performance and productivity. Using the fit and flexibility model proposed by Wells and Snell (1980), this paper analyzes the need to fit the IS staffing level in an organization in conjunction with IS budget to achieve the optimal hospital performance. In the next section, we review related literature and propose our hypotheses. Then, we describe our methodology and empirical methods. At last, we present our results and follow with a discussion, limitations and concluding comments.

LITERATURE REVIEW

IS investment are considerable in many industries, and there are a number of studies that studied the benefits of using information technology in different sectors and organizations (Dewett and Jones (2001), Lin and Pervan (2003), Subramani (2004), Shekelle, Morton, and Keeler (2006), Buntin, Burke, Hoaglin, and Blumenthal (2011), Ølnes, Ubacht, and Janssen (2017), Paaske, Bauer, Moser, and Seckman (2017)). However, these studies have not been able to establish the relationship between IT and an organization's performance (Devraj and Kohli, 2000), and those investments do not always guarantee suitable returns.

Nevertheless, there are some studies that were successful in establishing relationship between IS and some measures of organizational performance such as inventory management, cost reduction (Melville et al., 2004), productivity and customer service (Melián-González & Bulchand-Gidumal, 2016), as well as innovation and entrepreneurship (Yunis, El-Kassar, & Tarhini, 2017). Many studies have been focused in the field of healthcare (Williams, Asi, Raffenaud, Bagwell, and Zeini (2016), Y. Wang and Hajli (2017), Angst, Block, D'arcy, and Kelley (2017), T. Wang, Wang, and McLeod (2018)). It was shown by Kohli and Tan (2016) that the use of IS does have a positive and significant impact interaction between hospitals with their patients and other stake holders. Specifically, investigating the relationship between IS investment and hospital financial performance has attracted much interests from many scholars. For example, Li and Collier (2000) first established a significant relationship with the use of IS in hospitals. In addition, Zengul, Weech-Maldonado, Ozaydin, Patrician, and O'Connor (2018) conducted a longitudinal study to

examine the impact of high-tech medical services on financial performance of U.S. hospitals. Their findings revealed that both breadth and rareness of high-tech services were associated with high total margin among not-for-profit hospitals. There are other factors such as clinical quality and other intermediate business process that lead to the positive impact of IS on the financial performance.

These studies assume that the IS systems, in particular EHR, are interconnected, interoperable, widely adopted, and are used effectively (Grieger et al., 2007), therefore implying that health information exchange is an important factor in realizing the positive impact of IS on financial performance of a hospital. In addition, some studies such as the one conducted by Devraj and Kohli (2000) shows that IS has a limited impact on financial performance of a hospital. One notable point is that, the studies that find limited impact do not consider the lag effect of IS investment on the hospital financial performance (Lee and Kim, 2006). Even though Lee and Kim (2006) managed to establish a lag effect between IS investment and hospital performance, T. Wang et al. (2018) added that there are intermediate business processes, such as the use of EHR and the hospital bed utilization rates that impact the direct relationship. However, all these studies take IS investment as a whole and do not break it down into the multiple facets that are included in the investment. For example, IS investment includes the purchase of IS equipment and the IS support staff required to operate those EHR and other HIT related systems.

This study will present that the IS investment does not necessarily have a direct and positive impact on hospital productivity. Instead it will assert that various factors such as staffing levels in different IS functions, management plans to implement and deploy various IS functions will have a significant impact on hospital productivity. Therefore, this study will use additional factors such as management plans/initiatives, number of staff deployed in various IS related departments and number of IS components and systems currently in use to analyze the impact of IS investment on hospital productivity. To sum up, this study will analyze the importance of staff-strategy/plan fit to achieve the target hospital productivity.

THEORETICAL FRAMEWORK AND HYPOTHESIS

This study utilizes the fit and flexibility model proposed by Wright & Snell, 1998, which helps in developing an organization's strategic formulation. This model presents a framework to provide a guideline for the successful implementation of an organization's mission and goals, examination of the internal strengths and weaknesses, and identification of an organization's external opportunities and threats. This model presents two possible methods to achieve strategic fit for an organization and the first methods utilizes the use of HR practices. In our study we will implement the first method to develop our hypothesis. This method matches the skills and behaviors of a staff with the organization's strategic goals. The greater the match or congruence, the more likely it is that the firm will realize higher productivity. This method is a three-step process where the HR will first convert the strategies into required skills and behaviors. In the second stage, it will list the types of HR practices needed and then finally design, develop and deliver those practices. Extant literature has classified fit into three types: person- job match, person-organization match, and strategy-staffing match. In our study we will focus on the third type of fit which is the strategy-staffing match.

The organizations that practice stable and being operated in predictable environments will benefit from the strategy-staffing fit (Wright and Snell, 1998) of staffing practices to organizational strategy. This study focuses on IS staff only and does not consider other medical staff of a

hospital. Even though, hospitals mostly have a unstable and unpredictable environment in terms of required staffing level, this is mainly for the medical staff and not the IS staff. In fact, for IS staff the working condition is mostly considered to be stable and predictable. Hence, the Wright and Snell's model can be used to explain the staffing level strategy for IS staff in a hospital. Furthermore, the authors claim that one of the major responsibilities of the staffing fit is to be able to develop a proper understanding of the organization's current and future strategy and objectives as well as an organization's practices and policies. However, it is also evident that the assumptions of the staffing fit such as environmental stability, identifiable skills, accurate forecasts will be impacted by economic shifts, technological changes, globalization, and other workforce trends (Cascio & Aguinas, 2008; Howard, 1995). In short, this staffing-strategy fit will enable the decision makers to identify the skills and behaviors required for the implementation of the organization's strategies by providing an estimate of the quality (required skills, behaviors) and quantity (number of positions) of human resources required. Similarly, this study will extend the staffing-strategy fit to other resources in an organization, such as IS components and IS budget available for the use of the staff. IS components are also organization resource, thus this study hypothesizes that a fit between IS component and organization's strategies is needed to achieve the organization's performance goal.

Therefore, it can be implied that in order to get a proper benefit from the organization strategies to achieve a performance level, staffing fit is a key factor. In a hospital since most of the concern is about the medical staffing, the IT staff gets neglected and hence, the performance return from the implementation of health information system is not necessarily realized. This study hypothesizes that, in terms of IT staff, there are some distinction that have positive impact on the hospital productivity compared to other staff. Hence, this study will help in identifying those key IT staff (as shown in Fig. 1) and redesign the staffing level of a hospital to meet the fit between staffing level and the strategies to achieve desirable hospital productivity.

Hypothesis 1: The level of IS Budget has a positive impact on the productivity of the hospital.

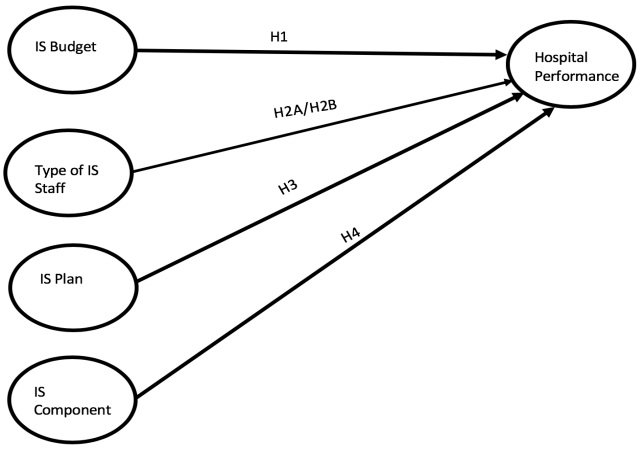
Hypothesis 2A: The level of IS staffing for each group of IS staff will have different impact on hospital productivity.

Hypothesis 2B: The level of IS staffing has a positive impact on the productivity of the hospital.

Hypothesis 3: The level of IS plans/initiatives use has a positive impact on the productivity of the hospital.

Hypothesis 4: The level of IS components use has a positive impact on the productivity of the hospital.

Fig. 1: Conceptual Model Framework for Hypothesis Testing.



METHODOLOGY AND MODEL

The data was collected from 2014 HMSS database. The study is based on 60 hospitals from 27 states (Fig. 2). There are 14 independent variables and one dependent variable (Productivity) in this study. The variables are defined in Table 1. The size of the hospital in terms of overall hospital bed and the location of hospital are control variables.

Fig. 2: Number of Hospitals per State

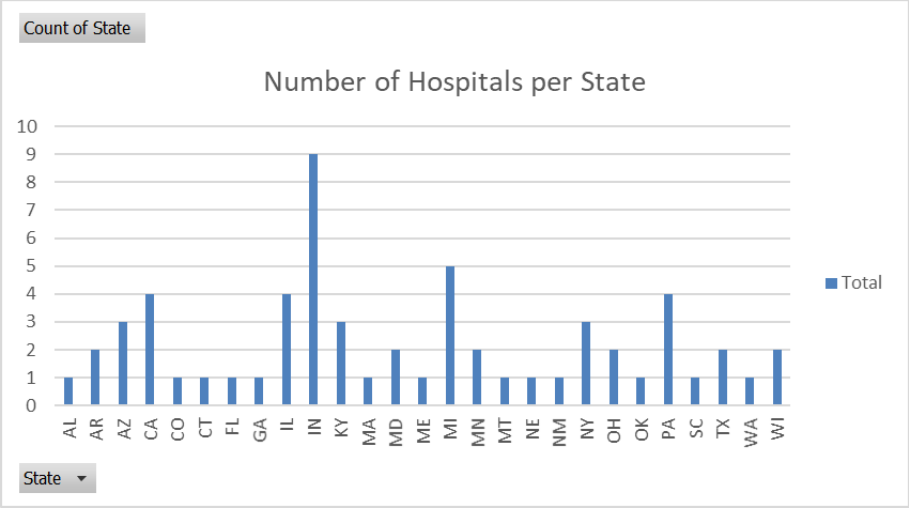


Table 1. Definitions of the Dependent and Independent Variables

Variable	Definition
Productivity (HP)	Net Revenue per million over the number of staffed beds [1].
IS budget (ISBudget)	The total amount budgeted by the IS department at the healthcare systems for the

	current fiscal year end. This amount is a percentage of all operating expenses.
FTE Project Management	Number of IS FTEs that are Project Managers.
FTE Management	Number of IS FTEs in Management
FTE Programmer	Number of IS FTEs that are Programmers
FTE Operations	Number of IS FTEs working in operations
FTE Network Administrator	Number of IS FTEs that are network administrators
FTE HelpDesk	Number of IS FTEs at the Help Desk
FTE Security (FTESecur)	Number of IS FTEs in IS Security
FTE EMR Support (FTEEMRSu)	Number of IS FTEs that support EMR applications
FTE PC Support (FTEPCSup)	Number of IS FTEs providing PC support
Number of IT components "(NoITComp)	Total number of IS devices in full operation
ISplan (ISPlan)	Total number of IS Strategic plans implemented
ISstaff (ISStaff)	Total number of IS FTEs including those categorized "Other IS FTEs"

To test the hypothesis, two models were developed: the first model only included IS budget, IS Plans, IS staff and Number of IT components as shown in equation 1 below. IS staff was calculated by adding all FTEs described in Table1 such as FTE operations, FTE management. All of the FTE's are different sub-groups of IS staff in the hospital. The next model included, all of the above variables except for the IS staff, as it was replaced with each individual FTE's in its original form as shown in equation 2. This helped us test our hypothesis, that if the strategy makers have information about the level of IS staff only then it is not as important and insightful as having insight on the level of different sub-group of IS staff in the hospital. The models are created in SPSS using the simultaneous multiple regression method for Model 1, in which all variables are entered and kept in the model regardless of their significance. Model 2 is created from a backward selection method in which all variables are first entered and then, removed one at a time according to their significance; the final model selection was based on the R-square adjusted.

$$HP = \beta 0 + \beta 1 * ISBudget + \beta 2 * ISStaff + \beta 3 * NoITComp + \beta 4 * ISPlan + \varepsilon \qquad (Model\ 1)$$

$$HP = \beta 0 + \beta 1 * ISBudget + \beta 2 * FTEPCSup + \beta 3 * FTESecur + \beta 4 * FTEEMRSu + \varepsilon \qquad (Model\ 2)$$

The adjusted R^2 for these models are listed in Table 2. The regression result and statistical significance are shown in table 3. According to the results shown in Table 3, IS budget, FTE Programmer and FTE PC Support have a significant linear relationship with Hospital Productivity. While the different IS FTEs components have significant linear relationships with Hospital Productivity, the number of IS plan does not have any significant linear relationship with the dependent variable.

RESULTS

Table 2. R^2 and Adjusted R^2

Model	R Square	Adjusted R Square
1	.133	.07
2	0.276	.224

Table 3. Regression Results and Hypothesis Testing

Model	Path	Significance	Path	Significance	Path	Significance
1	ISBudget→HP	.013(**) H1(+)	ISStaff→HP	0.993 (not sig) H2B(-)	NoITComp →HP	.702 (not sig) H4(-)
	ISPlan→HP	0.439(not sig) H3(-)				
2	ISBudget→HP	.010 (**) H1(+)	FTEPCSup →HP	.002(**) H2A(+)	FTESecur →HP	0.039(**) H2A(+)
	FTEEMRSu →HP	0.055(*) H2A(+)				

DISCUSSION

The results show that, the first model where the IS staff is not divided into different categories does not provide a proper analysis as all, but the IS budget is statistically insignificant. Furthermore, in that model, only 7% of the variations in Productivity are explained by the predictors variations, which suggested that a strategy developed based on the total number of staff does not provide an optimal solution. It has been shown that when the staff are divided into their subsequent sub-groups then it provides a better understanding of the model. 22.4% of the variations in Productivity are explained by the variations in the predictors hence, giving the decision makers clearer picture of how many staff to hire and the qualities, qualification and experience those staff need to possess. As shown, in our study not all sub-groups of the IS staff have statistical significance towards the hospital productivity. In our study IS staff in PC Support and Security have the most statistically significant impact on hospital productivity. The decision makers can therefore identify a proper staffing-strategy fit to identify the skills and behaviors required for the implementation of the hospital's strategies by providing an estimate of the quality (required skills, behaviors) and quantity (number of positions) of human resources required. It can then be implied that to have a proper benefit from the organization strategies for achieving a performance level, staffing fit is a key factor. The notion of having a fit between other hospital resources such as IS components were found to be statistically insignificant. Therefore, supporting our claim that the fit of IS staff and strategy is one of the most important factors impacting the hospital productivity.

The organization fit-strategy model developed by Wright and Snell 1998 is found to be compatible with hospitals in our study and we were also able to extend the fit-strategy model to distinguish

between different groups of staff. Hence, it can be seen that not all staff are needed in the same number and not all staff have similar contribution to the productivity of the hospital. This is particularly important in the case of IS staff. Therefore, in future studies it will give additional insight if the IS staff analysis is conducted by grouping different types of IS staff into different groups.

RESEARCH LIMITATIONS

This study includes 60 hospitals from 27 states in the U.S, hence does not include all the states in the U.S. In addition to that 60 hospitals in itself is not a big enough sample size if we were to have a more impactful research generalization. Furthermore, the samples are from 2014 after which HIMSS has reported 5 more years of data set. Hence, this study only captures the older version of the hospital data, if it were to capture more latest data, the results might have been more insightful since the health information exchange and health information technology were fairly new during the 2014 data reporting.

Furthermore, this study does not use any control variable in its model. In future, the regression model will be tested including control variables such as location and size of the hospital. The authors understand that, a larger hospital will definitely have a much larger budget. This can directly impact its IS investment and number of IS staff it can hire as well as the quality of IS staff. Highly trained IS staff will result in better hospital performance and productivity (Gorkhali, 2019).

REFERENCES

Angst, C. M., Block, E. S., D'arcy, J., & Kelley, K. (2017). When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches (January 24, 2016). Angst, CM, Block, ES, D'Arcy, J., and Kelley, K*, 893-916.

Brueckner, G., Howard, R., Koomen, M., Korendyke, C., Michels, D., Moses, J., . . . Llebaria, A. (1995). The large angle spectroscopic coronagraph (LASCO). In *The SOHO Mission* (pp. 357-402): Springer.

Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. (2011). The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health affairs, 30*(3), 464-471.

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology, 93*(5), 1062.

Centers for Medicare & Medicaid Services (CMS), 2016. 2016 Program Requirements. Retrieved on 10/28/2016 from. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/2016ProgramRequirements.html

Devaraj, S., Kohli, R., 2000. Information technology payoff in the health-care industry: a longitudinal study. J. Manag. Inf. Syst. 16 (4), 41–67.

Dewett, T., & Jones, G. R. (2001). The role of information technology in the organization: a review, model, and assessment. *Journal of management, 27*(3), 313-346.

Fichman, R., Kohli, R., Krishnan, R., 2011. Editorial overview – the role of information systems in healthcare: current research and future trends. Inf. Syst. Res. 22 (3), 419–428.

Gorkhali, A. (2019). Early Information Access to Alleviate Emergency Department Congestion.

Grieger, D.L., Cohen, S.H., Krusch, D.A., 2007. A pilot study to document the return on investment for implementing an ambulatory electronic health record at an academic medical center. J. Am. Coll. Surg. 205 (1), 89–96.

HealthIT.gov, 2015. Meaningful Use Definition & Objectives. Retrieved on 11/9/2017 from. https://www.healthit.gov/providers-professionals/meaningful-use- definition-objectives.

Holland, E., 2015. CMS Medicare and Medicaid Electronic Health Record (EHR) Incentive Programs Final Rule Overview. Retrieved on 11/9/2017 from. https://www.cms.gov/eHealth/downloads/Webinar_eHealth_October8_FinalRule.pdf.

Kohli, R., Tan, S.S., 2016. Electronic health records: how can IS researchers contribute to transforming healthcare? MIS Q. 40 (3), 553–573.

Lee, S., Kim, S.H., 2006. A lag effect of IT investment on firm performance. Inf. Resour. Manag. J. 19 (1), 43–69.

Li, L., Collier, D., 2000. The role of technology and quality on hospital financial performance: an exploratory analysis. Int. J. Serv. Ind. Manag. 11 (3), 202–224.

Lin, C., & Pervan, G. (2003). The practice of IS/IT benefits management in large Australian organizations. *Information & Management, 41*(1), 13-24.

Melián-González, S., & Bulchand-Gidumal, J. (2016). A model that connects information technology and hotel performance. *Tourism Management, 53*, 30-37.

Melville, N., Kraemer, K., Gurbaxani, V., 2004. Review: information technology and organizational performance: an integrative model of IT business value. MIS Q. 28 (2), 283–322.

Menon, N.M., Byungtae, L., Eldenburg, L., 2000. Productivity of information systems in the healthcare industry. MIS Q. 11 (1), 83–92.

Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. In: Elsevier.

Paaske, S., Bauer, A., Moser, T., & Seckman, C. (2017). The benefits and barriers to RFID technology in healthcare. *On-Line Journal of Nursing Informatics, 21*(2).

Shekelle, P. G., Morton, S. C., & Keeler, E. B. (2006). Costs and benefits of health information technology. *Evid Rep Technol Assess (Full Rep), 132*(April), 1-71.

Subramani, M. (2004). How do suppliers benefit from information technology use in supply chain relationships? *MIS quarterly*, 45-73.

Wang, T., Wang, Y., & McLeod, A. (2018). Do health information technology investments impact hospital financial performance and productivity? *International Journal of Accounting Information Systems, 28*, 1-13.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research, 70*, 287-299.

Williams, C., Asi, Y., Raffenaud, A., Bagwell, M., & Zeini, I. (2016). The effect of information technology on hospital performance. *Health care management science, 19*(4), 338-346.

Wright, P. M., & Snell, S. A. (1998). Toward a unifying framework for exploring fit and flexibility in strategic human resource management. *Academy of management review, 23*(4), 756-772.

Yunis, M., El-Kassar, A.-N., & Tarhini, A. (2017). Impact of ICT-based innovations on organizational performance. *Journal of Enterprise Information Management*.

Zengul, F. D., Weech-Maldonado, R., Ozaydin, B., Patrician, P. A., & O'Connor, S. J. (2018). Longitudinal analysis of high-technology medical services and hospital financial performance. *Health care management review, 43*(1), 2-11.

DECISION SCIENCES INSTITUTE

Implementing a predictive model equipment failure based on currently available data : case of biomass power plants

Yuta Inoue
Tokyo University of Science
Email: 7419502@ed.tus.ac.jp

Tomoki Oshima
Tokyo University of Science
Email: 7419505@ed.tus.ac.jp

Aya Ishigaki
Tokyo University of Science
Email: ishigaki@rs.tus.ac.jp

ABSTRACT

In this study, we deal with the inventory management problems of the pump used for material transfer in the biomass plant. The plants need to have spare parts in stock to be able to take immediate action when pump has a trouble. However, the spare parts are very expensive and precise, so the plant tries to manage the spare parts with "Just-in-Time inventory" strategy. This study proposes a method of inference the failure probability of the pump from the operation data of the past equipment and the blending ratio of the raw materials, and managing the inventory of the pump.

KEYWORDS: Inventory management, Just-in-Time, Dynamic predictive maintenance, Bayesian inference, Biomass blending

INTRODUCTION

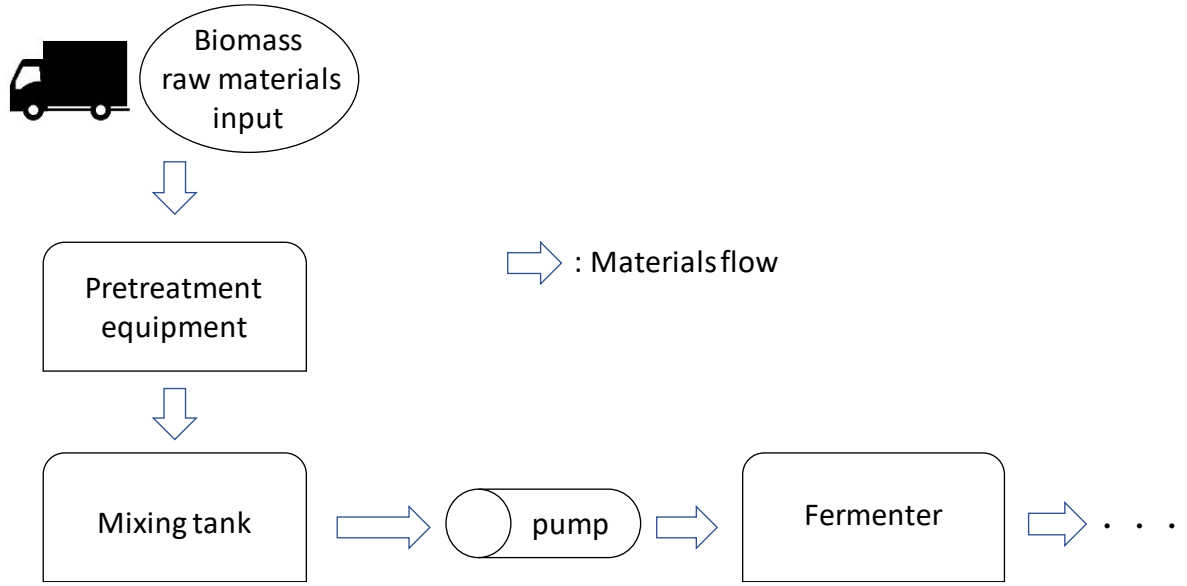
Many equipment is used in factories, schools, hospitals and public buildings. They are used based on the intended application. In recent years, by utilizing the Internet of Things (IoT) for industrial equipment, it has become possible to accumulate and analyze operational data of equipment and facilities in factories. In addition, an intelligent system is being developed that can visualize accumulated data and make effective use of it. (NEDO white paper, 2014).
As the system becomes more intelligent, the machine turns into a skilled worker and moves autonomously according to the situation, enabling productivity and quality improvement and Mass Customization. Therefore, expectations for the realization of smart factories are increasing (Sakakibara, 2017). In such a situation, manufacturing companies are considering large-scale investments for the development of IoT environments, while trying to solve the problems by making effective use of available data. Einabadi, Badoli and Ebrahimi (2019) proposed a methodology for dynamic predictive maintenance for a real case in automotive industries.
The prediction of the remaining service life (RUL) of machine parts is made by artificial neural network method considering sensor data. Lee et al. (2019) developed Artificial Intelligence (AI)-based algorithms for predictive maintenance. These algorithms are applied to monitor two important machine tool system elements: the cutting tool and the spindle motors. However, in an environment where each factor influences each other, and when there is a time lag in

response, a simple noise removal cannot detect a failure and the number of alerts cannot be optimized.
This study deals with the case of replacing parts of a pump used for transferring raw materials in a biomass power plant. We propose a method to predict the failure by correcting the failure probability of the pump from the latest data. This paper describes the procedure for predicting a failure by expressing the relationship between the frequency data for controlling the flow rate and the input material in a time series model in a situation where the inside of the pump cannot be visually inspected.

ANALYSIS TARGET

Figure 1 shows a conceptual diagram of the biomass plant in this study. In this case, biomass raw materials are charged every day, and the charged raw materials are sent to the mixing tank and fermenter. Finally, syngas is produced in the fermenter. The pump targeted in this study is the equipment shown as" pump" in Figure1.

Figure 1: Conceptual diagram of model



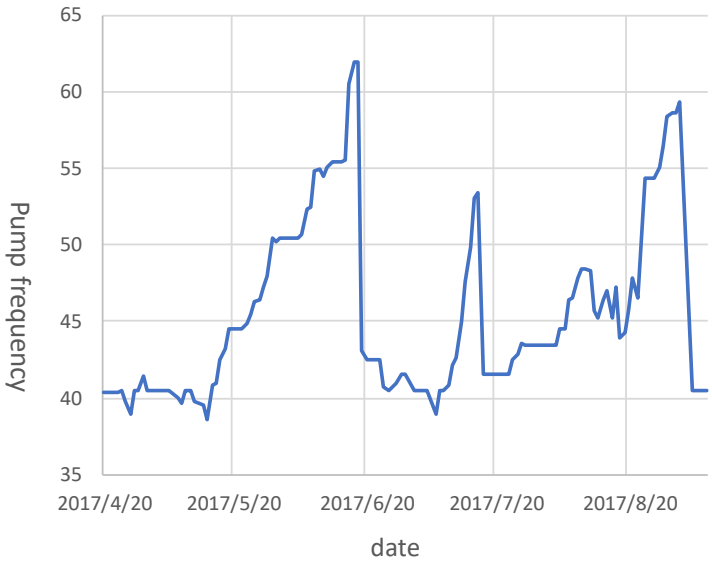
PUMP FAILURE DETECTION

Figure 2 shows the change over time of the pump frequency in the case treated in this study. The frequency of the pump fluctuates between approximately 40 Hz and 60 Hz, and facility managers try to keep the flow rate in the device constant by adjusting the frequency every day. As a general failure detection method, a method of setting a threshold value based on statistics such as an average value and a variance value and comparing the threshold value with an observed value to judge an abnormality is adopted. However, several types of raw materials were mixed and pumped. In addition, the mixing ratio of raw materials and the types of solids and liquids varied daily. Therefore, damage to the pump cross section varies from day to day. It also contained different combinations and quantities of raw materials, and there was little data to analyze them individually. In addition, since the inside of the pump cannot be visually confirmed, the facility manager not only changes the frequency according to the flow rate, but

also predicts the operating condition inside the pump while observing the response of the operation data to the change in frequency.
Since the raw material adheres to the inside of the pump, the diameter of the rotor gradually decreases, and also amount of discharge decreases. Therefore, the facility managers attempt to maintain the flow rate by increasing the frequency, but if the flow rate cannot be maintained, they determine that a failure has occurred and the pump parts are replaced.

So, high frequencies can be considered failure signals, and simple threshold settings do not cause problems such as excessive maintenance or failure detection. Here, we propose a method for equipment managers to predict pump operating conditions based on operating data, frequency trends, and differences in raw material and raw material distribution.

Figure 2: Changes in pump frequency (Excerpt)



PUMP FAILURE PREDICTION SYSTEM

The characteristics and blending ratio of the biomass feedstock not only affect the damage to the pump, but also the power generation. And, blending ratio of raw materials with less damage to pump does not always improve power generation. Therefore, it is necessary to optimize the blending ratio of the raw materials in consideration of these trade-offs. In this study, we first sought a method to determine the blending ratio of raw materials that would reduce damage to the pump.
In this study, the moving average method was used to take into account the time lag of damage to the pump caused by the raw material pretreatment process. The amount of raw materials through the pump is calculated as following equations:

$$I_t = \sum_{i=1}^{I} b_{it} \tag{1}$$

$$MI_t = \frac{I_t + I_{t-1} + \cdots + I_{t-(m-1)}}{m} \tag{2}$$

where $\boldsymbol{b_{it}}$ is weight of raw material *i* (*i* = 1,..., *I*) brought into the plant on date *t* (*t* = 1,..., *T*), $\boldsymbol{I_t}$ is cumulate amount of raw material brought into the plant on date *t,* $\boldsymbol{MI_t}$ is moving average at *m* on date *t.*
The flow from pump state monitoring to failure prediction and maintenance is as follows:

(*i*) Acquire raw material import data and equipment operation data
(*ii*) Data cleansing (ex. outliers, missing values)
(*iii*) Data analysis, rule-based prediction
(*iv*) Replacement of parts

In step (*i*), new data is read from daily reports and the database is updated every day. In step (*ii*), missing values and abnormal values due to equipment stoppage or inspection are detected and corrected. Smoothing is performed by applying a moving average method. In step (*iii*), the relationship between the frequency and the input raw material is analyzed, and rules are created. However, since the frequency is artificially reset by replacing parts of the pump, it is difficult to observe the relationship with the frequency increase as it is. Therefore, in this study, we use frequency accumulation to simplify the frequency change in the increasing direction and analyze the relationship with the input raw material. In step (*iv*), the failure occurrence probability at the current stage is predicted by Bayesian estimation. Eventually, the failure occurrence probability and the frequency increase prediction based on the input raw material in step (*iii*) are combined to predict the replacement time of parts. This paper mainly introduces data analysis performed in step (*iii*).

Pump frequency analysis

11 kinds of raw materials (expressed as A to K) are put into the equipment every day, mixed in the tank, and sent to the next equipment through the pump. The main raw materials are A, B and D, accounting for over 98% of the total. On the other hand, each of E to K has a different injection timing and number of times, and their influence on the frequency trend is also limited. Therefore, the results of investigating the effects of the main raw materials (A, B, and D) on the frequency will be shown first. The first period and the period when J and K were input are treated separately as the second period. The table below shows the breakdown of raw materials by period in this case. Table 1 shows the raw material breakdown for 275 days for each period. Due to confidentiality, the unit of raw material input cannot be disclosed, but the unit is unified. Since C was not used during these periods, the following discussion will be made except for raw material C. Table 2,3 shows the relationship between the cumulative input amount of the main raw materials and the cumulative frequency in each period. As described above, since the frequency once drops due to the replacement of the pump parts, and this time we want to analyze the tendency to increase in frequency, we decided to analyze using the cumulative frequency. Further, regarding the input amount of the raw materials, the procured amount of the raw materials is not stable due to the nature of the plant, and the input amounts and the mixing ratios of the raw materials are different every day. Furthermore, since the content of impurities and the like also differ depending on the source of the raw material, a more detailed analysis of the influence of the raw material is originally required. Therefore, first, the overall tendency was analyzed using the cumulative input amount of each raw material, and the impact of the raw material type on the frequency was investigated. When compared at the same cumulative frequency, the larger the rotor diameter, the more materials can be processed, and thus the larger the cumulative input amount.

Table 1: Breakdown of raw materials for each period		
materials	Input amount	
	1st period	2nd period
A	10477.2	22975.1
B	49942.0	48279.5
C	0.0	0.0
D	40790.3	65418.4
E	5033.7	0.0
F	329.3	23.5
G	35.3	0.0
H	776.4	0.0
I	51.7	0.0
J	2.0	86.2
K	0.0	567.1
total	107437.9	137349.7

Table 2: Relationship between cumulative frequency and cumulative input (1st period)			
Cumulative frequency	Materials		
	A	B	D
. . .	. . .	. . .	. . .
2993.8	3053.12	12442.98	6306.96
. . .	. . .	. . .	. . .
6019.2	6727.34	26180.84	21757.92

Table 3: Relationship between cumulative frequency and cumulative input (2nd period)			
Cumulative frequency	Materials		
	A	B	D
. . .	. . .	. . .	. . .
3001.5	7271.19	13757.12	19565.28
. . .	. . .	. . .	. . .
5984.6	12713.45	26276.41	33336.95

As a result of comparing the first period and the second period when cumulative frequency is about 6000.0, it was found that the influence on the frequency of the input amount of the main raw material B did not change. This is because the main raw material B is a solid substance and has a higher viscosity than other raw materials, so it was not affected by other raw materials and the damage to the inside of the pipe did not change. On the other hand, the amount of main raw materials A and D (liquid) supplied to the pump in the second period was larger than that in the first period. It can be seen that the increase in the ratio of the main raw materials A and D and the decrease in the ratio of the main raw material B suppress the damage to the pump, and as a result, maintain the input amount of the main raw material B.

CONCLUSION

In this study, we dealt with the case of replacing parts of the pump and created model for estimating invisible pump internal conditions by associating frequency data for controlling the flow rate with input raw materials. If the flow rate is constant, the system was evaluated by showing the relationship between the input amount of the raw material and the frequency of the pump. In addition, a method of predicting a pump failure from the operation state of the pump based on the influence of each raw material on the frequency was used. Since the ratio of the main raw material is high, it is necessary to remove the influence of the main raw material and evaluate the input amount of other raw materials. In addition, the input amount of each raw material was evaluated as an explanatory variable from the viewpoint of frequency control of the pump, but the correlation between the raw materials and restrictions on the raw materials were not considered. Future studies should take account for these constraints and relationships.

REFERENCES

Einabadi, B., Badoli, A. & Ebrahimi, M. (2019). Dynamic Predictive Maintenance in industry 4.0 based on real time information: Case study in automotive industries. *IFAC-PapersOnLine*, 52(13), 1069–1074.

NEDO. (2014). 2014 White Paper on Robotization of Industry, Business and Our Life. Retrieved 2020. 1.13 from https://www.nedo.go.jp/content/100563893.pdf.

Sakakibara, S. (2017). Factory Automation with Intelligent Robot Augmented by IoT and AI.*Systems, Control and Information*, 61(3), 101–106 (in Japanese).

Lee, W. J., Wu, H., Yun, H., Kim, H., Jun, M. B. G. & Sutherland, J. W. (2019). Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data, *Procedia CIRP*, 80, 506–511.

DECISION SCIENCES INSTITUTE

Extrinsic and Intrinsic Motivations in Reward-based Crowdfunding: Comparing Potential and Actual Backers

Trang Nguyen
University of Economics and Law
Email: trangnth@uel.edu.vn

Dinh Hieu Truong
National Cheng Kung University
The University of Danang – University of Economics
Email: hieutd@due.udn.vn

ABSTRACT

Based on cognitive evaluation and construal level theory, this study investigates the impacts of both motivation patterns (i.e., intrinsic and extrinsic) on funding intention, simultaneously, compare these impacts in two distinct groups of backers (actual and potential). This study also explores the moderating role of extrinsic motivation in the effect of intrinsic motivation on funding intention. The online survey will be distributed to crowd funders on crowdfunding platforms' fan pages or social network groups of crowd funders. This study is expected to contribute to crowdfunding literature by showing how intrinsic and extrinsic motivations differently influence funding intention in reward-based crowdfunding.

KEYWORDS: Crowdfunding, Extrinsic motivation, Intrinsic motivation, Potential backer, Actual backer

INTRODUCTION

In recent years, reward-based crowdfunding has increasingly attracted the special attention of businesses, which is considered as the firm's innovative mode to achieve the targeted fundraising through social network sites for its manufacturing and promotional activities. The recent development of Internet technology has led to the growing popularity of crowdfunding which facilitates project owners to raise sufficient financial resources (Fleming & Sorenson, 2016; Xiao & Yue, 2018). Among the diverse forms of crowdfunding, reward-based crowdfunding has attracted great attention from both creators and backers because of its capability to raise funds through various types, including donations, or rewards exchanging (Belleflamme, Lambert, & Schwienbacher, 2014). For instance, Kickstarter, the world's largest e-crowdfunding platform, has raised $3.5 billion and made funds for 139, 169 projects in the nine years from its launch in 2009 to January 2018.

The incentives behind backers' pledging and sharing behaviors are not entirely perceived despite the widespread of reward-based crowdfunding platforms. Specifically, previous studies revealed that rewards could be collected via various types of funder's motivation such as the prosocial motivation of pure altruism and impure altruism (for instance, warm glow, image concerns), and the economic motivation (Berglin & Strandberg, 2013; Bretschneider & Leimeister, 2017; Cholakova & Clarysse, 2015; Gerber & Hui, 2013). Nevertheless, the role of each individual type of motivation driving funders is still undiscovered in the context of reward-based crowdfunding. In other words, previous studies only examined the mixed roles of

motivation towards crowdfunding. Besides, there is a conflict existing among researchers, while some suggest prosocial motivation is very powerful (Berglin & Strandberg, 2013; Bretschneider & Leimeister, 2017), others conclude that economic incentives play a dominant role. Hence, it is necessary to conduct comprehensive research to provide a full understanding of motivations driving funder's intention as well as the distinct role of each form in the consideration of both potential and actual funders.

The philanthropic economic literature typically indicates that donors' joint fundraising efforts are motivated by three-factor combinations: intrinsic, extrinsic, and image enhancement motivations (Ariely, Bracha, & Meier, 2009). The purely internal satisfactions that arose from the act of the donation represent intrinsic motivations, such as altruism, warm-glow, or other types of prosocial preferences (Andreoni, 1989; Harbaugh, 1998). The desire to achieve material rewards or other privileges, such as CDs, t-shirts, or event tickets represents extrinsic motivations; while concerns relating to people's expectations and the need to be valued and accepted represents image enhancement. Nevertheless, the question should be considered by event creators and owners is: whether extrinsic motivations possess hidden costs or unpredictable effects? For both event managers and researchers in the areas of information systems (IS), this question is a critical issue. In specific, Zwass (2010) calls for research examining the unintentional impact of extrinsic motivations.

In this research context, the focus is on providing advanced understandings of factors motivating backers by drawing on cognitive evaluation theory. More specifically, this theory emphasized that extrinsic motivations might lead to positive or negative changes in intrinsic motivations to some certain extends (Liberman & Trope, 2014). Accordingly, the current research clarifies the roles of undiscovered dimensions of extrinsic motivations.
Besides, by relying on construal level theory (CLT), this study also aims to examine the differences in the effects of extrinsic motivation, as well as intrinsic motivation on funding intention between two distinct groups - actual and potential backers based on their direct experience and psychological distance from the reward-based crowdfunding event. Following the CLT theory, psychological distance is proposed, then, it plays an important role in the way how individuals think about and interpret a target event or activity (in this case, crowdfunding). The principle of CLT theory presents that an individual's distance to an object has a positive impact on their construal level (Trope & Liberman, 2010). Hence, the current research aims to examine the differences between potential and actual backers in construing a target reward-based crowdfunding event based on their " egocentric" perception of how large the distance is to an event or activity regarding time/temporal, space, social distance, and hypothetical alternatives to reality.

Indeed, it is important to distinguish between actual and potential groups of members and understand the differences in their perception and experience on their target activities (Kankanhalli, 2015). Taking the research's results as a salient example, due to the lack of direct experiences, the intention of potential users to innovate on mobile data services is not influenced by those factors which are only recognized by actual users who experience the innovation's activity directly. Accordingly, the present paper fills the gap that there is a lack of previous studies examining the differences between these two groups of backers on funding intention in the context of reward-based crowdfunding up to now.

LITERATURE REVIEW

In the domain of organizational behavior, Gagné and Deci (2005) conceptualized motivation as the psychological impacts on individuals' actions/ behaviors. Specifically, in crowdfunding contexts, self-determination theory is identified as the foundation of motivation research, from which two motivation types are established: intrinsically-oriented and extrinsically-oriented (Ryan & Deci, 2000a). According to incentive theory, extrinsic rewards including financial benefits and aspiration for reinforcement are critical motivating factors of behaviors (Steers, Mowday, & Shapiro, 2004). Nevertheless, individuals who appreciate both self-interest and prosocial values are not regularly affected by extrinsic motivations (De Dreu & Nauta, 2009). This implies that intrinsic motivations (i.e. personal concerns/ values, enjoyment, and self-development) can control behaviors regardless of whether extrinsic outcomes exist or not (Deci & Ryan, 1985; Ke & Zhang, 2009; Ryan & Deci, 2000a, 2000b).

It is undeniable that there are numerous motives driving people to become funders, which may be reaping rewards, broadening social networks, or supporting project founders (Agrawal, Catalini, & Goldfarb, 2014; Gerber, Hui, & Kuo, 2012). However, in incentive-based crowdfunding literature, Bretschneider and Leimeister (2017)'s study demonstrated that backers are inclined to have egoistic motivations instead of prosocial ones. Particularly, they mainly give priority to tangible or monetary rewards (Gerber et al., 2012; Gleasure & Feller, 2016). On the other hand, a growing number of researchers (e.g., Agrawal et al., 2014; Allison, Davis, Short, & Webb, 2015; Cholakova & Clarysse, 2015) have asserted that altruistic motivations urge investors to assist creators and support their ideas, even when extrinsic rewards are not assured. With funders' contributions in terms of charitable donations and other limited amounts of investment, they are still promised to be rewarded deservingly, which is the unique characteristic of reward-based crowdfunding. To provide an in-depth understanding of crowdfunding literature, both intrinsic and extrinsic motivations are simultaneously considered as driving forces of backers in our research.

HYPOTHESES/MODEL

Intrinsic Motivation for Crowdfunding

Altruism Motivation

Altruism refers to the willingness to help others. This perception encourages individuals to fund since they enjoy supporting others. The research findings (Moysidou & Spaeth, 2016) indicated that personal connections between fundraisers and investors lead to mutual beliefs and interests which encourage funding participation. In addition, those personal connections are supposed to link with compatibility in social background, cultural identities, religious beliefs, and so on (Lin & Viswanathan, 2015; Moysidou & Spaeth, 2016). A personal connection is considered to be a source of funding motivation since it helps the backers and creators invest in suitable campaigns in a positive way (Moysidou & Spaeth, 2016).

Based on those above arguments, this research adopts the perception that altruism motivation is ‘‘helping someone at some cost to oneself" or the reverse of selfishness (Ozinga, 1999). Al-Jabri and Al-Khaldi (1997) state that one of the causes of altruism motivation is a personal

connection. In this research context, the altruistic motive could lead to crowdfunding intention. Hence, we propose:
Hypothesis 1: In the reward-based crowdfunding context, backers' altruism motivation is positively associated with funding intention.

Liking Motivation

Liking refers to individuals' affections to a person or an object. In the marketing field, Unger, Johnson, and Rohrbach (1995) explore consumers' affection for advertisements by utilizing this concept. In IS researches, liking is exploited to measure users' attitudes and adoption of computers as well as the intention to use or actual behavior (Unger et al., 1995). From this basis, to study on user attitude and behavior, the researchers in IS discipline are more likely to use liking as a variable. Sabherwal, Jeyaraj, and Chowa (2006) indicate that liking plays a fundamental role in the success of information systems (IS). Liking is considered as an attitude's sub-construct which works as related influences in investigating user's behavior on online shopping. For instance, Pavlou and Fygenson (2006) reveal that the liking among users in online shopping platforms significantly influences them to shop on these platforms.
Besides, liking also holds an important role in the field of start-ups and angel investing. Brettel (2003), Feeney, Haines Jr, and Riding (1999) and Mason and Stark (2004) suggest that investors tend to seek a promising investment based on the linkage between a likable impression and a certain personal appeal toward entrepreneurs and their ventures at the first step.

In line with the above, we indicate that liking could be a critical influencing factor of funding intention in the context of reward-based crowdfunding. We believe that backers may invest in a crowdfunding project due to their impression of the project. Hence, we propose*:*
Hypothesis 2: In the reward-based crowdfunding context, backers' liking motivation is positively associated with funding intention.

Extrinsic Motivation for Crowdfunding

Reward Motivation

In reward-based crowdfunding, backers get non-monetary compensations for their investment which can be a priority to use products or coupons on products (Ahlers, Cumming, Günther, & Schweizer, 2015). In contrast, regarding lending- and equity-based crowdfunding, investors receive beneficial reimbursement from the project they invested (Ahlers et al., 2015). The reward could be expected as the major and apparent motivation of backers to participate in funding in all of the reward-based crowdfunding types. Specifically, the participant's motivation to fund is similar to business angels. That explains the reason why start-ups are funded by business angel investors, it is because of the priority goal of earning profit and/or capital gains on the invested capital (Feeney et al., 1999).
According to the incentive theory – one of the foundations of motivation research, extrinsic reward motivation is identified. It affirms that reinforcement or incentives could play a role as a deliberate design to drive behavior. In particular, it assumes that external targets such as reward or money could drive behavior (Hockenbury & Hockenbury, 2003). In the context of

reward-based crowdfunding, we assume that reward motivation also significantly influences funding intention because backers could expect rewards when the project is successfully funded. Hence, we propose:
Hypothesis 3: In the reward-based crowdfunding context, backers' reward motivation is positively associated with funding intention.

Backer's Image Motivation

In the crowdfunding system, the backers' information is displayed on the user profile webpage (UPW) including their profiles and the amount of money they have funded. Moreover, the exposure of funding behavior on systems is more likely to encourage the backers to fund projects since their online images are created and polished. Jabr, Mookerjee, Tan, and Mookerjee (2014) argued that the virtual status is an important motive for members in the online community to actively create their image and increase their status both in the real and virtual world. It is true that in an offline context, the more identities perceived, the more effective social communication is (Goffman, 1978). According to Ma and Agarwal (2007), in the IS field, researchers have more and more investigated in the online image as an important construct to explain user behavior in social networks such as LinkedIn or Facebook.
Liberman and Trope (2014) also proposed that personal image also urges programmers to participate more in open-source projects to enhance their reputation. Therefore, the image motive can increase funding behavior, from which the backer's profile is explicit, and his/her virtual status is highlighted. Hence, the above discussion leads us to develop the following hypothesis:
Hypothesis 4: In the reward-based crowdfunding context, backers' image motivation is positively associated with funding intention.

Study 1: Comparison Between Actual and Potential Backers

Construal Level Theory

To provide the basis for understanding funding motivation between actual and potential backers, the current research builds on the construal level theory (CLT). The main purpose is to explain the differences between potential and actual user backers based on the relationship between the direct experience and psychological distance from the funding activity. In general, CLT depicts that psychological distances, including temporal distances, spatial distances, social distances, and hypothetical distances, affect the way people perceive an event, person, or idea in the high or low degree of construal (Trope, Liberman, & Wakslak, 2007). Taking a broader view, the underline mechanism of CLT is that the more distant an object is from the individual, the higher the construal level is, and vice versa.

As aforementioned, past research has demonstrated that several different forms of psychological distance are depended on the "egocentric" perception of how far an object or activity is from the self. Supporting this claim, Lin and Viswanathan (2015) differentiates among four dimensions of psychological distance: time/temporal (time distance from object or event; the object belongs to the past or future), spatial (geographical distance from object or event), social (distance to others; they are similar/dissimilar, familiar/unfamiliar, ingroup/outgroup), and

hypothetical (e.g., high versus low probability, real versus hypothetical). More importantly, beyond these four dimensions, (Kleemann, Voß, & Rieder, 2008) singled out additional distance dimensions that also impact on construal level, namely experiential distance (e.g., experience versus no experience). CLT proposes to view the issue of psychological distance as an object or activity whenever it is not part of one's direct experience (Unger et al., 1995).

Based on the psychological distance concept, the present research explores the distal perspective in the context of crowdfunding and tends to identify the difference between potential and actual user backers influencing crowdfunding intention which might be affected by experiential distance. Specifically, actual backers are defined as those who have experienced funding based on their knowledge and high interest/needs in crowdfunding events. While potential backers are the ones who only have a basic understanding which can lead to their interest in crowdfunding events, however, they have not experienced yet.

Another important difference between potential and actual backers is the intensity of the above effect. Although many CLT studies point out that distal versus proximal perspectives toward an activity or event cause the different importance of desirability and feasibility concerns (Unger et al., 1995). Further, on the one hand, a distal, high-level construal is defined to emphasize desirability (e.g., interest, enjoyment) concerns of an individual. On the other hand, proximal, low-level construal stress on feasibility (e.g., convenience, ease) concerns toward performing the activity. In addition, it is important to note that, psychological distance from an event or activity (in this case, crowdfunding) is also associated with idealistic versus pragmatic concerns with respect to the activity (Unger et al., 1995). Nevertheless, a distal perspective (such as for potential backers) has been defined to shift attention inward toward the core values that describe a person, activating the idealistic self for the focused object or activity.

In another perspective, proximity to the target (such as for actual backers) is proposed to highlight extrinsic inducements (e.g., financial benefits) by activating the pragmatic self (Kleemann et al., 2008). Supporting this claim, several studies have been investigated in the effect of pragmatic versus idealistic dimension on different behavioral responses and choices (Trope & Liberman, 2010). Altogether, we argue that actual backer's greater proximity to the funding activity will trigger their pragmatic self, while extrinsic rewards could lead to strong funding intention as compared to the idealistic self that is evoked for potential user backers, which makes external rewards less salient for them. From these arguments, the hypotheses are stated as:

Hypothesis 5a: Altruism motivation is a stronger driver for potential backers' funding intention than for actual backers.

Hypothesis 5b: Liking motivation is a stronger driver for potential backers' funding intention than for actual backers.

Hypothesis 5c: Reward motivation is a stronger driver for actual backers' funding intention than for potential backers.

Hypothesis 5d: Image motivation is a stronger driver for actual backers' funding intention than for potential backers.

Study 2: Effects of Extrinsic on Intrinsic Motivation

Cognitive Evaluation Theory

Finding the most persuasive theory to explain for this topic, cognitive evaluation theory (CET) is selected. The theory developed from the 1970s argues that extrinsic motivation diminishes the role of intrinsic to some certain extents. Liberman and Trope (2014) developed the Cognitive Evaluation Theory to explain the results. In this theory, two aspects of intrinsic motivation are employed consisting of autonomy and competence in which the changes of those factors could be positive or negative related to extrinsic rewards.

Taken psychological processes into account, extrinsic motivation could influence to self-determination, and self-esteem. Specifically, previous studies argued that individual self-determination is decreased because of external interference credited by reward as extrinsic motivation (Deci, 1971; Frey & Jegen, 2001). Under the effects of extrinsic motivation, individuals tend to switch their focus from inside to the outside of themselves (Deci & Ryan, 1985) which is controlled by external influence (Deci, 1971). Receiving rewards as extrinsic motivation, people may feel that their enthusiasm and engagement have not existed. Therefore, an individual does not perceive that intrinsic factors motivate their behavior or even those factors may be eroded because of extrinsic motivation.

Accordingly, under the extrinsic motivations such as rewards, backers may invest in the project with other incentives rather than they enjoy in supporting others (i.e. Altruism) or the one they like (i.e. Liking motivation). As a result, the association between altruism and liking motivation to funding behavior is undermined because of extrinsic motivation. Hence, the following hypotheses are proposed:
Hypothesis 6a: The motivation to obtain rewards motivation negatively moderates the effect of altruism in helping others on an individual's attitude toward funding intention
Hypothesis 6b: The motivation to obtain image motivation negatively moderates the effect of altruism in helping others on an individual's attitude toward funding intention
Hypothesis 6c: The motivation to get rewards motivation negatively moderates the impact of liking motivation on funding intention.
Hypothesis 6d: The motivation to get image motivation negatively moderates the impact of liking motivation on funding intention.

Figure 1: Research model study 1

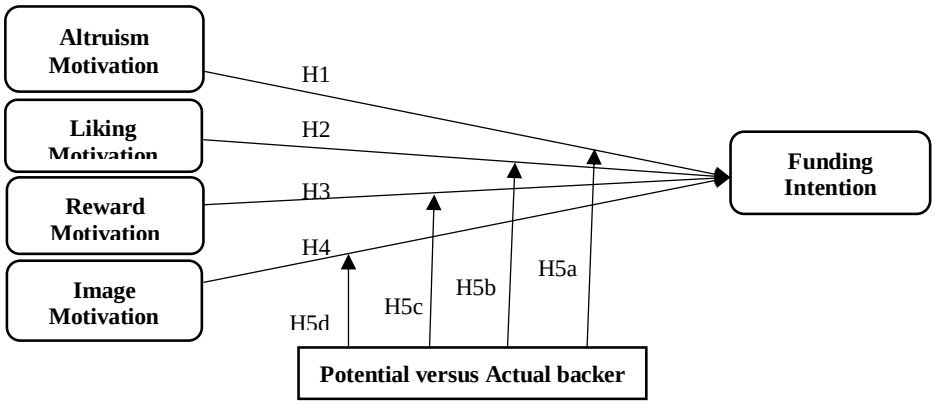
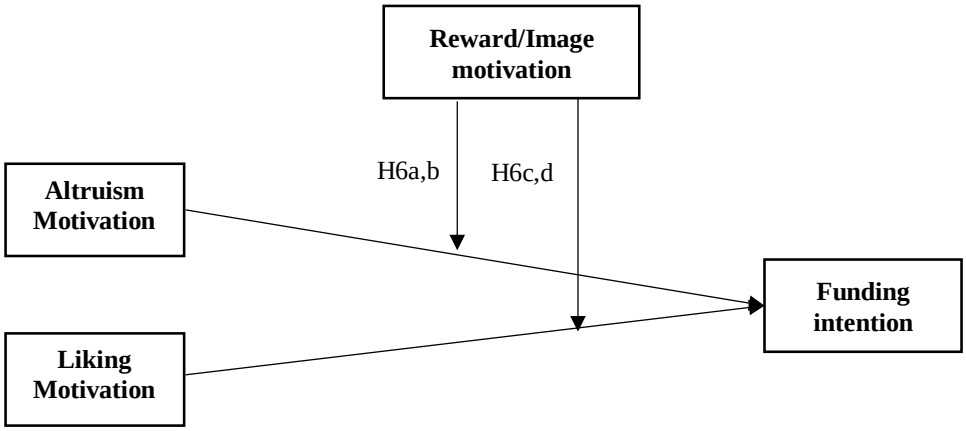


Figure 2: Research model study 2



METHODS

Measurements

The items of altruism, liking, reward, and image motivation are adapted from (Bretschneider & Leimeister, 2017). For funding intention, the measurements are adopted from (Olsen & Johnson, 2003).To rate the evaluation of respondents, the 7- point Likert scale is used ranging from 1 (strongly disagree) to 7 (strongly agree).

Data collection and analysis

The purpose of this study is to examine the motivations of funding intention and the differences in the antecedents of such intention between potential and actual funders. In line with the research objective, we target respondents who are either actual and potential crowd funders. To reach respondents, the online questionnaire will be posted on the fan page of crowdfunding platforms and also distributed in social network groups of crowd funders. Because of the popularity of crowdfunding, the recruited respondents should not be limited to any nationality. The proposed conceptual model will be tested by using a PLS-SEM approach.

EXPECTED IMPLICATIONS

The present study is expected to contribute to both academics and event organizers in reward-based crowdfunding. On the one hand, our findings contribute to the literature of crowdfunding. First, we provide an in-depth understanding of funder's motivations in reward-based crowdfunding by indicating how intrinsic and extrinsic motivations drive the funding intention of participants. Second, our research is one of the earliest studies uncovering extrinsic motivation's unexpected consequences in the context of reward-based crowdfunding. Particularly, extrinsic motivation plays a hidden role effecting the link between intrinsic motivation and funding intention. Third, our study investigates in both potential and actual backers to conceptualize the alternative theoretical model to compare the divergence among their motivation to funding intention. Eventually, by examining the psychological distance factor

which is widely ignored in previous studies (Mollick, 2014; Skirnevskiy, Bendig, & Brettel, 2017)**,** our study fills this gap to deliver a thorough view on funding behavior.

On the other hand, this research has important implications for both platform designers and project creators in the context of reward-based crowdfunding. First, the comprehensive understanding of funder's motivations, especially, the role of distinct motivations provided in this research could allow event organizers to design better platforms as well as the effective process of crowdfunding projects to attract more intention of potential funders. Second, we differentiate the participants into potential and actual funders instead of treating them as a homogeneous group. Thus, our findings provide platform designers and project creators suggestion that practitioners should distinguish their strategies for different purposes, specifically, create gravitation for potential funders to join and promote actual funders to contribute to the project.

APPENDIX A: **Measurement items**

Altruism	*(Bretschneider & Leimeister, 2017)*
Altruis1	I want to help people with good ideas on a crowdfunding platform without expecting any compensation
Altruis2	I like doing something for project initiators on a crowdfunding platform at some cost to myself
Altruis3	I deeply enjoy helping others on a crowdfunding platform – even if I have to make sacrifices
Liking	*(Bretschneider & Leimeister, 2017)*
Like1	In my eyes, there are some projects on a crowdfunding platform that has their own personal appeal. That is why I would like to support these projects in some way
Like2	I really like the idea behind some projects on a crowdfunding platform and, therefore, I would like to support these projects in any way
Like3	I would like to support some of the projects that I find particularly likeable
Reward	*(Bretschneider & Leimeister, 2017)*
Rew1	I contribute to projects to get a reward
Rew2	I engage in some of the projects on a crowdfunding platform in the prospect of a reward
Rew3	I aim at receiving rewards from projects I engage in on a crowdfunding platform
Image	*(Bretschneider & Leimeister, 2017)*
Image1	I want to impress others on a crowdfunding platform
Image2	I am aiming to give an impression to other backers on a crowdfunding platform
Image3	I want project initiators to have a positive image of me based on my behavior on a crowdfunding platform
Funding intention	*(Olsen & Johnson, 2003)*
Int1	The probability that I would consider funding a crowdfunding project on a crowdfunding platform is high.
Int2.	If I were to fund a project, I would consider a crowdfunding project on a crowdfunding platform.
Int3.	The likelihood of my funding to a crowdfunding platform is high
Int4.	My willingness to fund for a crowdfunding platform project is high

REFERENCES

Agrawal, A., Catalini, C., & Goldfarb, A. (2014). Some simple economics of crowdfunding. *Innovation policy and the economy, 14*(1), 63-97.

Ahlers, G. K., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice, 39*(4), 955-980.

Al-Jabri, I. M., & Al-Khaldi, M. A. (1997). Effects of user characteristics on computer attitudes among undergraduate business students. *Journal of Organizational and End User Computing (JOEUC), 9*(2), 16-23.

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice, 39*(1), 53-73.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of political Economy, 97*(6), 1447-1458.

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review, 99*(1), 544-555.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of business venturing, 29*(5), 585-609.

Berglin, H., & Strandberg, C. (2013). Leveraging customers as investors: The driving forces behind crowdfunding. In.

Bretschneider, U., & Leimeister, J. M. (2017). Not just an ego-trip: Exploring backers' motivation for funding in incentive-based crowdfunding. *The Journal of Strategic Information Systems, 26*(4), 246-260.

Brettel, M. (2003). Business angels in Germany: a research note.

Cholakova, M., & Clarysse, B. (2015). Does the possibility to make equity investments in crowdfunding projects crowd out reward–based investments? *Entrepreneurship Theory and Practice, 39*(1), 145-172.

De Dreu, C. K., & Nauta, A. (2009). Self-interest and other-orientation in organizational behavior: implications for job performance, prosocial behavior, and personal initiative. *Journal of Applied Psychology, 94*(4), 913.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and social psychology, 18*(1), 105.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality, 19*(2), 109-134.

Feeney, L., Haines Jr, G. H., & Riding, A. L. (1999). Private investors' investment criteria: insights from qualitative data. *Venture Capital: An international journal of entrepreneurial finance, 1*(2), 121-145.

Fleming, L., & Sorenson, O. (2016). Financing by and for the Masses: An Introduction to the Special Issue on Crowdfunding. *California Management Review, 58*(2), 5-19.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys, 15*(5), 589-611.

Gagné, M., & Deci, E. L. (2005). Self‐determination theory and work motivation. *Journal of Organizational behavior, 26*(4), 331-362.

Gerber, E. M., & Hui, J. (2013). Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI), 20*(6), 1-32.
Gerber, E. M., Hui, J. S., & Kuo, P.-Y. (2012). *Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms.* Paper presented at the Proceedings of the international workshop on design, influence, and social technologies: techniques, impacts and ethics.
Gleasure, R., & Feller, J. (2016). Does heart or head rule donor behaviors in charitable crowdfunding markets? *International Journal of Electronic Commerce, 20*(4), 499-524.
Goffman, E. (1978). *The presentation of self in everyday life*: Harmondsworth London.
Harbaugh, W. T. (1998). What do donations buy?: A model of philanthropy based on prestige and warm glow. *Journal of Public Economics, 67*(2), 269-284.
Hockenbury, D., & Hockenbury, S. (2003). Motivation and emotion. *Psychology*, 319-365.
Jabr, W., Mookerjee, R., Tan, Y., & Mookerjee, V. S. (2014). Leveraging philanthropic behavior for customer support: The case of user support forums. *MIS quarterly, 38*(1), 187-208.
Kankanhalli, A. (2015). Comparing Potential and Actual Innovators: An Empirical Study of Mobile Data Services Innovation. *MIS quarterly, 39*(3).
Ke, W., & Zhang, P. (2009). Motivations in open source software communities: The mediating role of effort intensity and goal commitment. *International Journal of Electronic Commerce, 13*(4), 39-66.
Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utiliza-tion of consumer work through crowdsourcing. *Science, technology & innovation studies, 4*(1), PP. 5-26.
Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences, 18*(7), 364-369.
Lin, M., & Viswanathan, S. (2015). Home bias in online investments: An empirical study of an online crowdfunding market. *Management science, 62*(5), 1393-1414.
Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information systems research, 18*(1), 42-67.
Mason, C., & Stark, M. (2004). What do investors look for in a business plan? A comparison of the investment criteria of bankers, venture capitalists and business angels. *International small business journal, 22*(3), 227-248.
Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing, 29*(1), 1-16.
Moysidou, K., & Spaeth, S. (2016). *Cognition, emotion and perceived values in crowdfunding decision making.* Paper presented at the 14TH INTERNATIONAL CONFERENCE.
Olsen, L. L., & Johnson, M. D. (2003). Service equity, satisfaction, and loyalty: from transaction-specific to cumulative evaluations. *Journal of Service Research, 5*(3), 184-195.
Ozinga, J. R. (1999). *Altruism*: Greenwood Publishing Group.
Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS quarterly*, 115-143.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology, 25*(1), 54-67.
Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55*(1), 68.
Sabherwal, R., Jeyaraj, A., & Chowa, C. (2006). Information system success: Individual and organizational determinants. *Management science, 52*(12), 1849-1864.
Skirnevskiy, V., Bendig, D., & Brettel, M. (2017). The influence of internal social capital on serial creators' success in crowdfunding. *Entrepreneurship Theory and Practice, 41*(2), 209-236.
Steers, R. M., Mowday, R. T., & Shapiro, D. L. (2004). The future of work motivation theory. *Academy of Management review, 29*(3), 379-387.
Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review, 117*(2), 440.
Trope, Y., Liberman, N., & Wakslak, C. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of consumer psychology, 17*(2), 83-95.
Unger, J. B., Johnson, C. A., & Rohrbach, L. A. (1995). Recognition and liking of tobacco and alcohol advertisements among adolescents: relationships with susceptibility to substance use. *Preventive medicine, 24*(5), 461-466.
Xiao, S., & Yue, Q. (2018). Investors' inertia behavior and their repeated decision-making in online reward-based crowdfunding market. *Decision Support Systems, 111*, 101-112.
Zwass, V. (2010). Co-creation: Toward a taxonomy and an integrated research perspective. *International Journal of Electronic Commerce, 15*(1), 11-48.

DECISION SCIENCES INSTITUTE

Customer type or customer preference: A modeling approach to evaluate the value of customer information in a sales scenario

Hisashi Kurata
Yokohama National University
Email: kurata-hisashi-zc@ynu.ac.jp

ABSTRACT

This work analyzes the relative values of two types of information—customer type and the customer's service preference, while assuming two types of customers: those that have already made a purchase decision and those that have not. After developing a relevant mathematical model, we found that if a retailer has a high brand value and a high-cost structure, correctly identifying the customer type will be more profitable than identifying their service preference. Additionally, between state-of-the-art information technology and simple industrial engineering tools, we determined the most suitable methodology for obtaining each type of information.

KEYWORDS: Value of information, Webrooming, Showrooming, Heterogenous consumers, Personal sales, IT or IE selection

INTRODUCTION

Personal sales support offered by staff on the shop floor is a valuable tool for retailers to increase their sales. In fact, PR Newswire (July 8, 2015), a popular press release distributor, reported that the service offered by sales staff is the top factor influencing customer satisfaction at particular electronic appliance shop; moreover, an experimental study concluded that storytelling and salesperson's communication with the customers have the greatest impact on a customers' intent to purchase the given product/service (Gilliam & Zablah, 2013). In practice, the salesperson must consider two aspects when communicating with a customer: understanding the customer's needs and wants and addressing their inquisitiveness regarding a product of interest by offering adequate information. Thus, a salesperson is expected to simultaneously "get to know" the customer and provide them with the necessary information. This study analyzes how the value of the information related to these two tasks contributes to a retailer's profit increase when the customers' preference for personal service is heterogenous, where a mismatch between the customer's preference and the quality of the actual service exists such that the customer's sense of satisfaction is reduced.

Skilled sales staff can recognize the unexpressed and often minor needs and tastes of customers better than the customers themselves by picking up on the latter's nonverbal cues, such as small actions, slight gestures, or a minor change in expression. However, while this skill of studying the customer's body language is crucial to the retailer's success, it is a difficult skill to master. Thus, having experience in retail has historically been deemed crucial to being a sales expert. However, the rapid development in information technology (IT) has transformed this situation. For example, radio frequency identification (RFID) technology, which can trace the in-store movement of a shopper and evaluate their hidden purchasing behavior, helps to increase both the store's profitability and customer satisfaction (Hui et al., 2009, 2013a, 2013b). Understanding customer behavior using IT has thus become a recent trend in retail operations.

While both skilled sales staff and state-of-the-art IT enhance the in-store service quality, intentional limiting of the human touch is often considered an effective technique, even for the sale of high-end products. One such example is the “service as you like it” stress-free sales style offered by Clinique, a major cosmetics and skincare product brand in USA. At Clinique, customers can experience products without the assistance of a salesperson, creating a so-called “open” sell environment, which, as the media have consistently reported has become a recent trend in cosmetics retailing (Holmes, 2012). Another example is the “no-sales-talk shopping bag” environment provided by Urban Research, a casual fashion apparel chain in Japan. At Urban Research, customers who find sales pitches annoying and unnecessary can enjoy a quiet shopping experience by displaying a particular “shopping bag.” Further, sales assistance offered by machines, including those based on artificial intelligence (AI), is another trend in modern retailing. In fact, an empirical study concluded that 73% customers prefer to use self-service systems, whereas 80% businesses have adopted or plan to adopt AI in difference forms, such as chatbots, to deal with customer concerns (Layun, 2017). Therefore, one challenge for any retail manager relates to deciding whether or not to offer comprehensive human sales assistance to customers by accurately judging their preference Additionally, if a retailer identifies that a certain customer prefers the human touch, the next challenge is to determine the optimal service for this customer. In short, understanding the customer type and determining the appropriate form of service for each customer are the current predominant issues in retail.

To address these issues, this work considers two types of customers who might visit a physical store. The first type is the *undetermined customer* who has not yet decided whether to buy a product of interest because they are uncertain of the product’s quality and functionality. This type of customer requires personal sales assistance to obtain the necessary information. This customer group also includes the so-called *showroomers*. This label is given to customers who tend to evaluate the product at a physical store before buying it online. Fassnacht et al. (2019) stated that *showroomers* use sales staff to obtain the necessary product information so that they can decide on the most suitable product for them. As such, the sales staff can change the *showroomer’s* purchase intention and encourage them to buy the product at the physical store (Fassnacht et al. 2019). Hence, determining the appropriate level of personal service is critical for any physical store. The second customer type is the *predetermined customer*, who has decided which product to buy before they visit a physical store. The so-called *webroomers* collect product information online and then buy the product at a physical store, and they are typical predetermined customers. Ferdandez et al. (2018) stated that *webroomers* are in the dominant segment of research shoppers and concluded that sales staff find it challenging to influence the purchase intention of such customers.

Therefore, herein, we focus on two aspects that a successful personal customer service must consider: identifying the type of the customer as “predetermined” or “undetermined” and determining the appropriate service to be provided, which is particularly important in the case of “undetermined” customers. When considering the aforementioned customer types, a customer’s preference for personal service, and a scenario wherein the sales staff offer a personal service (sales pitch) to a customer to increase the sales of a certain product, the following questions must be addressed.

1. Which information is more valuable for a retailer: the correct identification of customer type or the correct level of customer service? Further, the sales staff must identify whether a given customer is undetermined or predetermined. A good sales pitch could convince the undermined customer whereas sales assistance will be ineffective to the predetermined customer, or in some cases, such sales assistance might be regarded as annoying.
2. What is the optimal service quality for the customers in the retail situation?

3. Which method should be recommended for obtaining customer-related information: an expensive state-of-the-art IT (e.g., the adoption of RFID to trace a shopper's in-store movement) or an inexpensive, simple industrial engineering (IE) method (e.g., the service pokayoke system).

The discussion derived from our analytical model analysis suggests that a firm can decide which approach is more important, identifying the customer type or identifying their service preference, depending on two aspects of uncertainty: an inaccurate estimation of the optimal service level and the customer's uncertainty regarding product quality. Specifically, we find that if a retailer has high brand value and a high-cost structure, correctly ascertaining the customer type would be more profitable than providing information on a product. Following this inference, we derive several managerial implications in terms of identifying which method is more suitable for obtaining customer-related information, state-of-the-art IT or a traditional, simple IE method. The focus on the selection of IT or IE is a significant contribution of this work.

LITERATURE REVIEW

This paper relates to several research subjects, the first of which is the determination of an adequate level of personal service quality. This subject is researched as oftentimes a high-quality service including personal sales assistance is not always favored by consumers. Using the game theory model, Ba et al. (2010) determined the optimal mix of human service and digital service under the assumption that consumers preference in terms of an IT-based service is largely heterogenous. Elsewhere, Pierce and Aguinis (2013) proposed a meta-theoretical principle labeled the "too-much-of-a-good-thing effect," a phenomenon where a trend involving largely positive relations becomes asymptotic and negative. Furthermore, Bellos and Kavadias (2019) explored a service system design wherein the service provider optimally places its service processes under the customer's control. Therein, the authors found that a customer's role in controlling the service process is more important in non-routine services than in routine services.

A customer's "supersaturation" in terms of over-serving is also related to this subject. Kurata and Nam (2010) considered a situation in which a retailer offers an optional after-sales service to the buyer of an electronic appliance in addition to the manufacturer's service, which they claims causes customer supersaturation. The authors subsequently extended this research by including the aspect of uncertainty (Kurata & Nam 2013). After identifying two customer groups, one with a preference for high-quality and one with a preference for moderate-quality, Kurata (2019) then analytical studied the value of the information related to customer preference and the value of determining optimal service quality.

The second research subject involves the timing of the customer when deciding to a purchase. In an omnichannel scenario, a customer could purchase online in advance and then visit a physical store to collect the ordered item, which is a process known as buy-online-and-pick-up-in-store (BOPS). Gao and Su (2016) studied the effect of BOPS-related behavior on a retailer's performance, whereas Jin et al. (2018) explored a store's service area for BOPS collections as part of an omnichannel strategy. Further, using a two-period model, Shi et al. (2018) analyzed the relationship between customers' pre-orders and product returns concerning to an omnichannel approach. The predetermined customer group includes the *webroomers*, who collect product's information online and then buy the product at a physical store. Ferdandez et al. (2018) investigated the concept of research shopping, which involves researching a product's attributes through one channel before buying it through another. Note that BOPS customers and

webroomers are different. The former already purchased an item online before visiting a physical store whereas the latter purchase an item at a physical store after assessing product information online. However, both groups have in common that they collect and evaluate product information online before going to a brick-and-mortar store. In practice, certain customer groups visit a store without making a purchase decision in advance. Here, the *showrooming* a customer makes an online purchase after examining a product's quality and functionalities at a physical store and is a typical example of undecided store visiting. An experimental study by Fassnacht et al. (2019) found that high-quality interaction between *showroomers* and sales personnel could increase the probability of the former making an in-store purchase. Thus, in such a case, two types of purchase decision making come into play: deciding before visiting a store and deciding after visiting a store. This work classifies these two types of decision making as predetermined and undermined purchase decisions.

The third research subject relates to the heterogeneity of customer behavior and correctly identifying a consumer's type. Davidow and Uttal (1989) claimed that the selection of an optimal service that is in accordance with various customer segments is critical, with "too much" service or "too little" service ultimately causing loss of customers. In an empirical study, Díaz-Martín et al. (2000) stated that segmenting customers according to their service expectations contributes to a service provider's performance. Additionally, using data related to the tourism industry, Fuentes-Blasco, et al. (2017) examined how retail-related technological and marketing innovations affect customer satisfaction and word-of-mouth. Technology and marketing innovations affect customer satisfaction and word-of-mouth publicity. Therein, the authors concluded that technological innovation has a stronger impact on store image and customer satisfaction than marketing innovation. Elsewhere, Kurata (2019) analyzed whether customer-type information or service preference information more significantly impacts a retailer's performance.

This article encompasses all the above research subjects. However, the novelty of this work lies in the following points. First, as a part of a retailer's personal service provision to a largely heterogenous customer base, we consider the choice between a minimal service without a sales pitch and an elaborate personalized service. Second, we consider the consumers' purchase decision sequence, i.e., whether it is "predetermined" or "undetermined." This classification is associated with two common consumer segments in the modern information society, one of which includes *webroomers* and the other includes *showroomers*. Third, we discuss managerial issue of determining the most appropriate method for obtaining customer information, i.e., whether data collection should be IT-based or IE-based. To the best of our knowledge, this paper is the first to simultaneously address the abovementioned issues.

MODEL DEVELOPMENT

This work models a sales situation in which a salesperson offers face-to-face service to a customer with the expectation that the customer will purchase a certain product. Personal service can help address the uncertainty and inquisitiveness of a customer regarding a specific product, increasing the probability of purchase. However, the disparity between the customer's service preference and the actual service may decrease the probability of purchase. Two types of customers are considered; the first of which is the "predetermined" customer who has already decided which product to buy and fully understands the quality of the product thereby requiring minimal personal service. In such a case, the predetermined customer is indifferent to the unnecessary attention by sales staff. This unnecessary attention could decrease the probability of purchase if the customer is sensitive. The second type is the "undetermined" customer, who

has not decided whether to buy a product of interest and has some uncertainty regarding the quality of the product. Thus, an adequate level of personal service will be required to satisfy the customer's inquisitiveness to increase the probability of a purchase. All the notations are listed in Table 1.

Table 1: Notations and symbols

Symbols			
x	Customer's preference to personal service, $x \sim Uni(0,1)$.	U_i	Utility of a type *i* customer; $i \in \{P, U\}$.
s	The service quality for the undetermined customers, $s \in [0,1]$.	D_j	Demand of the situation *j*; $j \in \{P, U, MIS\}$.
δ	Estimation effort of the value s, $\delta > 0$.	S_j	Actual sales of the situation *j*; $j \in \{P, U, MIS\}$.
$g = g(s)$	An undetermined customer's uncertainty regarding a product's quality.	π_I	Profit of Scenario *h*; $h \in \{I, II, III, IV\}$.
α, β	The negative impacts of over quality ($\alpha > 0$) and under quality ($\beta > 0$).	k_s	Coefficient of service offering cost; $k_s > 0$.
γ	A general effect of service quality, $\gamma > 0$.	Δ_{A-B}	$\equiv \pi_{II} - \pi_{III}$; $A, B \in \{I, II, III, IV\}$.
θ	The probability that a customer type is wrongly judged, $\theta \in [0.5, 1]$.		
Superscripts and subscripts			
U, P	Undetermined and predetermined customers, respectively	I, II, III, IV	Scenarios I, II, III, and IV, respectively.
MIS	A situation in which a customer type is misjudged.	*	The optimal value
$\overline{\ }$, $\underline{\ }$	Upper-limit and lower-limit of the range of non-negative utility, respectively.		

Key Assumptions

We used the following assumptions to build a model for the above research questions. Notably, our model includes the assumptions and formulation of the utility, demand, and profit functions and predominantly follows the work of Kurata (2019).

[1] Two types of customers are considered herein: "predetermined" customers (P) and "undetermined" consumers (U).
[2] Market sizes of the two customer groups are the same and both markets are standardized.
[3] Customer preference regarding personal service (x) is heterogenous and is assumed to be uniformly distributed over the 0–1 interval (i.e., $x \sim Uni(0,1)$). Here, $x = 0$ represents a situation that a shop offers a minimum level of personal service, and $x = 1$ represents the shop offering the highest possible level of personal service.

[4] The most favored service quality of "predetermined" customers is assumed to be 0. Thus, a minimal service is adequate for them because they require no further information about the product. The favored service for "undetermined" customers is denoted as $s(\in [0,1])$, which means a retailer must determine the most appropriate service.

[5] Two situations are considered regarding s: a non-uncertain situation wherein a retailer knows the exact value of s and an uncertain situation wherein the value of s is estimated using the Bernoulli distribution of $\pm\delta$ $(\delta > 0)$, with a 50% occurrence for each $-\delta$ or $+\delta$.

[6] A "predetermined" customer may buy a product provided the utility of the personal service provided by a salesperson is non-negative. In this case, product uncertainty resolution via personal service is not required.

[7] When an "undetermined" customer has some uncertainty regarding a product's quality $(g = g(s))$, the product's utility may decrease for the customer. However, good customer service can overturn this negative effect (i.e., $\partial g/\partial s < 0$) such that the utility becomes non-negative, and the "undetermined" customer buys the product.

[8] The disparity between a customer's service preference (x) and the actual service (s) results in some disutility. The impacts of over quality (α) and under quality (β) differ as confirmed by several behavior-related studies (e.g., Kahneman & Tversky, 1979).

[9] The probability that a staff member misinterprets the customer type is denoted by a probability of $\theta \in [0.5, 1]$ where $\theta = 1$ implies that the judgment is perfect, whereas $\theta = 0.5$ implies that the staff member has randomly interpreted the customer type (i.e., the worst-case scenario).

[10] It is common for high-quality service to generate a better brand image and excellent reputation, thereby positively affecting a customer's perceived quality. Personal selling as a part of marketing communication builds brand equity, including the brand image (Kotler et al., Chap.19). Various empirical studies have shown that a company's image is formulated based on the perception of service quality (e.g., Lai et al. 2009 and Nyuyen and LeBlanc, 1998), that service quality significantly influences brand equity (e.g., Esmaeilpour et al. 2016), and that high service quality generates customer loyalty (e.g., Das 2014). Therefore, we reasonably assume that high-quality service would increase the customer's utility (denoted as γ).

[11] The cost of the service provision is assumed to be a quadratic function of s, which represents the diminishing return of investment on personal service. Quadratic cost functions are common in analytical research on retail modeling (e.g., Iyer, 1998; Tsay & Agrawal, 2000; Desai et al. 2001; Ha et al. 2011).

Utility Functions

Utility functions of the "predetermined" (Eq. 1) and "undetermined" customers (Eq. 2) are defined using the utility setting of Kurata (2019)

$$U_P = m_P - \alpha x + \gamma x. \quad (1)$$

$$U_U = m_U - \alpha(s - x)^+ - \beta(x - s)^+ + \gamma x, \text{ where } A^+ \equiv max(0, A). \quad (2)$$

Here, the potential utility or reservation price is denoted by m_i, $i \in \{P, U\}$; α and β represent the impact of over quality and under quality, respectively (see Assumption 8). Thus, as the offered service quality increasingly deviates from the customer's preference, the utility becomes smaller. The general effect of service quality (see Assumption 10) is denoted by γ. Notably, the "predetermined" customers are assumed to prefer minimum service because they have already collected enough product information via the internet before coming to a physical store. Thus, for

simplicity, we set the optimal service quality for them as zero. Hence, only the issue of over quality denoted as $-\alpha x$ in Equation (1) pertains to predetermined consumers in (1). However, we assume that the optimal quality for the undetermined customers is between 0 and 1. Hence, both over quality and under quality occur for them, as shown in the terms $-\alpha(s-x)^+$ and $-\beta(x-s)^+$, respectively, in (2). Equations (1) and (2) yield the points where utility is zero: $\underline{x}_U = \frac{\alpha s - m_U}{\alpha+\gamma}$, $\bar{x}_P = \frac{m_P}{\alpha-\gamma}$, and $\bar{x}_U = \frac{m_U+\alpha s}{\alpha-\gamma}$ as shown in Figure 1. We assume $0 < \underline{x}_U < \bar{x}_P < \bar{x}_U < 1$ to eliminate a trivial outcome.

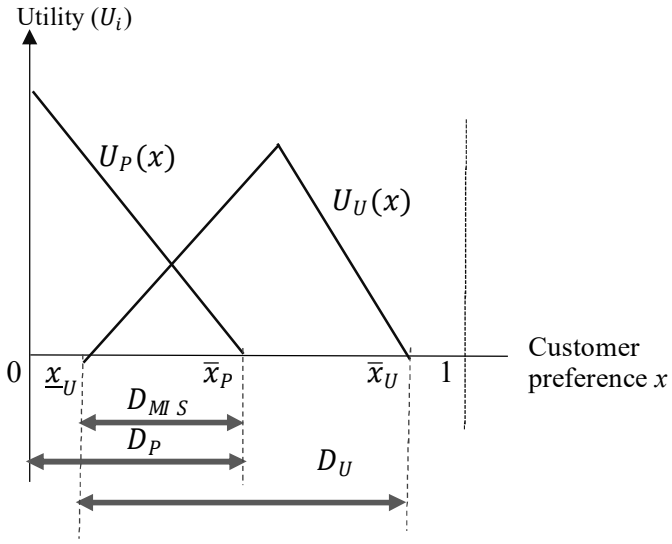


Figure 1. Utility functions and customer demands

Demand Functions and Sales Amounts

Because we assume that only a customer with a non-negative sense of utility buys a product, the demands of the "predetermined" and "undetermined" groups are defined as shown in Figure 1. Note that this customer demand modeling follows the work of Kurata (2019).

$$D_P = \bar{x}_P - 0 = \frac{m_P}{\alpha-\gamma}. \quad (3)$$

$$D_U = \bar{x}_U - \underline{x}_U = \frac{m_U+\alpha s}{\alpha-\gamma} - \frac{\alpha s - m_U}{\alpha+\gamma}. \quad (4)$$

In addition, when the customer's type is incorrectly determined, customer demand can be considered as the distance between $\bar{x}_P$ and $\underline{x}_U$:

$$D_{MIS} = \bar{x}_P - \underline{x}_U = \frac{m_P}{\alpha-\gamma} - \frac{\alpha s - m_U}{\alpha+\gamma}. \quad (5)$$

In our model setting, an "undetermined" customer has uncertainty regarding a product's quality, which diminishes the customer's sense of utility. The actual sales amount, $S_j, j \in \{P, D, MIS\}$, becomes less than the original demand, $D_j, j \in \{P, D, MIS\}$, because it diminished its sense of

utility; for tractability, we model this linearly, $\tau_0(1-s)$, where τ_0 represents the maximum impact of the diminished sense of utility (See Figure 2). This model implies that personal service can resolve the negative impact of product uncertainty. The diminished sense of utility changes the potential utility in Equations (1) and (2) from m_U to $m_U - \tau_0(1-s)$ for "undetermined" customers. Hence, the actual sales amounts can be determined as follows:

$$S_P = D_P = \frac{m_P}{\alpha-\gamma}. \quad (6)$$

$$S_U = \frac{m_U-\tau_0(1-s)+\alpha s}{\alpha-\gamma} - \frac{\alpha s-m_U+\tau_0(1-s)}{\alpha+\gamma}. \quad (7)$$

$$S_{MIS} = \frac{m_P}{\alpha-\gamma} - \frac{\alpha s-m_U+\tau_0(1-s)}{\alpha+\gamma}. \quad (8)$$

Notably, half of the misinterpreted customers are assumed to be influenced by product uncertainty because we have assumed that the size of the two customer groups is the same.

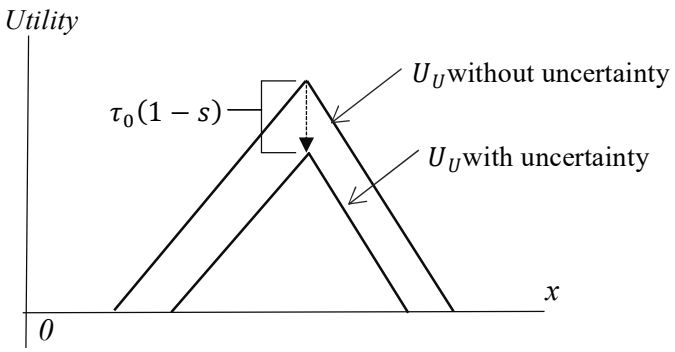


Figure 2. Utility reduction due to uncertainty regarding product quality.

Profit Functions and Four Scenarios

A profit function comprises a revenue defined as the product of unit revenue and the sales amount added to a quadratic cost function denoting the cost of offering a personal service, $k_s s^2$:

$$\pi = r(\text{total sales size}) - k_s s^2.$$

Here, we set up four scenarios to evaluate the value of information.

Scenario I.

This scenario is the ideal situation in which a store can precisely determine the customer type and where uncertainty regarding product quality does not exist. In this scenario, the profit is determined as follows:

$$\pi_I = r(S_P + S_U) - k_s s^2. \quad (9)$$

Scenario II.

This scenario represents a situation wherein a store makes an erroneous judgment in terms of customer type, represented by the probability $\theta(\in [0.5,1])$. Note that $\theta = 0.5$ indicates the worst-

case scenario in which the customer type is randomly decided, and Scenario II is equivalent to Scenario I when $\theta = 1$. The profit function of Scenario II is determined as follows:

$$\pi_{II} = r\{\theta(S_P + S_U) + (1-\theta)S_{MIS}\} - k_s s^2. \quad (10)$$

Scenario III.
In this scenario, we consider the uncertainty regarding the determination of optimal service level. Here, we assume that the service level deviates from the actual optimal service level by either $\delta(>0)$ or $-\delta$. As Figure 3 shows, this deviation negatively affects the purchase intention of the "undetermined" customers by δ, whereas that of the "predetermined" customers remains unaffected by this deviation. The profit function of Scenario III is determined as follows:

$$\pi_{III} = r(S_P + S_U - \delta) - k_s s^2. \quad (11)$$

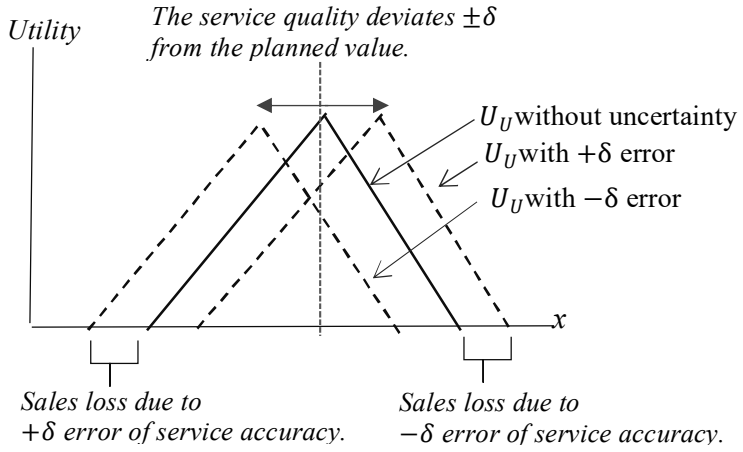


Figure 3. Effect of the inaccuracy of the service provision.

Scenario IV.
Scenario IV is set as the worst-case scenario including two erroneous factors, the misjudgment of customer type and deviation from the optimal service quality. Hence, the profit function of Scenario IV is determined as follows:

$$\pi_{IV} = r\left\{\theta(S_P + S_U - \delta) + (1-\theta)\left(S_{MIS} - \frac{\delta}{2}\right)\right\} - k_s s^2. \quad (12)$$

MODEL ANALYSIS

Here, we analyze the profit functions by comparing the four scenarios.

Comparison When the Service Quality is Exogenously Fixed

This subsection compares the profits of the four scenarios when the personal service quality level s is exogenously fixed. Often mass retailers set the level of personal customer service

according to standardized employee manuals and training programs; thus, implementing any change to the personal service quality is difficult or impractical. Performance comparison under the four scenarios in a fixed service quality situation can be summarized as Proposition 1.

Proposition 1.
(a)
The order of the profits $\pi_i, i \in \{I, II, III, IV\}$ *is determined as*

$$\pi_I > \pi_{II} > \pi_{IV} \text{ and } \pi_I > \pi_{III} > \pi_{IV}.$$

(b) Let us define Δ_{I-III} *as*

$$\Delta_{I-III} \equiv \pi_{II} - \pi_{III} = r\{(1-\theta)(S_{MIS} - S_P - S_U) + \delta\}. \quad (13)$$

Thus,

$$\text{if } \delta > (1-\theta)(S_P + S_U - S_{MIS}), \text{ then } \Delta_{I-III} > 0.$$
$$Other\ wise, \Delta_{I-III} \leq 0.$$

Proof. A simple calculation of the profit functions (9) through (12) results in this proposition.

Lemma 1
The probability of $\Delta_{I-III} > 0$ *tends to be high when* δ *and* θ *are relatively high but r and* τ_0 *are relatively low.*

Proof. A simple calculation of the sensitivity of the profit difference (13) in terms of the parameters of interest results in this Lemma.

We know that if the uncertainty is relatively high (i.e., δ and θ have high values) and the service quality level as well as the salesperson's resolution of customer inquisitiveness are relatively low (i.e., s and τ_0 have low values), improving the accuracy of the determination of customer type better increases profitability compared with improving the accuracy of service determination. This inference can be interpreted as follows: for a mass retailer, such as a large discount shop chain with low resources, focusing on correctly identifying customer type would be more reasonable than focusing on service preference.

Optimal Service Quality Levels and Their Comparison.

This subsection determines the optimal service quality levels ($s_i^*, i \in \{I, II, III, IV\}$) of the four scenarios for a profit-maximizing retailer and subsequently compares them. First, we determined the optimal service quality of each scenario.

Proposition 2.
(a) The optimal service quality levels were determined as follows:

$$s_I^* = s_{III}^* = \frac{r}{2k_s}\left(\frac{\alpha+\tau_0}{\alpha-\gamma} - \frac{\alpha-\tau_0}{\alpha+\gamma}\right)$$

$$s_{II}^* = s_{IV}^* = \frac{r}{2k_s}\left(\frac{\alpha+\tau_0}{\alpha-\gamma}\theta - \frac{\alpha-\tau_0}{\alpha+\gamma}\right)$$

(b) The relationships among the four optimal profits can be represented as follows:

$$s_I^* = s_{III}^* > s_{II}^* = s_{IV}^* \text{ for } 0.5 \le \theta < 1.$$

Proof. We obtained these results by differentiating the profit function in terms of s.

Proposition 2 reveals that when a retailer frequently errs in identifying the customer type, the quality of the personal service for "predetermined" customers decreases. In a sense, this outcome is perfectly logical: if one group of shoppers comprises "predetermined" customers who require a minimum level of personal service, the service quality should be set relatively low for the other groups, which may mitigate the negative impact of erroneous customer type identification. In contrast, in our framework, high-quality personal service is positively related to the store's reputation and brand value. Hence, while an intentionally low-setting service quality is acceptable from the perspective of individual shoppers, this can be risky from the store's perspective. If a retailer can accurately identify customer type, the micro-vs.-macro conflict will never occur.

Comparison of Maximum Profit among the Four Scenarios

This subsection examines the order of the maximum profits of the four scenarios when the optimal service quality is endogenously determined according to Proposition 2. Note that we considered the worst-case scenario of customer type determination (i.e., $\theta = 0.5$) following the analysis conducted by Kurata (2019). Proposition 3 summarizes the answer to this scenario.

Proposition 3.
Let us assume the worst-case scenario of $\theta = 0.5$.
(a)

$$\pi_I^* > \pi_{III}^* \text{ and } \pi_I^* > \pi_{III}^* \text{ for } \delta > 0.$$

(b)

$$\pi_{III}^* > \pi_{IV}^* \text{ and } \pi_{III}^* > \pi_{IV}^* \text{ for } \delta > 0.$$

(c) Let us define $F(\tau_0)$ *as*

$$F(\tau_0) \equiv r(\alpha + \tau_0)^2 - 2k_S(\alpha - \gamma)(m_U - \tau_0) + 4k_S\delta(\alpha - \gamma)^2.$$

(c-i)

If $\delta > \frac{m_U}{2(\alpha-\gamma)} - \frac{r\alpha^2}{4k_S(\alpha-\gamma)^2}$, *then* $\pi_{III}^* > \pi_{III}^*$

(c-ii)

If $\delta \le \frac{m_U}{2(\alpha-\gamma)} - \frac{r\alpha^2}{4k_S(\alpha-\gamma)^2}$,

then if $0 \le \tau_0 < \bar{\tau}_0$, *then* $\pi_{III}^* < \pi_{III}^*$; *if* $\tau_0 = \bar{\tau}_0$, *then* $\pi_{III}^* = \pi_{III}^*$; *otherwise,* $\pi_{III}^* \ge \pi_{III}^*$

Notably, we define $\bar{\tau}_0$ *as the greater root of the quadratic equation* $F(\tau_0) = 0$.

Proof. The answers in (a) and (b) are self-evident. For (c-i), we set $\Delta\pi_{III-III}^* \equiv \pi_{III}^* - \pi_{III}^*$ Here, a simple calculation results in the following expression:

$$\Delta\pi^*_{I\text{–}III} = \frac{r}{4k_s} \cdot \frac{\alpha+\tau_0}{\alpha-\gamma} \cdot \frac{r(\alpha+\tau_0)^2 - 2k_s(\alpha-\gamma)(m_U-\tau_0) + 4k_s\delta(\alpha-\gamma)^2}{(\alpha+\tau_0)(\alpha-\gamma)}. \quad (14)$$

We know from (14) that the sign of $\Delta\pi^*_{I\text{–}III}$ is the same as that of $F(\tau_0)$. Here, $F(\tau_0)$ can be rewritten as follows:

$$F(\tau_0) = r\tau_0^2 + 2\{\alpha + k_s(\alpha-\gamma)\}\tau_0 + [r\alpha^2 - 2k_s(\alpha-\gamma)m_U + 4k_s\delta(\alpha-\gamma)^2]. \quad (15)$$

From (15), if the value inside the square bracket [ ] is positive, which is equivalent to $\delta > \frac{m_U}{2(\alpha-\gamma)} - \frac{r\alpha^2}{4k_S(\alpha-\gamma)^2}$, then $F(\tau_0)$ is always positive, implying $\pi^*_{II} > \pi^*_{III}$ (c-ii) If $\delta < \frac{m_U}{2(\alpha-\gamma)} - \frac{r\alpha^2}{4k_S(\alpha-\gamma)^2}$, $F(0) < 0$. Thus, demonstrating that if $0 \le \tau_0 < \bar{\tau}_0$, *then* $\pi^*_{II} < \pi^*_{III}$; if $\tau_0 < \bar{\tau}_0$, *then* $\pi^*_{II} = \pi^*_{III}$; otherwise , $\pi^*_{II} > \pi^*_{III}$ is easy. The same logic can be used in the case of $F(0) = 0$.

Figure 4 presents a visual representation of Proposition 3c. Here, if a store's uncertainty in terms of service level (i.e., δ) and/or the customer's uncertainty regarding product quality (i.e., τ_0) are relatively large, the profit in Scenario 2 will tend to be greater than that in Scenario 3. In contrast, if both the uncertainties are relatively small, Scenario 3 will prove more profitable for a retailer than Scenario 2. The question remains as to which aspect is more valuable for the retailer, customer type or customer preference. Here, the value of the information is measured as the difference between Scenarios 2 and 4 for the value related to customer type and between Scenarios 3 and 4 for the value related to service preference.

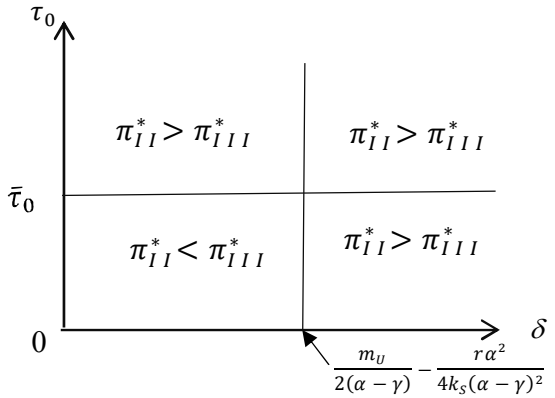


Figure 4. Comparison between π^*_{II} and π^*_{III} in relation to τ_0 and δ.

The implications identified through Figure 4 are as follows. First, when ascertaining customer preference (i.e., high δ) and/or ascertaining the product quality is sufficiently difficult (i.e., high τ_0), the value of the customer-type information becomes higher than that of the customer preference information. Thus, in such a case, a retailer must focus on ascertaining customer type. Second, when a retailer has already correctly ascertained customer type and the customer has already sufficiently determined the attributes of a product of interest, the retailer must focus more on accurately ascertaining the customer's preference than their type.

Next, Lemma 2 is the sensitivity analysis of the value of $F(\tau_0)$ in relation to several key parameters. Lemma 2 demonstrates how the value of information is influenced by environmental change.

Lemma 2.

Kurata	Customer type or customer preference

If the value of parameters δ, r, k_S, or m_U increases, the probability that the maximum profit is greater in Scenario 2 than in Scenario 3 also increases. Thus, the value of accurately identifying customer type becomes greater than the value of ascertaining their preference.

Proof. This result can be easily obtained using the function $\frac{m_U}{2(\alpha-\gamma)} - \frac{r\alpha^2}{4k_S(\alpha-\gamma)^2}$ presented in Proposition 2.

Here, the parameters δ and k_S can be interpreted as service quality variation and cost of service provision, respectively, both of which denote the negative impact of inadequate personal service provision. In contrast, the parameters r and m_U represent the brand value of a store as having a high retail price, and a high reservation price are deemed a sign of high brand reputation. Lemma 2 states that if a retailer is a luxury product store with a high brand reputation, Scenario 2 will tend to yield higher profits than Scenario 3; i.e., the value of identifying customer type is higher than that of ascertaining customer preference. On the contrary, if a store is positioned as part of a large discount store chain, the operations of which are highly efficient and the brand image is moderate, our finding suggests that Scenario 3 will tend to yield more profit than Scenario 2; i.e., ascertaining customer preference would be more valuable than identifying their type.

IT OR IE: DISCUSSION ON MANAGERIAL DECISION MAKING

This section discusses which method is more suitable for obtaining the required information: sophisticated and expensive IT or a simple and inexpensive IE-based tool. Modern technology is widely applied in marketing to identify consumers' unexpressed preferences and tastes. For example, one trend in the 4.0 era is in-store monitoring that often incorporates various sensor-based strategies, such as beacon, near-field communication, and RFID (Kotler et al., 2017). RFID technology that can trace the in-store movement of a shopper is applied to ascertain their purchasing behavior (Hui et al. 2009; Hui et al. 2013a; Hui et al. 2013b).

However, traditional IE methods continue to be applied in the manufacturing and service sectors to improve efficiency and prevent unnecessary failures, including the pokayoke, 5S methods, and as the seven tools of quality control. Pokayoke is a simple mechanism applied to equipment and processes to prevent careless mistakes. While originally developed in the manufacturing sector, pokayoke is now widely used in service industries (Chase & Stewart, 1994). For example, Urban Research, a casual apparel chain in Japan, has devised a "no-sales-talk" in-store shopping bag, which has a specific design different from a normal shopping bag. Therefore, this special type of shopping bag can serve as a simple pokayoke-type signal for store employees, informing them that a sales pitch is unwanted. Thus, if a customer chooses to bring such a bag, the sales staff will not talk to them unless the customer asks for help. Thus, customers who do not need or dislike a sales pitch can quietly enjoy their shopping experience.

First, we must ascertain how store managers should allocate their limited budget to obtain shopper information, using either IT or IE, from the perspective of the store's marketing–operations status. As derived from Lemma 2, the statements in Table 2 summarize the issue of IT-or-IR selection from two perspectives: brand status (i.e., marketing perspective) and cost structure (i.e., operations perspective). Here, we interpret high values of r and m_i $(i \in \{P, U\})$ as high brand values and high values of $k_i (i \in \{P, U\})$ and δ as high cost structurers. The statements in Table 2 imply that a store with high brand value, such as a luxury goods store, strongly focuses on customer type (e.g., appropriate selection of customers through approaches

such as a membership system or store positioning that allows consumers to self-select). In contrast, if a store has a moderate-level brand value but efficient operations, investing in IT to ascertain customer preferences (e.g., in-store measurement of a customer's trajectory using RFID) and adopting a simple method (e.g., "no-sales-talk" shopping bags) to determine customer type are appropriate options.

Table 2: IT or IE selection regarding a retailer's managerial status

		How high is the store's brand reputation?	
		Relatively low.	*Relatively high.*
Service operation type	*Relatively high-cost structure*	• IT for identifying customer type • IE for ascertaining customer preference	• IT for identifying customer type • IE for ascertaining customer preference
	Relatively low-cost structure.	• IT for ascertaining customer preference • IE for identifying customer type	• IT for identifying customer type • IE for ascertaining customer preference

Table 3: IT or IE selection regarding a retailer's information availability

		How well does the store know a customer's preference?	
		Relatively poorly	*Relatively well.*
Extent of customer's knowledge of the product	*Relatively high.*	Generally inconclusive	• IT for identifying customer type • IE for ascertaining customer preference
	Relatively low.	• IT for ascertaining customer preference • IE for identifying customer type	Generally inconclusive

Second, we must interpret Proposition 3 in terms of IT or IE selection. Table 3 summarizes how a retailer should choose IT or IE considering two aspects of uncertainty: customer's preference (i.e., the value of δ) and the customer's understanding of the quality and attributes of a product (i.e., the value of τ_0). The statements in Table 3 imply that if a store already has a sufficient level of customer information and the customer already understands the product well, it is appropriate to use IT to determine the customer's type and IE to ascertain their service preference. A large

grocery store chain can be used as an example here as this type of chain will likely understand its customer base well and the customers will be familiar with the commodities that the chain sells. As such, the adoption of RFID in grocery stores to ascertain the shoppers' unexpressed behavior (Hui et al., 2009, 2013a, 2013b) is consistent with our analytically derived managerial decisions.

Finally, we can summarize our managerial decision-making as follows:

1. Any inaccuracy in customer-type identification may reduce the quality of the personal service, which will negatively affect the purchase intension of "undetermined" customers, such as showroomers. In contrast, inaccuracy in service preference identification will not affect the service quality.
2. Sales related to "predetermined" customers such as webroomers are insensitive to the inaccuracy of customer-type identification and the uncertainty in customer service preference.

CONCLUDING REMARKS

Personal customer service on the shop floor is a powerful tool for increasing sales. Therein, sales staff interacts with customers to alleviate the latter's uncertainty regarding the quality and attributes of a product that they may be unsure of purchasing. However, some customers will have already decided which product to buy, meaning a sales pitch will be useless or even annoying. Thus, sales personnel must deal with two challenges: identifying the customer type and ascertaining what level of personal service is required for each type of customer. This paper analyzed the values of two forms of information, customer type and customer preference, for personal consumer service while assuming the existence of two main customer groups: those who have already decided which product to buy before visiting the store (e.g., *webroomers*) and those that have not yet decided and are unsure of a product's quality (e.g., *showroomers*).

We developed a mathematical model and compared the profitability among four different scenarios based on the aforementioned information types. Here, we found that if a retailer has a high brand value and a high-cost structure, correctly identifying customer type will be more profitable than helping the customer with details regarding the product. We also provided managerial recommendations regarding which methodology is suitable for obtaining each type of information, be it a sophisticated and expensive IT system or a simple and inexpensive IE tool. Our IT-or-IE discussion will provide support retailers in making resource-allocation decisions.

Several directions are available for further research. First, for tractability, we simplified our model by assuming that the "predetermined" customers prefer minimum service provision as they are already aware of the product's attributes. Although in practice, relaxing these assumptions would serve better, it would require various numerical methods owing to the complexity of the analysis. Second, we only considered two main methods for obtaining the information, i.e., expensive IT and simple IT; it would perhaps be interesting to consider various combinations of these two different methods. Third, given that our analysis was static, long-term effects related to the quality of service could be analyzed by formulating a specific dynamic model.

ACKNOWLEDGEMENTS

The authors are grateful for all comments and suggestions from participants at the POMS 2019 annual conference, the JIMA 2019 Spring conference, and the DSI 2019 annual meeting. This research is supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B) 18H00903.

REFERENCES

Ba, S., Stallaert, J., & Zhongju Zhang. (2010). Balancing IT with the human touch: Optimal investment in IT-based customer service. *Information Systems Research*, *21*(3), Sep, 423-442. https://doi.org/10.1287/isre.1100.0282

Bellos, I., & Kavadias, S. (2019). When should customers control service delivery? Implications for service design. *Production & Operations Management*, *28*(4), Apr, 890-907. https://doi.org/10.1111/poms.12956

Chase, R.B., & Stewart, D.M. (1994). Make your service fail-safe. *MIT Sloan Management Review*, 35(3), Spring, 35-44.

Das, G. (2014). Linkages of retailer personality, perceived quality and purchase intention with retailer loyalty: A study of Indian non-food retailing. *Journal of Retailing & Consumer Services*, *21*(3), May, 407–414. https://doi.org/10.1016/j.jretconser.2013.11.001

Davidow, W. H., & Uttal, B. (1989). Service companies: Focus or falter. *Harvard Business Review*, 67(4), Jul/Aug, 77-85.

Desai, P., Kekre, S., Radhakrishnan, S., & Srinivasan, K. (2001). Product differentiation and commonality in design: Balancing revenue and cost drivers. *Management Science*, 47(1), Jan, 37-51. https://doi.org/10.1287/mnsc.47.1.37.10672

Diaz-Martin, A. M., Iglesias, V., Vazquez, R., & Ruiz, A. V. (2000). The use of quality expectations to segment a service market. *Journal of Services Marketing*, *14*(2/3), 132-146.

Esmaeilpour, M., Mohamadi, Z., & Rajabi, A. (2016). Effect of Dimensions of Service Quality on the Brand Equity in the Fast Food Industry. *Studies in Business & Economics*, 11(3), Dec, 30–46. https://doi.org/10.1515/sbe-2016-0033

Fassnacht, M., Beatty, S. E., & Szajna, M. (2019). Combating the negative effects of showrooming: Successful salesperson tactics for converting showrooming into buyers. *Journal of Business Research*, 102, Sep, 131-139. https://doi.org/10.1016/j.jbusres.2019.05.020

Fernández, N. V., Pérez, M. J. S., & Vázquez-Casielles, R. (2018). Webroomers versus showroomers: Are they the same? *Journal of Business Research*, 92, Nov, 300-320. https://doi.org/10.1016/j.jbusres.2018.08.004

Fuentes-Blasco M, Moliner-Velázquez B, Servera-Francés D, & Gil-Saura I. (2017). Role of marketing and technological innovation on store equity, satisfaction, and word-of-mouth in retailing. *Journal of Product & Brand Management*, 26(6), 650-666. doi:10.1108/JPBM-07-2016-1279.

Gao, F., & Su, X. (2017). Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science*, *63*(8), Aug, 2478-2492. https://doi.org/10.1287/mnsc.2016.2473

Gilliam, D. V., & Zablah, A. R. (2013). Storytelling during retail sales encounters. *Journal of Retailing and Consumer Services*, 20, Sep, 488-494.

Ha, A. Y., Tong, S., & Zhang, H. (2011). Sharing demand information in competing supply chains with production diseconomies. *Management Science*, *57*(3), Mar, 566-581. https://doi.org/10.1287/mnsc.1100.1295

Holmes, E. (2012, June 28). Leave me alone, I'm shopping --- Top cosmetics brand tells sales staff to back off; 'Our product is here to play with. *The Wall Street Journal*, June 28, 2012.

Hui, S.K., Bradlow, E.T., & Fader, P.S. (2009). Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of Consumer Research*, 36(3), Oct, 478-493. https://doi.org/10.1086/599046

Hui, S.K., Yanliu, H., Suher, J., & Inman, J.J. (2013a). Deconstructing the “first moment of truth”: Understanding unplanned consideration and purchase conversion using instore video tracking. *Journal of Marketing Research*, 50(4), Aug, 445-462. https://doi.org/10.1509/jmr.12.

Hui, S. K., Inman, J. J., Yanliu, H., & Suher, J. (2013b). The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Journal of Marketing*, 77(2), Mar, 1-16. https://doi.org/10.1509/jm.11.0436.

Lai, F., Griffin, M., & Babin, B. J. (2009). How quality, value, image, and satisfaction create loyalty at a Chinese telecom. *Journal of Business Research*, 62(10), Oct, 980–986. https://doi.org/10.1016/j.jbusres.2008.10.015.

Iyer, G. (1998). Coordinating channels under price and nonprice competition. *Marketing Science*, *17*(4), 338-355. https://doi.org/10.1287/mksc.17.4.338

Jin, M., Li, G., & Cheng, T. C. E. (2018). Buy online and pick up in-store: Design of the service area. *European Journal of Operational Research*, *268*(2), Jul, 613-623. https://doi.org/10.1016/j.ejor.2018.02.002

Kahneman, D., & Tversky, A., (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), Mar, 263-292. https://doi.org/10.2307/1914185.

Kotler, P., Kartajaya, H., & Setiawan, I. (2017). *Marketing 4.0: Moving from traditional to digital*. Hoboken, NJ: John Wiley & Sons.

Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T., (2016). M*arketing Management (2nd ed.)*, Harlow, Essex, UK: Pearson Education.

Kurata, H. & Nam, S. H. (2010), After-sales service competition in a supply chain: Optimization of customer satisfaction level or profit or both?, *International Journal of Production Economics*, Volume 127(1), Sep, 136-146. https://doi.org/10.1016/j.ijpe.2010.05.005.

Kurata, H. & Nam, S. H., (2013), After-sales service competition in a supply chain: Does uncertainty affect the conflict between profit maximization and customer satisfaction?, *International Journal of Production Economics*, 144(1), Jul, 268-280. https://doi.org/10.1016/j.ijpe.2013.02.014

Kurata, H. (2019). Is the information of customer types and preferences to personal selling worth the investment in innovative technology? A modeling approach, *Journal of Retailing and Consumer Services*, 49, Jul, 371-379. https://doi.org/10.1016/j.jretconser.2019.01.020

Layun R. (2017). Top customer care trends to win in the age of the consumer. *CRM Magazine*, 21(10), WP14-WP15. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=125475453&lang=ja&site=ehost-live, May 23, 2020.

Nguyen, N., & LeBlanc, G. (1998). The mediating role of corporate image on customers' retention decisions: an investigation in financial services. *International Journal of Bank Marketing*, *16*(2), Apr, 52-65. https://doi.org/10.1108/02652329810206707

Pierce, J.R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, 39(2), Feb, 313–338. https://doi.org/10.1177/0149206311410060.

PR Newswire (2015, July 8), Lowe's ranks highest in customer satisfaction with appliance retailers; Sales and promotions drive retail sales; Staff and good service have greatest impact on customer satisfaction. *PR Newswire*. Retrieved from https://www.prnewswire.com/news-releases/lowes-ranks-highest-in-customer-satisfaction-with-appliance-retailers-300109183.html, May 23, 2020.

Shi, X., Dong, C., & Cheng, T. C. E. (2018). Does the buy-online-and-pick-up-in-store strategy with pre-orders benefit a retailer with the consideration of returns? *International Journal of Production Economics*, *206*, Dec, 134-145. https://doi.org/10.1016/j.ijpe.2018.09.030

Tsay, A. A., & Agrawal, N. (2000). Channel dynamics under price and service competition. *Manufacturing & Service Operations Management*, *2*(4), Fall, 372-391. https://doi.org/10.1287/msom.2.4.372.12342

DECISION SCIENCES INSTITUTE

Comparison of Suggested and Competing Regression Growth Models for Peppermint Production

S. K. Yadav
Babasaheb Bhimrao Ambedkar University
Email: drskystats@gmail.com

Dinesh K. Sharma
University of Maryland Eastern Shore
Email: dksharma@umes.edu

Julius A. Alade
University of Maryland Eastern Shore
Email:ajalade@umes.edu

A.K. Shukla
Chhatrapati Shahu Ji Maharaj University
Email: alokshukladav@gmail.com

ABSTRACT

In this paper, we have suggested three different regression models for the production of the peppermint yield at block level of Barabanki District of Uttar Pradesh State in INDIA. We have checked the models adequacy on the basis of various measure of goodness of fit. We have compared the suggested models with the competing regression models for yield production. The theoretical efficiencies of the suggested models over the competing models have been verified using the real primary data collected from the Banki block of the Barabaki District of Uttar Pradesh State in India. It has been shown that the suggested models are better than the competing models for the production of peppermint yield.

KEYWORDS: Response variable, Regressor, Regression model, Residual, Coefficient of determination.

INTRODUCTION

Peppermint has many medicinal properties; it has high menthol content and is used as a flavour in tea, ice cream and in many other products. Because of all these derived uses of Peppermint, it is very important to estimate its production very close to the actual one and it is very necessary to find the best predictive model so that we may plan to cultivate it as per the need and demand in the market. If we estimate the production of peppermint very efficiently through the best fitted model, then we can predict and calculate the revenue contribution to district and state as well on the basis of appropriately scribbling out various schemes for district as well as for State. An efficient estimation of actual production of peppermint crop is paramount today because it has very high medicinal value with great potential to contributing to the economic prosperity of the district as well as the state.

Therefore, it becomes very imperative to search for the best fitted model for the production of this crop. This kind of proposed research always seeks the attention of governmental agencies to

support the research activities in order to disseminate wider accessibility of research findings. Without governmental support, an investigator may face various sorts of complexities at official and administrative levels for conducting the research. The dependent variable is the production (Yield) of peppermint oil in kilogram and the auxiliary variable is the area of the field in one Bigha (2529.3 Square Meter). There are 20 Biswa (one Biswa =126.5 Square Meter) in one Bigha. Similar type of works has been conducted by many authors including the latest one by Satoh (2019) and Wen et al. (2019).

LITERATURE REVIEW

Various authors have given different models to establish the most appropriate relationship between area and the production of different crops. The considerable work by different authors is presented in Table 1.

Table 1: Different growth models suggested by various authors

Year	Authors	Journal
1938	Smith	Journal of Agricultural Science
1988	Haq *et al.*	Annals of Agricultural Research
1989	Young and Ord	International Journal of Forecasting
2010	Al-Kassie	Agriculture and Biology Journal of North America
2011	Kumar *et al.*	Agricultural Economics Research Review
2015	Zhao *et al.*	Open Journal of Animal Sciences
2017	Scarneciu *et al.*	Pakistan Journal of Medical Sciences
2018	Kaplan *et al.*	Journal of Applied Animal Research
2019	Wen *et al.*	Journal of Applied Animal Research

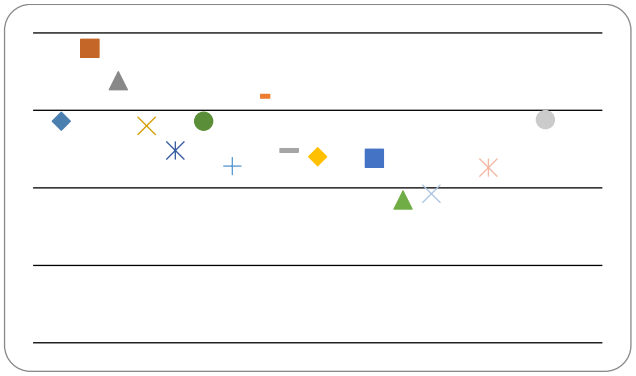
MATERIALS AND METHODS

A data set has been used for the study, collected from the Banki blocks of Barabanki district with 15 farmers. These farmers were selected using simple random sampling scheme. The data on yield (Kilo Gram per Biswa) as the dependent variable (Y). The data on the area (Biswa) of the field as the independent or the auxiliary variable (X) were collected at the time of harvesting. The data set collected is presented in Table1. The scatter plots of the data set is also presented in Figure 1.

Table 1: The data Collected from Banki Block of Barabanki District of Uttar Pradesh

X	5	10	15	20	25	30	35	40
Y	0.143	0.19	0.169	0.14	0.124	0.143	0.114	0.159
X	45	50	60	65	70	80	90	
Y	0.124	0.12	0.119	0.092	0.096	0.113	0.144	

Figure 1: Scatter Plot along with the Fitted line with (Y) on the Y-Axis and (X) on the X-Axis



From Figure 1, we can see that the fitting of a straight line seems not to be of the best fitted model or curve. The scatter plot looks like nonlinear. Thus, we search for the most appropriate nonlinear growth model for the best estimation and prediction of the yield of the peppermint at block level of Barabanki district of Uttar Pradesh in India for making the best policies.

COMPETING GROWTH MODELS

The competing growth models suggested by Smith (1938) and Haque et al. (1988) are the best fitted models in almost all areas of applications. The growth models suggested by Smith (1938), and Haque et al. (1988) are as follows,

$$Y = a\ b^{X} \quad (1)$$

$$Y = a\ X^{b} \quad (2)$$

$$Y = a\ \exp(-b\ X) \quad (3)$$

$$Y = a + b\ c^{X} \quad (4)$$

where, a, b and c are the parameters of the above growth models.

SUGGESTED GROWTH MODELS

Motivated by the above authors and keeping in view the shape of the scatter plot, we have suggested the following three growth models for the better estimation and prediction of peppermint yield.

$$Y = a + b\ X + c\ d^{X} \quad (5)$$

$$Y = a + \frac{b}{X} + c\ d^{X} \quad (6)$$

$$Y = a + \frac{b}{X^{2}} + c\ d^{X} \quad (7)$$

where, a, b, c and d are the parameters of the suggested growth models.

FITTING OF GROWTH MODELS

Draper and Smith (1998) have broadly classified these growth models into two classes. First class contains the intrinsically linear-nonlinear models, while the other contains purely nonlinear models. A model is called intrinsically linear-nonlinear model if by means of any transformation, it can be transformed into linear model. A model is called linear in its parameters if all of its partial derivatives with respect to all parameters of the model are free from the parameters otherwise it is nonlinear. According to Draper and Smith (1998), the models (1), (2) and (3) are the intrinsically linear-nonlinear as they can be transformed into linear models through logarithmic transformation. The model (4) is purely nonlinear model as it cannot be transformed into a linear model. Similarly, the suggested models (5), (6) and (7) are purely nonlinear models. Generally growth models are nonlinear models. The ordinary least square (OLS) method is applied for the estimation of the parameters of the linear models but it is not an appropriate estimation method for nonlinear models. Thus the parameters of the models (1), (2) and (3) can be estimated using OLS method but not for the models (4), (5), (6) and (7). There are many methods for the estimation of parameters of the nonlinear models such as nonlinear least square, steepest descent method, method of three selected points and Levenberg-Marquardt's method. Among the above methods of estimation of parameters of nonlinear models, Levenberg-Marquardt's method is the best as its estimates hold almost all properties of the good estimators. Thus the parameters of the models (4), (5), (6) and (7) are estimated using Levenberg-Marquardt's method.

GOODNESS OF FIT OF DIFFERENT GROWTH MODELS

There are various measures of Goodness of Fit for the growth models. Different measures of goodness of fit, their formulae and descriptions are given in the following Table 3.

Table 3: Different Measures of Goodness of Fit for the Models

S.No.	Measure of Goodness of Fit
1.	Coefficient of Determination - R^2
2.	Adjusted Coefficient of Determination - R^2_{Adj}
3.	Residual Mean Square- s^2
4.	Mean Absolute Error (M.A.E)
5.	Mean Absolute Percentage Error (M.A.P.E)
6.	Test for auto correlation of errors
7.	Test for independence of errors
8.	Test for normality

EMPIRICAL STUDY

To check the adequacies of the competing and the suggested models, we have used the data set of primary data collected from Banki block of Barabanki district of Uttar Pradesh in India on the yield of the peppermint crop. The estimates of the parameters and various measures of goodness of fit for the competing and the suggested estimators are presented in Table4 and Table 5 respectively.

Table 4: Parameter Estimates for Various Models				
Model	a	b	c	d
Model (1)	0.1608	0.9953	-	-
Model (2)	0.2159	-0.1406	-	-
Model (3)	0.1629	0.0050	-	-
Model (4)	0.1084	0.0718	0.9681	-
Model (5)	-1.6335	0.0058	1.8141	0.9956
Model (6)	0.1309	-1.0780	0.3323	0.9285
Model (7)	0.1181	-3.5015	0.2651	0.9099

Table 5: Goodness of Fit of Models & Residuals Analysis							
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)
R^2	0.3787	0.3807	0.3802	0.4446	0.4973	0.6029	0.6146
R^2_{Adj}	0.3309	0.3331	0.3325	0.3521	0.3602	0.4946	0.5095
s^2	0.000474	0.000472	0.000473	0.000459	0.000453	0.000358	0.000347
M.A.E.	0.0166	0.0166	0.0166	0.0154	0.0150	0.0124	0.0116
M.A.P.E	12.7083	12.5857	12.6633	11.7003	11.4052	9.9358	9.4138
DW#	1.7233	1.8638	1.7343	1.9236	2.0275	1.7833	1.7945
R*	0. 556 (0.578)	0. 556 (0.578)	0. 556 (0.578)	0. 556 (0.578)	0. 000 (1.000)	0. 000 (1.000)	0. 000 (1.000)
SW^	0. 908 (0. 128)	0. 921 (0. 203)	0.910 (0.134)	0.921 (0.199)	0.963 (0.749)	0.960 (0.692)	0.946 (0.465)

Durbin & Watson test values, * Run test values, ^ Shapiro-Wilk test values, and the p-values (in parentheses).

RESULTS AND DISCUSSION

From Table 5, it can be easily observed that the value of R^2 for the models in competition with the suggested models ranges from [0.3787 0.4446], while for the proposed model it ranges from [0.4973 0.6146] respectively. Similarly the values of R^2_{Adj} for the models in competition ranges from [0.3309 0.3521] and for the proposed models it ranges from [0.3602 0.5095] respectively. The value of the residual sum of squares s^2 for the competing models ranges from [0.000459 0.000474] while for the suggested models, it ranges from [0.000347 0.000453] respectively. The value of M.A.E. ranges from [0.0154 0.0166] for the competing models while for the proposed models it ranges from [0.0116 0.0150] respectively. The value of M.A.P.E. ranges from [11.7003 12.7083] for the competing models while it ranges from [9.4138 11.4052] for the proposed models respectively. Other measures given in Table 5, Durbin-Watson statistic, Run test and Shapiro-Wilk test are also better for proposed models in comparison to the competing models.

CONCLUSION

In this paper, we suggested three new growth models for the estimation and prediction of the yield of the peppermint crop. The statistical measures of the suggested models are studied. The suggested models are compared with the competing growth models on the basis of these fitting measures of R^2, R^2_{Adj}, s^2, M.A.E., M.A.P.E., Durbin-Watson statistic, Run test and Shapiro-Wilk test. It is evident from Table 5 that the above measures are better for the suggested models in comparison to the competing models. As the suggested models of goodness of fit is better than the competing models, they will predict better than the competing models. Thus the suggested models will be more appropriate for better prediction of peppermint yield prediction and may be used for necessary policy prescription for peppermint crop production for good yield and market demand as well as economic benefit to the Barabanki district of India.

Acknowledgement

The first author Dr. S. K. Yadav (PI) is thankful to Council of Science and Technology, Uttar Pradesh, Lucknow for financial assistance in carrying out this research under the Project No. CST/4052.

REFERENCES

Al-Kassie, G.A.M. (2010). The role of peppermint (Mentha piperita) on performance in broiler diets. *Agriculture and Biology Journal of North America*, 1, 5, 1009-1013.

Draper, N.R., & Smith, H. (1998). *Applied regression analysis*, 3rd Ed., John Wiley & Sons.

Haque H. N., Azad N. K., Jha R. N., & Singh S. N. (1988). Optimum Size and Shape of Plots for Wheat, *Annals of Agricultural Research*, 9, 2, 165-170.

Kaplan, S., & Gurcan, E.K. (2018). Comparison of growth curves using non-linear regression function in Japanese quail. *Journal of Applied Animal Research*, 46, 1, 112-117.

Kumar, S., Suresh, R., Singh, V., & Singh, A. K.(2011). Economic Analysis of Menthol Mint Cultivation in Uttar Pradesh: A Case Study of Barabanki District. *Agricultural Economics Research Review*, 24, Dec-July.

Shukla, A.K., & Yadav, S.K. (2020). New linear model for optimal cluster size in cluster sampling, *Statistics in Transition New Series*, 21, 2, 189-200.

(A complete list of references is available upon request)

DECISION SCIENCES INSTITUTE
Exploring The Most Creative and Innovative Business Leaders at Top 100 U.S. Companies

Tuncay Bayrak
Western New England University
College of Business
Springfield, MA, USA.
tbayrak@wne.edu

ABSTRACT

Numerous studies have repeatedly demonstrated a positive relationship between innovative leadership and gaining and sustaining long term competitive advantages. In addition, as various business and organizational problems call for innovative and creative minds, an increasing number of companies are turning to creative minds for innovation. Creative and innovative leadership has the potential to inspire employees, and transform businesses, communities and organizations. Therefore, innovative and creative leadership is arguably one of the most important forces that drive the 21th century business world.

KEYWORDS: Innovative leader, Creativity, Leadership, CEO.

INTRODUCTION

Innovative and creative leadership may be vital to any organization's survival in an interconnected and highly complex global business world. To stay competitive and survive, one would agree that today's organizations need creative leaders who can craft innovative solutions to challenging business problems. Thus, it may be argued that being innovative and creative is one of the most important characteristics of modern business leaders.

Creative and innovative leadership has the potential to inspire trust, and transform businesses, communities, and organizations. Innovative and creative leaders create visions, coach people, and help navigate ambiguity in a fast-changing landscape of business world. They give their businesses leading edge, and foster success in a company. Further, innovative leaders address individual and organizational development needs simultaneously while building collaborative relationships and tailoring capacities and solutions to various business problems. Therefore, innovative and creative leadership is arguable one of the most important forces that drive the 21th century business world.

Successful, innovative, and creative minds often rely on their abilities to navigate ambiguity and to effectively address challenges facing their organizations. With their key capabilities, creative minds, leadership, and problem solving skills, innovative leaders successfully employ their visons and imaginations to generate new innovations, drive productivity, and solve complex business problems.

Innovative and creative leaders certainly possess particular qualities and characteristics that set them apart. They, for instance, have the abilities to turn risks, challenges, and ambiguity into growth opportunities. They can inspire and motivate people, generate ideas, turn the ideas into reality, and create values that trigger innovations. They focus their creative talents on

innovations, creating values and contributing to their societies through becoming change agents. Innovative leaders are often visionaries capable of inspiring others. To achieve their goals, they develop organizations that promote creative thinking. As Steve Jobs puts it "innovation distinguishes between a leader and a follower." Therefore, one would agree that in order for leaders to lead in a world of accelerating change, they need to think creatively and innovate.

While there are various traits that are essential to the mindset of innovative leaders, this study examines profiles of America's most innovative and creative leaders at top US companies as ranked by the Forbes magazine.

BACKGROUND

Innovative and creative leaders have left their mark, and pioneered and fueled new industries, products, and services. Classic examples of innovative leaders include Steve Jobs, Elon Musk, and Bill Gates. These leaders demonstrated the importance of innovative leaders within organizations on numerous occasions. Due to the fact that they possess a crucial set of qualities, companies increasingly understand the importance of innovative and creative minds. Companies that are run by innovative and creative leaders often gain and sustain their long-range competitive advantages. Not surprisingly, an increasing number of companies are turning to innovative and creative minds for innovation and creativity.

Various business and organizational problems call for innovative and creative minds. For instance, Horth and Buchner (2004) argue that making sense of complex challenges requires innovative leadership. Although traditional problem-solving tools, skills, and practices have been employed to solve well-defined business problems that involve no ambiguity, applying traditional business thinking to organizational challenges may not work as the direction may be unclear, and the problems to be solved may involve ambiguity. Therefore, shaping order out of ambiguity and reframing the challenge may call for an innovative and creative leadership.

Numerous studies have demonstrated a positive relationship between innovative leadership and creativity and innovation (Hughes et al., 2018). Moreover, similar studies undertaken in the leadership domain suggest that the terms "Innovative leader" and "creative thinker" have even been used interchangeably. The literature makes references to the term "creativity" as one of the most crucial traits of innovative leaders. For instance, IBM surveyed 1500 corporate heads and public sector leaders on what drives them in managing their companies in today's world. The study concluded that creativity was ranked the most important leadership quality for success in business, outweighing integrity and global thinking (IBM, 2010). Desjardins (2018) draws a similar conclusion arguing that creativity is among the top three skills needed to thrive in 2020, along with other leadership capabilities such as complex problem solving and critical thinking skills. Mumford et al., (2003) agree and assert that to direct innovative efforts, leaders must possess creative thinking skills.

Sternberg et al., (2003) claim creative leadership can be of three general kinds—leadership that accepts existing ways of doing things, leadership that challenges existing ways of doing things, and leadership that synthesizes different existing ways of doing things. However, regardless of their leadership styles, Sternberg (2008) argues that creative leaders generate novel and useful ideas, possess analytical intelligence to ascertain whether these ideas are good ideas, employ

practical intelligence to make ideas functional, and draw on their wisdom to ensure that the actions of the leader seek a common good.

Basadur (2004) maintains that individuals, teams, and organizations differ in their creative problem solving styles. How leaders manage these styles can have a significant effect on performance. Therefore, innovative and creative leaders of the 21st century, as pointed out by Basadur (2004), must employ what is called “applied creativity that involves coaching individuals and teams to coordinate and integrate their differing styles to continuously discover and define new problems, solve those problems, and implement the new solutions. Similar observations were echoed by Mink et al., (1989) who maintain that creative leaders consistently frame difficult interpersonal interactions as problems to be solved rather than as problems to be explored.

While various creative and innovative leadership models have been put forth, Sternberg (2008) developed the most commonly employed creative leadership model called WICS; an acronym for Wisdom, Intelligence, and Creativity, Synthesized. The WICS theory of leadership states that good and effective leaders possess a crucial set of developed characteristics. Building on the creative leadership model propped by Sternberg (2008), a number of studies focused on various qualities and characteristics that innovative and creative leaders are known to possess. For instance, Zenger and Folkman (2014) suggest that these leaders display excellent strategic vision, are persuasive, excel at setting goals, inspire and motivate through action, and create a climate of trust. Similarly, Marco (2018) claims innovative leaders take calculated risks, excite minds, look for problems, and collaborate to innovate. A study by Hill (2016) points out that innovative leaders learn by discovery, create better solutions, integrate diverse ideas successfully, and have a capacity for collaborative problem solving. The author further maintains that innovative leaders see themselves as social architects who create communities in which others are willing and able to innovate.

Some other factors pertaining to creative leaders have been examined. For instance, Mueller et al., (2011) points out that the negative association between expressing creative ideas and leadership potential underscores an important bias against selecting effective leaders. Jaussi and Dionne (2003) examines the relationship between a leader's unconventional behavior and followers' creative performance at both the individual and group level in an experimental setting. Their results indicate that unconventional leader significantly interacts with follower perceptions of the leader as a role model for creativity to explain follower creativity.

The aforementioned studies suggest that because of their agility, positive outlook, and abilities to transform businesses, and turn ambiguity into growth opportunities, creative and innovative leadership may be more important than ever before.

RESEARCH METHOD

This study is based on the data compiled and published by the Forbes Magazine in 2019 (Forbes, 2019). As explained on its website, it teamed up with business school professors Jeff Dyer, Nathan Furr, Mike Hendron, and consultant Curtis Lefrandth to rank America’s 100 most creative and successful business minds at top U.S. companies. They identified the most innovative CEOs based on four essential leadership qualities of top leaders and CEOs: media reputation for innovation, social connections, track record for value creation, and investor expectations for value creation.

As they pointed out, their sample of leaders includes the founders or CEOs or CEOs who have become chairman of the board within the past year. They further point out that this list is currently focused on U.S. companies because measures of media reputation for innovation and social capital were only available for leaders of U.S. companies. CEOs of firms in regulated industries such as utilities and defense contractors were not included. They did not rank CEOs of companies that have been acquired or who are CEOs of subsidiaries or divisions of larger companies. In addition, they co-ranked leaders of two firms because they were in visible top management positions. For instance, the founders of Google CEO Larry Page and President Sergei Brin were co-ranked. Therefore, the final list included a total of 102 innovative leaders.

Having retrieved the list published by the Forbes Magazine, we browsed the Internet to collect and compile as much information as possible on each CEO. We particularly searched for such information on each leader as their education level, their birth places, the colleges they attended, their major in college, and whether or not they had master's and Ph.D. degrees.

In the following section, we summarize and report our findings on the most innovative leaders at top 100 U.S. companies.

DATA ANALYSIS AND RESULTS

We fist explored the most successful business minds with respect to their ages across various sectors. As seen, across different sectors the most innovative leaders had an average age of 58. While the youngest CEO (35) works in the technology sector, the oldest one (82) leads a company in the health care sector. As summarized in table 1, CEOs in the Motor Vehicles & Parts sector have the lowest average age (48), whereas CEOs who are in charge of companies in the Wholesalers sector seem to have the highest average age.

Table 1: CEOs' Average Age by Sector

Sector	Average of Age	Max of Age	Min of Age
Aerospace & Defense	52	52	52
Apparel	64	64	64
Business Services	58	62	53
Chemicals	63	74	55
Engineering & Construction	62	62	62
Financials	55	55	55
Food, Beverages & Tobacco	64	69	61
Health Care	61	82	49
Hotels, Restaurants & Leisure	61	61	61
Household Products	58	62	55
Industrials	57	60	55
Materials	53	56	50
Media	62	68	56
Motor Vehicles & Parts	48	48	48
Retailing	58	67	54
Technology	55	67	35
Telecommunications	63	65	60

Wholesalers	70	70	70
Hotels, Restaurants & Leisure	52	52	52
Transportation	60	60	60
Grand Total	**58**	**82**	**35**

In this study, we speculated that the most innovative CEOs must have graduated from globally recognized highly reputable universities. Thus, we examined them with respect to the universities they graduated from. As seen, several CEOs have received their undergrad degrees from highly reputable universities such as Princeton University (5), Stanford University (3), Harvard University (2), and Dartmouth College (2). Overall, the overwhelming majority of the most innovative and creative leaders ranked by the Forbes magazine have received their undergrad degrees from less reputable universities. Hence, it may be concluded that to be an innovative and creative leader one does not need to be granted a degree by a top-ranked university.

Table 2: Universities CEOs attended	
University	Undergrad School
Princeton University	5
Stanford University	3
University of Michigan	3
University of Maryland	2
Bucknell University	2
University of Southern California	2
College of the Holy Cross	2
University of the Witwatersrand	2
Dartmouth College	2
Union College	2
Harvard University	2
Michigan State University	2

Similarly, we examined the CEOs in terms of their undergrad degrees they earned. As tabulated in table 3, a considerable number of the most innovative leaders have an undergrad degree in Business Administration (18), and Economics (16). A number of CEOs running various companies have a degree in mechanical engineering (9), electrical engineering (8), and chemical engineering (3). These results suggest that the most innovative leaders have diverse educational backgrounds in various domains.

Table 3: CEOs and their undergrad degrees	
Undergrad Major	Major
Business Administration	18
Economics	16
Mechanical Engineering	9
Electrical Engineering	8
Accounting	5

Engineering	4
Chemical Engineering	3
Medicine	2
Math	2
Political Science	2
Law	2
Finance	2

Having surveyed their master's degrees, it seems a total of 53 CEOs have a master's degree. Of those, 39 have an MBA degree. Looking at the institutions granted their master's degrees, we see that University of Pennsylvania (Wharton School of Business), and Harvard Business School are the top two higher education institutions from which most CEOs obtained their MBA degrees. Some other reputable universities such as Stanford, Duke, and Northwestern Universities made the list as well. It may be concluded that more than half of Forbes Magazine's most innovate CEOs are running their companies with a master's degree. Since 39 of the most innovative leaders have an MBA's degree, one might be tempted to suggest that there is a positive correlation between being an innovative leader and holding an MBA degree.

Table 4: CEOs and Master's degrees	
Degree Granting Institutions	MBA
University of Pennsylvania	7
Harvard University	5
Stanford University	2
Duke University	2
Northwestern University	2

An interesting observation was made after analyzing the countries in which the most innovative CEOs were born. While the overwhelming majority of the most innovative leaders were born in the USA (80), several CEOs were born in other countries such as India (4), South Africa (3), and Australia (2). The small number of CEOs who were worn abroad may seem negligible, however; these CEOs are in charge of some of the largest Fortune 500 companies such as Microsoft, Adobe, MasterCard, and Tesla.

Table 5: CEOs and country of birth	
Country of Birth	Number of CEOs
USA	80
India	4
South Africa	3
Australia	2

Having surveyed the countries in which they were born, we examined the USA born CEOs with respect to the state in which they were born. As tabulated in table 6, 9 CEOs were born in New York, 4 were born in California, and 3 were born in Illinois.

Table 6: CEOs and place of birth	
State	City of Birth
New York	9
California	4
Illinois	3
Pennsylvania	2
Hyderabad	2
Michigan	2

We further analyzed the data to find out which sector is home to the highest number of innovative leaders. As summarized in table 7, Technology sector is home to the highest number of innovative minds (29), followed by the Health Care sector with 23 CEOs. Although it appears that innovative and creative leadership is sector specific, a more detailed and comprehensive study should be undertaken to draw such a conclusion.

Table 7: Sectors and CEOs	
Sector	Sector
Technology	29
Health Care	23
Retailing	8
Business Services	8
Household Products	5
Industrials	4
Chemicals	4
Materials	3
Food, Beverages & Tobacco	3
Media	3
Hotels, Restaurants & Leisure	2
Telecommunications	2

We carried out a similar analysis to find out which industry is home to the highest number of CEOs. As summarized in table 8, the Medical Products and Equipment industry is home to the highest number of CEOs, followed by the Pharmaceuticals, and Computer Software industries, each with 8 CEOs. A similar conclusion may be drawn here since the top four industries are home to a total of 32 CEOs.

Table 8. Industry and CEOs	
Industry	Number of CEOs
Medical Products and Equipment	10
Pharmaceuticals	8
Computer Software	8
Semiconductors and Other Electronic Components	6
Specialty Retailers: Other	4
Chemicals	4

Internet Services and Retailing	4
Financial Data Services	4
Health Care: Pharmacy and Other Services	4
Information Technology	3
Household and Personal Products	3
Scientific, Photographic and Control Equipment	3
Entertainment	3
Industrial Machinery	3
Specialty Retailers: Apparel	2
Network and Other Communications Equipment	2
Food Services	2
Building Materials, Glass	2
Diversified Outsourcing Services	2
Telecommunications	2
Beverages	2
Internet Services	2

DISCUSSIONS AND CONCLUSIONS

Innovative leadership is about making sense of complex challenges, setting clear directions, navigating ambiguity, and driving results. Because of their abilities to transform businesses and turn uncertainties and challenges into growth opportunities, creative and innovative leadership may be more important than ever before. Particularly in certain industries where technological change is drastic, innovative leadership may be the most crucial trait.

In this study, we examined America's most innovative CEOs at top US companies ranked by the Forbes magazine. While we explored the most innovative leaders using various variables of interest such as their age, industries, and sectors, we're particularly interested in their education level, and their majors in college.

Statistically speaking, although our sample was small, our analysis yielded a number of important observations. For instance, it appears creative leadership is sector and industry specific. Out of 102 CEOs, a total of 52 CEOs lead various companies in the technology and health care sectors. With respect to their education, 53 CEOs had a master's degree out of which 39 had an MBA degree. One might even be tempted to suggest that one has to have an MBA degree to be recognized as an innovative and creative leader. In addition, a considerable number of CEOs had an undergrad degrees in Business Administration (18), and Economics (16).

As pointed out before, this study made use of a small sample. Future studies should employ a larger sample to draw more general results.

REFERENCES

Basadur, M. (2004). Leading others to think innovatively together: creative leadership, *The Leadership Quarterly*, *15* (1), 103-121.

Desjardins, J. (2018). 10 skills you will need to survive the rise of automation, Retrieved October 12, 2019 from https://www.weforum.org/agenda/2018/07/the-skills-needed-to-survive-the-robot-invasion-of-the-workplace

Forbes (2019). Retrieved October 17, 2019 from https://www.forbes.com/lists/innovative-leaders/#5699cf1e26aa.

Hill, L. A. (2016). Is This what it takes to be an innovative leader? Retrieved October 12, 2019, from https://www.weforum.org/agenda/2016/01/is-this-what-it-takes-to-be-an-innovative-leader.

Horth, D., and Buchner, D. (2004). Innovation leadership, how to use innovation to lead effectively, work collaboratively, and drive results. Retrieved October 12, 2019 from https://www.ccl.org/wp-content/uploads/2015/04/InnovationLeadership.pdf

Hughes, D.J., Lee, Al., Tian, A., Newman, Al., and Legood, A. (2018). Leadership, creativity, and innovation: a critical review and practical recommendations, *The Leadership Quarterly*, *29* (5), 549-569.

IBM. (2010). IBM 2010 global CEO study: creativity selected as most crucial factor for future success, Retrieved October 15, 2019, from https://www-03.ibm.com/press/us/en/pressrelease/31670.wss

Jaussi, K. S and Dionne, S. D. (2003). Leading for creativity: the role of unconventional leader behavior, *The Leadership Quarterly*, *14* (4–5), 475-498.

Marco, I. (2018). Five tips to become a more innovative leader. Retrieved October 12, 2019. From https://www.business.com/articles/become-innovative-leader.

Mink, O., Rogers, R., and Watkins, K., (1989). Creative leadership: discovering paradoxes of innovation and risk, *Contemporary Educational Psychology*, *14* (3), 228-240.

Mueller, J.S., Goncalo, J.A., and Kamdar, D. (2011). Recognizing creative leadership: can creative idea expression negatively relate to perceptions of leadership potential? *Journal of Experimental Social Psychology*, *47* (2), 494-498.

Mumford, M. D., Connelly, S., and Gaddis B. (2003). How creative leaders think: experimental findings and cases, *The Leadership Quarterly*, *14* (4–5), 411-432.

Sternberg, R. J., Kaufman, J. C., and Pretz J.E. (2003). A propulsion model of creative leadership. *The Leadership Quarterly, 14* (4–5), 455-473.

Sternberg, R. J. (2008). The WICS approach to leadership: stories of leadership and the structures and processes that support them. *The Leadership Quarterly*, *19* (3), 360-371.

Zenger, J., and Folkman, J. (2014). Research: 10 traits of innovative leaders. Retrieved October 12, 2019 from https://hbr.org/2014/12/research-10-traits-of-innovative-leaders.

DECISION SCIENCES INSTITUTE
Examining Fortune Global 100 Companies and Profiles of Their Chief Executive Officers

Tuncay Bayrak
Western New England University
College of Business
Springfield, MA, USA.
tbayrak@wne.edu

ABSTRACT

Chief Executive Officers (CEOs) are challenged to think strategically about the business and develop solutions to complex business problems. Therefore, CEOs must be equipped with the right set of skills to tackle various challenges that may arise in a highly complex global business world. Competent CEOs have been seen as fundamental to the overall well-being of a company. In this study, we examine Fortune Global 100 companies with a focus on the education levels of their CEOs. Fortune Global 100 companies examined in this study are scattered across four different geographic regions of North America, South America, Europe, and Asia.

KEYWORDS: Fortune Global 100, CEOs, Leadership, Management.

INTRODUCTION

Effective leadership and management is crucial for organizations to achieve organizational goals. In other words, leadership and management is a vital component that influences and drives the group efforts towards the accomplishment of goals. While organizations may employ various strategies to attain their goals, one would agree that managers' traits, qualities, and education levels certainly play a pivotal role in maximizing efficiency, achieving organizational goals, providing guidance, and communicating the plans and policies. Managers, particularly Chief Executive Officers (CEOs), formulate both short and long term plans and policies and execute them to move their organizations forward. Thus, CEOs must be well-educated and possess multidimensional traits.

One would agree that competent and well-versed CEOs are one of the most critical components of organizational success. They are often called upon to solve problems when navigating through the global business world. Hence, they must be equipped with the right set of skills to identify and tackle various challenges and problems that may arise in a highly complex global business world.

While there is no agreed upon way to lead an organization, the wrong type of leadership can certainly sink the whole organization. For any global company to be successful, the choice of well-educated, competent, and experienced CEO is important. In the absence of well-educated CEO, no Fortune 100 global company can work efficiently. Therefore, industry experience and knowledge should guide managers to tackle emerging problems. Although the environment and geographic region in which they operate vary, Fortune 100 global companies face similar challenges such as allocating resources properly, managing change, setting clear directions, identifying attainable goals, and achieving their targets.

This study examines Fortune Global 100 companies and the profiles of their CEOs. In this study, we are particularly interested in examining education levels of the CEOs across four different geographic regions of North America, South America, Asia, and Europe.

LITERATURE REVIEW

Fortune magazine surveys America's and the world's top companies every year. They are usually ranked by their total revenues for their respective fiscal years. In addition to the ranking performed by Fortune magazine, these companies have been analyzed and examined in numerous studies. For instance, a number of studies examined Fortune 100 or Fortune 500 in terms of how they utilize the internet and their websites to communicate with their customers, stakeholders, and other external entities. For instance, a content analysis of the websites of Fortune's "100 Best Companies to Work For" performed by Joyce (2003) suggests that the 100 Best are distinguished by employee development programs, diversity initiatives, and a fun work environment. Similarly, Castleberry (2002) analyzed Fortune 100 firms and concluded that while most Fortune 100 firms' web pages included general information (e.g., financials), few listed industry information or provided an easy contact method for more information. Jonsen et al., (2015) examined the values posted by Fortune 100 companies on their websites. They analyzed the performance implications associated with the espousal of values, and their stability or change over time. They conclude that it matters not just what companies do but also what they say.

Several studies surveyed how such companies employ social media. For instance, using a sample of more than 15,000 posts from Fortune 500 companies' Facebook accounts in a five-year time frame, a study by Liu et al., (2017) analyzed the effects of companies' social media communication strategies. Three communication strategies are examined including appeal, vividness, and interactivity. Results confirm the main effects of vividness and emotional appeals. Corkindale and Morgoch (2013) investigated and assessed mobile readiness of Fortune 500 companies' mobile websites and found that most Fortune 500 websites were not mobile ready. Swani et al., (2014) performed a longitudinal content analysis of a sample of more than 7000 tweets by Fortune 500 companies, and found that marketers in B2B and B2C settings exhibit significant differences in their branding and selling strategies, their use of message appeals, and the use of cues, links, and hashtags to support information searches.

Tian (2005) examined the mission statements of Fortune 100, and suggested that although 55 of the 100 companies at that time had a mission statement on their websites "companies with higher rankings are found to be more likely to post a mission statement on their web sites than those with lower rankings, which suggests that companies with higher rankings are integrating mission statements into their online image management strategies." King et al., (2011) evaluated the mission statements of Fortune 100 companies by examining the mission statements of the 25 largest businesses in the U.S., Australia, Canada, and Great Britain. The study found that the most frequently mentioned topics were quality, core values, and leadership. As a result, the literature pointed to the notion that Fortune 100 businesses were driven to communicate the quality of the products they were delivering to their customers.

The literature also cites references and alludes to a number of other constructs used to evaluate Fortune 100 or 500 companies. For instance, Clement (2006) conducted a study of reports in the business news between January 1, 2000 and June 30, 2005 and found that as of 1999, 40 corporations in the Fortune 100 had committed behaviors that can be considered unethical. Kumar et al., (2018) interviewed executives from a Fortune 100 high-technology company to

identify challenges firms face in managing the customer experience effectively in solutions delivery processes. They found that in solutions selling, the delivery phase is particularly critical to success. Schwaig et al., (2006) examined how well Fortune 500 firms address information privacy concerns and assess the substance and content of their stated information practices. Maldonado et al., (2018) collected information on organizational cultures of humility for a sample of Fortune 500 firms using data from company mission statements, company websites, and news articles and introduced the concept of a humble organizational culture that can be defined as a culture that promotes humility as a key success factor and a source of competitive advantage.

Few other studies focus on individual differences of CEOs such as their genders and cognitive capabilities. For instance, Sprangler et al., (2012) focus on individual differences of leadership among Fortune 100 CEOs and describe a general methodology for developing and validating measures of leader individual differences based on computerized language analysis of archival data. In terms of their genders, Spencer et al., (2019) interviewed 57 women CEOs of Fortune companies in effort to understand the common characteristics of female CEOs. They found common themes of female CEOs' courage risk taking, resilience, and ability to manage ambiguity. Glass and Cook (2016) surveyed women who have ever served as CEO in the Fortune 500 with a matched sample of men. They analyzed the conditions under which women are promoted to top leadership positions and explored the opportunities and challenges they face post-promotion. Their analysis reveals that women are more likely than men to be promoted to high risk leadership positions and often lack the support or authority to accomplish their strategic goals.

Wai and Rindermann (2015) examined Fortune 500 CEOs in terms of their education and cognitive abilities at four time points that spanned 1996 to 2014. They found that between 37.5% and 41.0% of these CEOs attended an elite school. Higher CEO education and cognitive ability was associated with higher gross revenue of the CEO's company.

The aforementioned studies suggest, while numerous studies explored Fortune 100 or 500 companies, few investigated CEOs in terms of their education levels and the field in which they majored. In the following sections, this study makes an attempt to survey Fortune Global 100 companies and education levels of their CEOs.

RESEARCH METHOD

Each year Fortune ranks the world's top companies based on a number of different criteria such as revenues and profits. The sample we used in this study was the companies on the Fortune Global 500 list published in 2018. It's an annual list of the 500 largest global publicly and privately held companies that were ranked by their total revenues for their respective fiscal years ended on or before March 31, 2018.

We compiled our Fortune Global 100 list using data available on Fortune magazine's website (Fortune, 2019). While the website lists the top 500 global companies, we picked and examined the first 100 global companies, which is a comprehensive list of the world's largest corporations as measured by their revenues, profits, assets, and employees. We examined each company with respect to its revenues, the number of employees, its sector, industry, and Chief Executive Officer (CEO). Once we identified each company's CEO, we then compiled data on their educational background with respect to the domain they majored in, and the highest degree they obtained.

DATA ANALYSIS AND RESULTS

Fortune Global 100 Companies

We first broke down the global companies into four different geographic regions as tabulated in table 1. As seen, both Asia and North America have 37 corporations on the Fortune Global 100 list. Overall, Asia and North America are home to 74 global 100 companies, while Europe is home to 25 global 100 list, and South America is home to 7 companies. Although the Fortune Global 100 included only one Latin American company (Petrobras, ranked 73), to be able to make a meaningful comparison among different regions, in this study, we included every Latin American company listed on the Fortune Global 500. As seen, USA dominates the list with 37 companies followed by China with 21 companies. The world's largest companies are scattered across the globe. Africa had no companies in the top 100.

The world's 106 largest companies generated $13,714,446 in revenues in 2017, employed 26,096,529 people worldwide, and are represented by 15 countries. While North American companies generated the highest amount of revenues, Asian companies employed the highest number of employees (table 1).

Table 1: Regions, Number of Companies, their Revenues and Employees

Region/Country	Count of Company	Sum of Revenues ($M)	Sum of Employee
Asia	**37**	**$4,907,532**	**11,922,774**
China	21	$2,974,018	8,205,284
Japan	9	$1,055,771	1,801,341
Russia	2	$205,880	573,200
Singapore	1	$136,421	3,935
South Korea	3	$380,743	535,888
Taiwan	1	$154,699	803,126
Europe	**25**	**$3,253,455**	**4,682,893**
Britain	3	$435,677	327,398
France	6	$669,923	987,217
Germany	8	$1,025,517	2,204,839
Italy	3	$264,692	167,161
Netherlands	2	$473,547	391,637
Spain	1	$87,401	198,960
Switzerland	2	$296,698	405,681
North America	**37**	**$5,173,269**	**8,817,804**
USA	37	$5,173,269	8,817,804
Latin America	**7**	**$380,190**	**673,058**
Brazil	7	$380,190	673,058
Grand Total	**106**	**$13,714,446**	**26,096,529**

Having summarized Fortuna Global 100 companies in table 1, we then examined their CEOs based on their gender and age. As summarized in table 2, only 3 (2.8%) out of 106 CEOs are females. As seen, North American companies are home to the female executives, while the other three regions have no female CEO. Further, in terms of their average age, CEOs in four

different regions are similar to one another. The youngest CEO who in charge of Pacific Construction Group of China is 32 years old, and the oldest CEO is 88 years old and is in charge of Berkshire Hathaway located in North America. Overall, a typical CEO is 58.90 years old, which may suggest that CEOs must have years of experience to hold such a position.

Table 2: Regions and Gender of CEOs

Region/Gender	Average of Age	Min of Age	Max of Age	Count of Gender
Asia	**59.97**	**32.00**	**74.00**	**37.00**
M	59.97	32.00	74.00	37.00
Europe	**57.00**	**42.00**	**65.00**	**25.00**
M	57.00	42.00	65.00	25.00
North America	**59.30**	**45.00**	**88.00**	**37.00**
F	59.00	57.00	61.00	3.00
M	59.32	45.00	88.00	34.00
South America	**57.86**	**46.00**	**65.00**	**7.00**
M	57.86	46.00	65.00	7.00
Grand Total	**58.90**	**32.00**	**88.00**	**106.00**

We further explored the Fortuna Global 100 companies in term of the sector in which they operate, the number of companies within each sector, the average age by sector, total revenues generated, and the number of employees employed by each sector. As tabulated in table 3, the Financials sector is home to 26 companies followed by the energy sector with 23 companies. It appears that the Financials sector employs the highest number of employees with a total number of 5,384,211employees, followed by the Energy sector with a total number of 4,754,906 employees. Similarly, the Energy sector generated the highest amount of revenues, followed by the Financials sector. While the Engineering & Construction sector has the lowest average age, the Telecommunications sector seems to have the highest average age.

Only one North American Aerospace & Defense company (Boeing, ranked 64) made the list. In addition, not surprisingly, several well-known motor vehicles companies such as Toyota, Volkswagen, Ford, Hondo, and General Motors are listed among the largest Fortune global 100 companies. Similarly, such energy giants as BP, Exxon Mobil, Chevron, and Royal Dutch Shell are ranked by Fortune among the largest corporations in terms of both the number of employees and the revenues they generated.

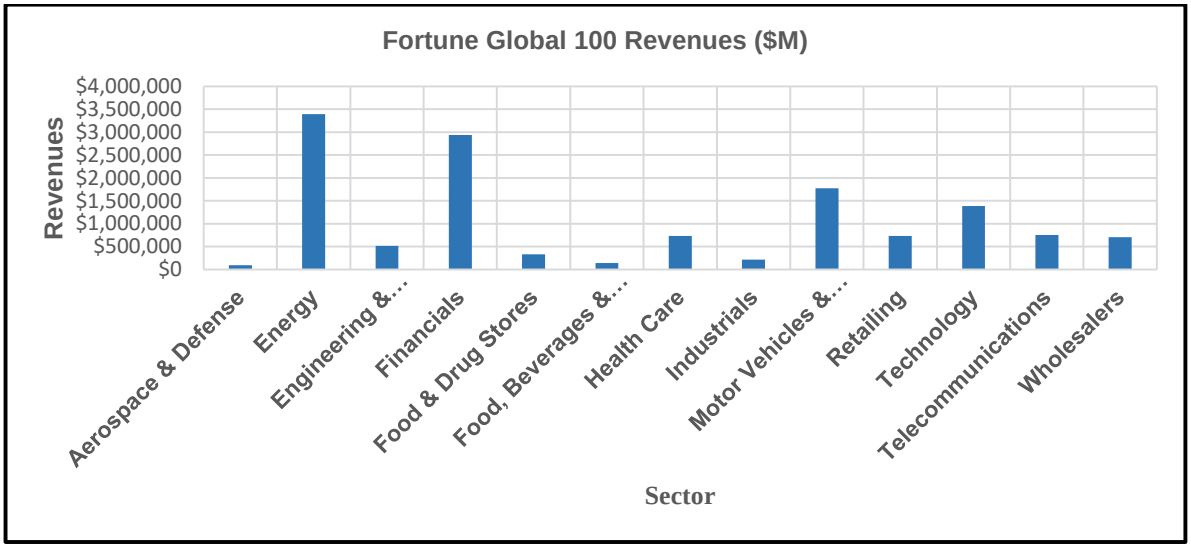
Table 3: Global 100 Companies by Sectors

Sector	Count of Company	Average of Age	Sum of Revenues ($M)	Sum of Employee
Aerospace & Defense	1	55.00	$93,392.00	140,800.00
Energy	23	59.70	$3,392,960.00	4,754,906.00
Engineering & Construction	5	51.40	$516,315.00	1,452,825.00
Financials	26	58.23	$2,937,955.00	5,384,211.00
Food & Drug Stores	3	61.00	$332,152.00	1,117,923.00
Food, Beverages & Tobacco	2	56.50	$142,340.00	558,000.00
Health Care	6	60.33	$734,662.00	1,102,769.00

Industrials	2	59.00	$213,859.00	685,000.00
Motor Vehicles & Parts	12	60.25	$1,777,636.00	3,030,762.00
Retailing	3	59.67	$730,272.00	2,895,000.00
Technology	11	59.27	$1,383,329.00	3,164,282.00
Telecommunications	7	61.00	$754,911.00	1,614,417.00
Wholesalers	5	57.40	$704,663.00	195,634.00
Grand Total	**106**	**58.90**	**$13,714,446.00**	**26,096,529.00**

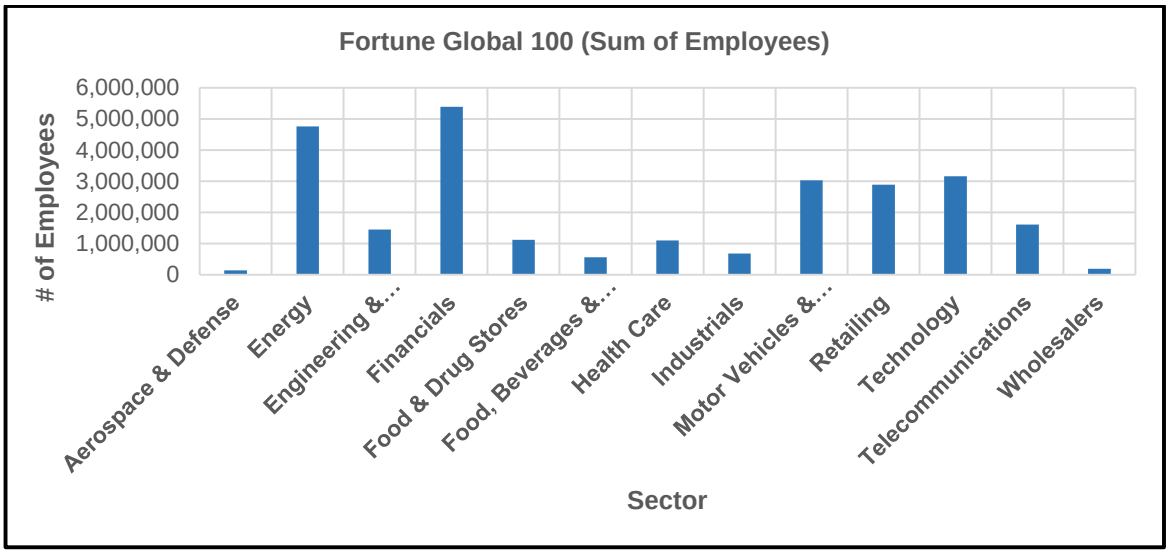
Figure 1 depicts the total revenues generated by each sector. As seen, while the Energy, Financials, and Motor Vehicles & Parts sectors are the top three sectors responsible for generating the highest amount of revenues, the Aerospace & Defense, Food, Beverage & Tobacco, and Industrials sectors are at the bottom of the global 100 list in terms of the total revenues they generated as of March 31, 2018.

Figure 1: Fortune Global 100 companies by sector



With respect to the total number of employees, the Financial, Energy, and Technology sectors are the top three sectors. The Aerospace & Defense and Wholesalers sectors employ the fewest number of employees (Figure 2).

Figure 2: Fortune Global 100 companies by the number of employees.



Looking at both figures 1 and 2, it may be argued that there is a positive relationship between the amount of revenues generated and the number of employees employed by a Fortune global 100 company. Although we are not running any hypothesis tests in this study, that assumption, from a statistical point of view, may be validated by running a simple correlation analysis. A similar conclusion may be gleaned from a scatter plot as well.

Table 4 tabulates the global 100 companies by their industry, the average age of the CEOs, sum of the revenues generated within each industry, and the total number of employees in each industry. As seen, the Banking and Petroleum Refining industries have the highest number of companies followed by the Motor Vehicles & Parts industry. While in terms of the total revenues generated the Petroleum Refining industry tops the list, with respect to the total number of the employees the Banking industry holds the highest position with 3,535,461 employees.

The average age column in table 4 indicates that Network and Communication Equipment and Insurance industries seem to be run by the oldest CEOs with 74 and 73 years of age, respectively. Similarly, CEOs working in the Internet Services and Retailing industry have the lowest average age of 49.5.

Although the Health Care Industry is a huge industry in at least North America, it's among the smallest industries with only 26,600 employees.

Table 4: Global 100 Companies by Industry

Industry	Count of Company	Average of Age	Sum of Revenues ($M)	Sum of Employee
Aerospace & Defense	1	55.00	$93,392.00	140,800.00
Banks: Commercial and Savings	15	57.00	$1,477,376.00	3,535,461.00
Computer Software	1	51.00	$89,950.00	124,000.00
Computers, Office Equipment	2	55.50	$307,894.00	268,000.00
Diversified Financials	2	50.50	$274,071.00	314,837.00
Electronics, Electrical Equip.	4	64.00	$528,314.00	1,548,372.00
Energy	2	53.50	$193,411.00	482,175.00

Engineering & Construction	5	51.40	$516,315.00	1,452,825.00
Food & Drug Stores	3	61.00	$332,152.00	1,117,923.00
Food Consumer Products	1	53.00	$91,222.00	323,000.00
Food Production	1	60.00	$51,118.00	235,000.00
General Merchandisers	2	59.00	$629,368.00	2,482,000.00
Health Care	3	57.67	$491,477.00	127,900.00
Health Care: Pharmacy and Other Services	1	59.00	$100,065.00	26,600.00
Industrial Machinery	2	59.00	$213,859.00	685,000.00
Information Technology Services	1	61.00	$79,139.00	397,800.00
Insurance and Managed Care	2	57.00	$291,198.00	316,000.00
Insurance: Life, Health	7	58.86	$866,040.00	1,091,249.00
Insurance: Property and Casualty	2	73.00	$320,468.00	442,664.00
Internet Services and Retailing	2	49.50	$288,721.00	646,110.00
Mining, Crude-Oil Production	3	61.67	$322,671.00	254,765.00
Motor Vehicles & Parts	12	60.25	$1,777,636.00	3,030,762.00
Network and Communications Equip.	1	74.00	$89,311.00	180,000.00
Petroleum Refining	15	59.73	$2,365,351.00	2,890,447.00
Pharmaceuticals	2	63.00	$158,634.00	557,169.00
Pharmacy and Other Services	1	63.00	$184,765.00	203,000.00
Specialty Retailers	1	61.00	$100,904.00	413,000.00
Telecommunications	7	61.00	$754,911.00	1,614,417.00
Trading	2	57.00	$213,186.00	67,734.00
Utilities	3	61.67	$511,527.00	1,127,519.00
Grand Total	**106**	**58.90**	**$13,714,446.00**	**26,096,529.00**

Education Profiles of CEOs of Fortune Global 100 Companies

In this study, we are particularly interested in examining the undergrad degrees CEOs of the Fortune Global 100 companies obtained. Table 5 tabulates various undergrad majors and regions. As seen, out of 106 CEOs, a total of 31 (29%) have a degree in engineering. Similarly, 13 (12%) CEOs have a degree in business administration, and 13 (12%) have an undergrad degree in economics. Table 5 shows that 24 (22.6%) CEOs have a bachelor's degree, however, what they majored in was not stated in their profiles available on the Internet. Thus, we just listed them as CEOs with a bachelor's degree. Looking at each region separately, we see that 11 (10%) CEOs in both Asia and North America have a bachelor's degree in engineering. Similarly, 8 (7.5%) European CEOs have a bachelor's degree in engineering. It appears both European and North American CEOs prefer a degree in engineering and business administration. Based on the data summarized in table 5, one can infer that no Asian CEOs have a degree in Business Administration.

Table 5: Undergrad Majors and Regions					
Undergrad Major	Asia	Europe	North America	Latin America	Grand Total
Accounting		1	3		4
Applied Sciences		1			1
Arts and Sciences	1				1
Bachelor's Degree	12	2	7	3	24
Business Administration		5	6	2	13
Economics	6	3	3	1	13
Engineering	11	8	11	1	31
English			1		1
Finance			2		2
Geology	1				1
History			1		1
Industrial and Labor Relations			1		1
Law	1	1			2
Management		1			1
Marine Technology	1				1
Pharmacy			1		1
Physics	1	2			3
Science & Technology	1	1	1		3
Sociology	1				1
Telecommunications	1				1
Grand Total	**37**	**25**	**37**	**7**	**106**

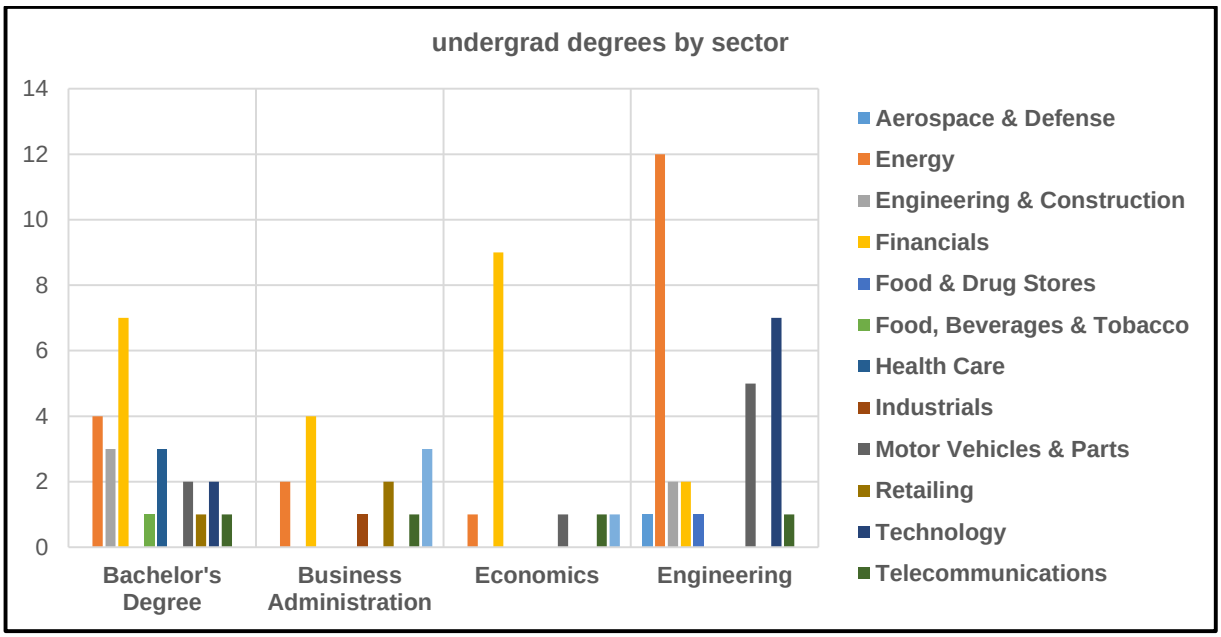
We further examined the sectors to discover which undergrad degree is more common in each sector. A close examination of the data tabulated in table 5 reveals that 81 out of 106 (76%) CEOs have an undergrad degree in 4 major fields: Bachelor's degree, Business administration, Economic, and Engineering. As seen in table 6, 12 out of 31 CEOs with an undergrad degree in engineering seem to be employed in the Energy sector. Similarly, 7 CEOs with the same degree are in charge of companies in the technology sector. The Financials sector seems to be dominated by CEOs with a degree in Economics as 9 out of 13 CEOs in the Financials sector have an undergrad degree in Economics.

Table 6: Undergrad Majors and Sectors					
Row Labels	Bachelor's Degree	Business Administration	Economics	Engineering	Grand Total
Aerospace & Defense				1	1
Energy	4	2	1	12	19
Engineering & Construction	3			2	5
Financials	7	4	9	2	22
Food & Drug Stores				1	1

Food, Beverages & Tobacco	1				1
Health Care	3				3
Industrials		1			1
Motor Vehicles & Parts	2		1	5	8
Retailing	1	2			3
Technology	2			7	9
Telecommunications	1	1	1	1	4
Wholesalers		3	1		4
Grand Total	**24**	**13**	**13**	**31**	**81**

Figure 3 illustrates the data tabulated in table 6. As seen, the Financials sector is also home to 7 CEOs with a bachelor's degree in unspecified majors.

Figure 3: Undergrad degrees by sector.



We further investigated CEOs with respect to their graduate degrees they earned. As seen in table 7, overall, 55 CEOs (52%) have a graduate degree, 30 of which have an MBA degree while 51 have no MBA degree. North America dominates the list with 12 CEOs with an MBA degree followed by Asian 8 CEOs with an MBA degree. Surprisingly, 4 out of 7 Latin American CEOs have an MBA degree. Overall, a total of 30 out of 106 CEOs (28%) have an MBA degree. Moreover, a total of 6 CEOs (5.6%) have a master's degree in economics, 4 of which are Asian CEOs. Looking at the females CEOs, 2 out of 3 have an MBA degree.

Table 7. Distribution of Master's Degrees by Region					
Master's Degree	Asia	Europe	North America	South America	Grand Total
Accounting	1		2		3
Aeronautics and Astronautics			1		1
Chemical Engineering		1			1
Computer Engineering	1				1
Computer Science	1		1		2
Economics	4	1	1		6
Electrical Engineering	1				1
Engineering		1			1
Law			2		2
Management	1	1			2
MBA	8	6	12	4	30
Mechanical Engineering		2			2
No Master's Degree	***18***	***12***	***18***	***3***	***51***
Petroleum Engineering	1				1
Physics		1			1
Science & Technology	1				1
Grand Total	**37**	**25**	**37**	**7**	**106**

.

In terms of the highest degree earned, out of 106 CEOs, only 14 (13%) have a Ph.D. degree. Breaking down the Ph.D. degrees, we see that while 7 CEOs in Asia, 4 in Europe, 3 in North America hold a Ph.D. degree, none of the Latin American CEOs has a Ph.D. degree. 5 CEOs have a Ph.D. degree in Economic, while 3 have a similar degree in Law. It appears, a Ph.D. degree is more popular with Asian CEOs. Surprisingly, 3 CEOs have a Ph.D. degree in Law.

Table 8: Distribution of Ph.D. Degree by Region					
Ph.D.	Asia	Europe	North America	South America	Grand Total
Economics	3	2			5
Engineering	1	1			2
Law			3		3
Management	2				2
Money and Banking	1				1
No Ph.D.	***30***	***21***	***34***	***7***	***92***
Physics		1			1
Grand Total	**37**	**25**	**37**	**7**	**106**

DISCUSSIONS AND CONCLUSIONS

Fortune magazine surveys and publishes a list of the largest companies in the US and in the world by such measures as revenue, number of employees, and profits. These corporations generate billions of dollars in revenues and employ millions of employees. While numerous studies performed in the pasted have examined such companies using various constructs, in this study we explored Fortune global 100 companies in four different regions and investigated their CEOs with respect to their education levels and the domains in which they earned their degrees.

USA dominates the list with 37 companies followed by China with 21 companies. While world's largest companies are scattered across the globe, Africa had no companies in the top 100. The world's 106 largest companies generated $13,714,446 in revenues in 2017 and employed 26,096,529 people worldwide and are represented by 15 countries.

Looking at their genders and ages, only 3 (2.8%) out of 106 CEOs are females all of whom are in charge of North American companies. In terms of their average ages, the CEOs in five different regions are similar to one another. While the youngest CEO is 32 years old and is in charge of Pacific Construction Group of China, the oldest CEO is 88 years old and is in charge of Berkshire Hathaway located in North America.

The Financials sector is home to 26 companies followed by the energy sector with 23 companies. It appears the Financials sector employs the highest number of employees with a total of 5,384,211employees followed by the energy sector with a total of 4,754,906 employees. Similarly, energy sector generated the highest amount of revenues followed by the Financials sector. While the engineering and construction sector has the lowest average age, the telecommunications sector has the highest average age.

We were particularly interested in examining the education levels CEOs of the Fortune global 100 companies obtained. It appears, both European and North American CEOs prefer a degree in engineering and business administration. With regards to a graduate degree, overall, 55 CEOs (52%) have a graduate degree. Although it's a little bit surprising that none of the Asian CEO has a degree in Business Administration, they seem, based on the data, to prefer a degree in engineering. In terms of the highest degree earned, out of 106 CEOs, only 14 (13%) have a Ph.D. degree. It appears, a Ph.D. degree is more popular with Asian CEOs.

REFERENCES

Castleberry, S.B. (2002). The Web as an information source for sales recruits: its effectiveness in aiding anticipatory socialization of salespeople, *Industrial Marketing Management*, *31* (7), 573-579.

Clement, R. W. (2006). Just how unethical is American business? *Business Horizons*, *49* (4), 313-327.

Fortune (2019). Fortune Global 500, Retrieved February 12, from http://fortune.com/global500/list/

Glass, C., and Cook, A. (2016). Leading at the top: understanding women's challenges above the glass ceiling, *The Leadership Quarterly*, *27* (1), 51-63.

Joyce, K. J. (2003). Lessons for employers from Fortune's "100 Best", *Business Horizons*, *46* (2), 77-84.

Jonsen, K., Galunic, C., Weeks, J., and Braga, T. (2015). Evaluating espoused values: does articulating values pay off? *European Management Journal*, *33* (5), 332-340.

King, D., Case, C., and Premo, K. (2011). A mission statement analysis comparing the united states and three other english speaking countries, *Academy of Strategic Management Journal*, *10* (51). 21-45.

Kumar, A., Steward, M. S., and Morgan, N. F. (2018). Delivering a superior customer experience in solutions delivery processes: seven factors for success, *Business Horizons*, *61* (5), 775-782.

Liu, J., Li, C., Jie, Y. G., North, M., and Yang, F. (2017). Like it or not: the Fortune 500's Facebook strategies to generate users' electronic word-of-mouth, *Computers in Human Behavior*, 73, 605-613.

Maldonado, T., Vera, D., and Ramos, N. (2018). How humble is your company culture? and, why does it matter? *Business Horizons*, *61* (5), 745-753.

McCorkindale, T., and Morgoch, M. (2013). An analysis of the mobile readiness and dialogic principles on Fortune 500 mobile websites, *Public Relations Review*, *39* (3), 193-197.

Schwaig, K.S., Kane, G.C., and Veda S. C. (2006). Compliance to the fair information practices: how are the fortune 500 handling online privacy disclosures? *Information & Management*, *43* (7), 805-820.

Spangler, W.D., Gupta, A., Kim, D.H., and Nazarian, S. (2012). Developing and validating historiometric measures of leader individual differences by computerized content analysis of documents, *The Leadership Quarterly*, *23* (6), 1152-1172.

Spencer, S.M., Blazek, S., and Orr, E. (2019). Bolstering the female CEO pipeline: equalizing the playing field and igniting women's potential as top-level leaders, *Business Horizons*, *62* (5), 567-577.

Swani, K., Brown, B. P., and Milne, G.R. (2014). Should Tweets differ for B2B and B2C? An analysis of Fortune 500 companies' Twitter communications, *Industrial Marketing Management*, *43* (5), 873-881.

Tian, Y., (2005). Being responsible to the society. a computer-assisted text analysis of the mission statements on the Fortune 100 companies' web sites, in proceeding of Annual Meeting of the International Communication Association, New York.

Wai, J., and Rindermann, H. (2015). The path and performance of a company leader: a historical examination of the education and cognitive ability of Fortune 500 CEOs, *Intelligence*, 53, 102-107.

DECISION SCIENCES INSTITUTE

Value From Big Data Analytics And Impact On Firm Performance: A Conceptual Framework

Vatsal Paghadal
University of Toledo, Ohio
Email: Vatsal.Paghadal@rockets.utoledo.edu

Md. Golam Kibria
University of Toledo, Ohio
Email: Md.Golam.Kibria@rockets.utoledo.edu

Tianling Xie
University of Toledo, Ohio
Email: Tianling.Xie@rockets.utoledo.edu

Seyedeh Golnoush (Julie) Joreimi
University of Toledo, Ohio
Email: SeyedehGolnoush.Joreimi@rockets.utoledo.edu

Jeen Su Lim
University of Toledo, Ohio
Email: Jeen.Lim@utoledo.edu

ABSTRACT

Nowadays big data analytics is highly considered as an important technological development in academic and business world. Though there are increasing number of firms that are participating in big data initiatives, there is still little understanding on how firms create the potential outcomes of such technological advancement into business value. Recent studies argue that to grasp big data analytics and experience performance gains, firms must build strong big data analytics capabilities. Therefore, this paper tries to develop a conceptual model that explains how big data analytics creates value to increase firm performance based on Resourced Based View (RBV) theory.

KEYWORDS: Big data analytics, Firm performance, Financial performance

INTRODUCTION

Big data analytics (BDA) can help examine huge amounts of data to find correlations, hidden patterns and other insights. Data analytics helps to obtain knowledge about market movements. Product recommendations can be provided after studying cyclic variations. Analytics of data helps to detect fraudulent cases. For promotion purposes, analytics can help in strategically placing advertisement (Yaqoob, 2016). Big data has been widely noticed as one of the most important areas of upcoming technology. It has attracted the attention of many industries, as it has the potential to provide high business value for firms of all sizes (Grover, 2018). While the potentiality of big data analytics to create high business value has been well recognized in the literature, there exists limited understanding in terms of business value that companies can create from big data analytics investments. This study intends to provide answers to the following research issues. First, this study identifies what determines the big data analytics

adoption by firms. Second, this study develops the process mechanism of business value creation by big data analytics using the resource-based view (RBV) as a backdrop. According to Bharadwaj (2000) and Wernerfelt (1984), the resource-based view (RBV) attributes better financial performance to the firm resources and capabilities. Based on this ground this study uses RBV to demonstrate that the business value achieved from investments in big data analytics leads to advantages in terms of the financial performance of a firm. Finally, the impact on firm's performance from value generated from big data analytics is described in the proposed conceptual framework. Theoretical and practical implications of the proposed conceptual framework are also presented in the paper.

LITERATURE REVIEW

Big Data and Big Data Analytics

Business intelligence and analytics (BI&A) and other similar field of big data analytics have become very important in both the academic and the business world over the past two decade (Chen, 2012). Big data has been defined as high-volume (large scale), high-velocity (moving/streaming), and high-variety (e.g., numerical, text, video, etc.) information assets that demand cost-effective, innovative forms of information processing for improved insight and decision making (Gillon, 2014). Big data analytics is becoming popular due to its greater ability to both capture vast amounts of data and apply more powerful analytical techniques to vast data sets. This recent ability of firms to be able to both collect varied data and also apply effective analytical techniques to such data enables the organization to automate highly complex decisions that have traditionally been dependent (primarily or solely) on human judgment and intuition (Brynjolfsson, 2011; Gillon, 2014).

The sophistication in data analysis has influenced organizations to implement and adopt BDA technologies that can process and analyze data to improve decision making. Because of its practical and methodological nature, it has many applications in almost all areas of business including e-commerce, marketing and finance (Chen, 2012).

The advancement of BDA is still in infancy. Our literature search reveals that implementation of BDA varies organization to organization depending on the resources and capabilities that organizations possess (Ghasemaghaei, 2015). So, we conceptualize BDA activity is a function of organizational resources that firms can leverage to gain maximum benefits from BDA adoption and implementation (Bajari, 2019). Therefore, the RBV theory is deemed appropriate to explain the influence of BDA adoption in organizations (Mikalef, 2017).

THEORETICAL FRAMEWORK AND HYPOTHESES

The conceptual framework developed in this study is presented in Figure 1. The Figure 1 framework describes the three antecedents of big data analytics activity by firms. The framework discusses the key components of big data analytics activity and the moderators of the relationships between the three antecedents and big data analytics activity. The values created by big data analytics activity leading to firm financial performance are presented in the framework. Next section presents the components of the conceptual framework and related hypotheses.

Financial Resources and Big Data Analytics

This section focuses on why there is need for an investment in big data. According to Wamba (2017), "Big data analytics is now considered as a game changer that can enable improved business efficiency and effectiveness because of its high operational and strategic potential". Currently big data technologies are highly recognized as one of the most leading areas of future

technology (Raguseo, 2017). According to Tambe (2014), investments associated with big data technologies has observed 3% faster productivity growth for firms with significant data assets. This has developed interest among various industries to engage in big data practices, since it has the strength to provide companies with high business value.

It is also very important to comprehend the extent to which big data analytics can indulge in the creation of a competitive advantage that can be evaluated by considering the resource-based view (RBV) (Wamba, 2017). In this study we examine the forms of business value that companies can create from big data analytics investments and observe the direct impacts it has on the financial performance of a firm. From the resource-based view theory, this study determines the business value that can be achieved from investments in big data analytics in terms of the financial performance of a firm.

H1: Financial resources will have positive influence on big data analytics activity.

Technical Resources and Big Data Analytics

To manage the large-volume digital flood of data that is being generated at very high velocities and varieties, firms needs to update their data handling technologies that will help them to improve their services and create new, more effective, business models (Lucas, 2013). An information technology (IT)-enabled alteration idea which consists of a series of progressive changes leading to the operational improvement and internal amalgamation of IT functionalities will create competitive advantage and improve financial performance through business redesign. With this in mind, we believe that the adoption and innovation of big data technologies and information systems holds great promise for better business performance.

To implement big data technologies, it requires advanced technical resources like setting up new high-performance IT infrastructure. Technological trends, like the Internet of Things, the proliferation of the cloud computing as well as the spread of smart devices, they require powerful systems and distributed applications that support such multiple connections systems (e.g., smart grid systems) (A. Kankanhalli, 2016). Implementing big data technologies like machine learning, Hadoop, YARN will help to sort and understand the data which will help managers to make real time decision improving firm’s efficiency to perform better (Ahmed Oussous, 2018).

H2: Technical resources will have positive impact on big data analytics activity.

Human Resource and Big Data Analytics

The concept of Big Data is defined as those “information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value” (Andrea De Mauro, 2018). The fast evolution of big data analytics is forcing companies to rethink on their Human Resource (HR) needs (Andrea De Mauro, 2018). Many earlier studies have shown the vital role of human capital in the success of companies, especially those that have a high degree of technology intensity. It requires highly skilled professionals to operate and perform operations on high tech technical resources for big data analytics (Siepel, 2017). The job roles that require to perform big data analytics practice are business analyst, data scientist, big data developer, big data scientist (Thomas H. Davenport, 2012).

H3: Human resources will have positice impact on big data analytics activity.

Moderating Effects of Environmental Uncertainty

Environmental changes can affect innovation related performance. For example, such changes can affect firm performance (Han, 1998). We focus here on two aspects of environmental uncertainty: (a) market uncertainty, which refers to rate of change in the composition of customers and their preferences (Jaworski, 1993), and (b) technological uncertainty, which refers to the rate of technological change in the value addition processes (Jaworski, 1993).

A business that accepts new information technology initiatives can perform better with environmental changes in information intensive environments (Weiss, 1993). When market uncertainty is high, businesses try to gather more information from the market in order to better forecast future market trends and to better manage their channel members in prediction of these trends. Higher prediction and coordination would enable the businesses to respond quickly and efficiently to changes in customer preferences and demand. In addition, in such environments, big data initiatives in communications and order taking will help businesses serve customers better and may result in higher revenues.

On the other hand, when technological uncertainty is high, it is often hard to predict whether the newly adopted technology applications will have an immediate effect on performance outcomes. When technological uncertainty is high, customers and suppliers may be less willing to use systems that do not have proper outcomes, i.e. using old methods. They may have updated their systems and may want their partners to use the same technology so that there is ease in information sharing and ease in performing business practices. Base on this ground we can hypothesize that:

H4: The relationship between (a) financial resources, (b) technical resources, and (c) human resources and big data analytics activity will be moderated by environmental uncertainty.

Big data analytics and Value Creation

The identification of value creation is complicated in firms (Leonel Furtado, 2017). In big data implementation, we should acknowledge the uncertainty of value creation in future, regarding the trends inferred by predictive analytics processes. Big data analytics can analyze large volume of data with wide variety of data types. Otherwise, this analyzing would be almost impossible by traditional statistical tools. As per Gartner (2012), large volume of data provides opportunities for to create value for firms that can exploit the opportunities adopting BDA. According to Davenport (2014), firms create value adopting BDA in following ways:

1. The reduction of uncertainty in the decision-making process, where data are used as information sources for explanatory and predictive models.
2. The generation of improvements in products and services based on insights taken from analyzed data.
3. The reduction of costs, where data are used to identify fraud or the least costly investments.

Based on the discussion, we hypothesize:

H5: Big data analytics activity will have positive impact on value creation.

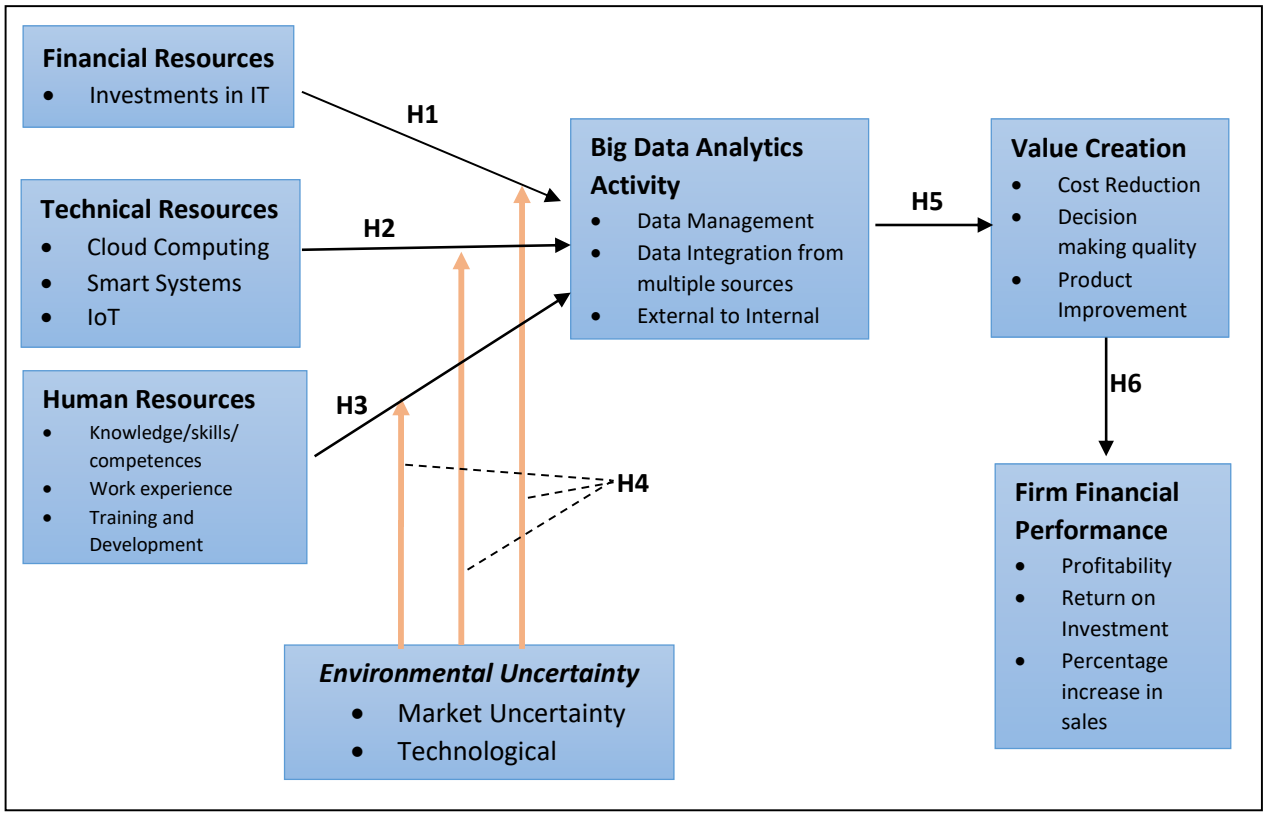
Value Creation and Firm Financial Performance

Big data analytics plays important role to improve firm performance (Wixom, 2013). Shahriar Akter (2016) found empirical evidence that big data analytics increased sales, profitability and

return on investment (ROI). In sum, big data analytics adds value for firms and that, in turn, increases firm performance. Therefore, we hypothesis,

H6: Value creation will have positive impact on firm financial performance.

Figure 1: Conceptual Framework



MANAGERIAL IMPLICATIONS AND LIMITATIONS

Being a new technology, big data analytics has tremendous importance for both academicians and practitioners. Theory-based research is needed to understand the factors that may drive or inhibit the use of big data analytics in organizations. In this study, we critically reviewed a prominent theory called resource-based view to explain how big data analytics activity of firms creates value and influences firm performance. Relevant constructs of the theory were discussed in respect to big data analytics. Finally, based on the discussion, we developed a conceptual framework and related hypotheses that can be tested empirically to guide future research in the area of big data analytics. This study is part of a growing area of research that is still in its infancy.

This study has several implications. First, we believe that our conceptual framework is a contribution to the existing knowledge because it adds unexplored factors that affect the acceptance of big data analytics. As a result, the proposed framework will help managers understand what organizational factors that drive the adoption and evolution of big data

analytics and how these factors are responsible for value creation and firm performance. Second, this paper contributes to literature proving the underlying constructs that act as antecedents to the adoption of emerging technologies such as big data analytics. Moreover, the hypothesized framework proposes the relationships between constructs surrounding big data analytics. Given the mature state of information systems literature and the embryonic stage of big data analytics research, we hope that this study will inspire and challenge others to investigate such new phenomenon in a manner that synthesizes existing theories of information systems, strategy, and competition and offers relevance to both researchers and practitioners.

REFERENCES

A. Kankanhalli, J. H. (2016). *Big data and analytics in healthcare: Introduction to the special section.* Inf. Syst. Front.

Ahmed Oussous, F.-Z. B. (2018). *Big Data technologies: A survey.* Journal of King Saud University – Computer and Information Sciences.

Andrea De Mauro, M. G. (2018). *Human resources for Big Data professions: A systematic classification of job roles and required skill sets.* Information Processing and Management.

Bajari, P. C. (2019). The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings*.

Bharadwaj, A. S. (2000). *A resource-based perspective on information technology capability and firm performance: an empirical investigation.* MIS Quarterly.

Brynjolfsson, E. a. (2011). Race against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy. *Lexington, MA: Digital Frontier Press*.

Chen, H. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188.

Davenport, T. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities.* Boston (USA): Harvard Business School Publishing Corporation.

Gartner, L. D. (2012). *Information Economics, Big Data and the Art of the Possible with Analytics.*

Ghasemaghaei, M. H. (2015). Impacts of big data analytics on organizations: a resource fit perspective. *Association for Information Systems*.

Gillon, K. A.-Y. (2014). Business analytics: Radical shift or incremental change? *Communications of the Association for Information Systems*, 287–296.

Grant, R. (1996). Prospering in dynamically competitive environments: Organization capability as knowledge integration. *Organization Science*, 375–387.

Grover, V. R.-P. (2018). *Creating strategic business value from big data analytics: A research framework.* Journal of Management Information Systems .

Han, J. K. (1998). Market Orientation and Organizational Performance: Is Innovation a Missing Link? . *Journal of Marketing 62 (October): 30-45.*

Jaworski, B. J. (1993). Market Orientation: Antecedents and Consequences. *Journal of Marketing 57 (July): 53-71.*

Leonel Furtado, M. D. (2017). *Value Creation in Big Data Scenarios: A Literature Survey.* Journal of Industrial Integration and Management.

Lucas, H. J. (2013). *Impactful Research on Transformational Information Technology: An Opportunity to Inform New Audiences.* MIS Quarterly.

Mikalef, P. F. (2017). Big data analytics capability: antecedents and business value. *Pacific Asia Conference on Information Systems (PACIS). Association For Information Systems.*

Nunnally J C, B. I. (1994). Psychometric Theory. *American Educational Research Journal* , 83.

Patrick Mikalef, M. B. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*.

Raguseo, V. (2017). *Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects.* International Journal of Production Research.

Shahriar Akter, S. F. (2016). *How to improve firm performance using big data analytics capability and business strategy alignment?* Int. J. Production Economics.

Siepel, J. C. (2017). *Non-founder human capital and the long-run growth and survival of high-tech ventures.* Technovation.

Tambe, P. (2014). *Big Data Investment, Skills, and Firm Value.* MANAGEMENT SCIENCE.

Thomas H. Davenport, D. P. (2012). *Data Scientist: The Sexiest Job of the 21st Century.* Harvard Business Review.

Wamba, S. F. (2017). *Big data analytics and firm performance: Effects of dynamic capabilities.* Journal of Business Research.

Weiss, A. M. (1993). The Nature of Organizational Search in High Technology Markets. *Journal of Marketing Research*.

Wernerfelt, B. (1984). *A resource‐based view of the firm.* Strategic management journal.

Wixom, B. Y. (2013). *Maximizing value from business analytics.* MIS Q. Exec. 12, 111–123.

Yaqoob, I. I. (2016). *Big data: From beginning to future.* International Journal of Information Management .

DECISION SCIENCES INSTITUTE
Existence of Gender Role Congruity by Social Provisions: A Moderation Based Analysis

Siddhartha Saxena
Ahmedabad University
Email: Siddhartha.saxena@ahduni.edu.in

ABSTRACT

Family businesses are paradoxical units, where the decisions are generally taken by the leader within the family business. Women seeking positions in family businesses often encounter considerable challenges in obtaining and retaining their roles. Role congruity was reported as barrier to the ability of women to successfully acquire and manage and roles within the organizations. Role congruity can be understood in terms of the involved perceptions of how an individual's gender role fits or achieves congruity with other roles held by the person. The paper tries to look at qualities under social provision and also understand the moderating role of gender which affects the family business roles and acceptance of them as successors.

KEYWORDS: Diversity Issues, Gender Studies, Social Issues, Participatory Management

INTRODUCTION

Research regarding the impact of role congruity suggested that there was a disconnect between follower and leader, perceptions of leader characteristics, and attributes of women. Garcia-Retamero and Lopez-Zafra (2006) asserted that women were commonly the objects of prejudice when they were placed in leadership roles. According to these authors, this prejudice against female leaders “stems from the incongruity between expectations about women and about leaders” (p. 51). In short, role congruity makes it difficult for followers and leaders to view women as having the qualities or attributes that were commonly associated with leadership positions. Although some may argue that gender roles are diminishing, particularly in the workplace, recent Gallop Poll responses confirm that stereotypical attitudes regarding leadership and gender still exist. For example, when it comes to choosing a boss, people still prefer men (Simmons, 2014). In addition, employees see women as emotional and affectionate, and men as more aggressive. Heilman (2014) has noted that the perception of gender roles is a natural consequence of gender perceptions and the subsequent expectations stereotypes produce about what women or men are like and how they should act.

LITERATURE REVIEW

Rationale for Choosing Family Businesses

Most of the identified family business best practices are aimed at ensuring not only family business success but, also, long-term family business continuity and succession. In the best practice approach, the success of family businesses appears to be so inextricably tied to, and dependent on, long-term continuity and succession has to be identified with it. Yet, not all family owned and managed businesses either seek long-term continuity and succession or are capable of achieving it. Research demonstrates that a high proportion of family business owners plan to sell their businesses (Smyrnios & Dana, 2006, 2007). In view of that, many family

businesses do not survive beyond the first generation which is not necessarily because of failure, nevertheless because their owners make a positive choice to harvest (Bierly, Ng, & Godfrey, 1999, p. 607). Small business management research largely has shown that the business planning leads to better financial performance (Ibrahim et al., 2004), but a close examination of that relationship reveals a more complex picture focusing primarily on type, nature, and extent of management (Danes et al., 2008b). Danes et al., (2007) acknowledged 10 business strategies which were factored into three categories viz., business management, quality management and financial management.

Although there is no agreement in the literature on how to measure the "Familiness" of the family business but, (Danes et al., 2007) it is found that family influence in the firm makes the family businesses unique entities (Chua et al., 1999). This is so because, in family businesses as in no other organizational form, the family and the business are "inextricably intertwined" (Perason et al., 2009, p 966)

Theories where the research will contribute

Family Systems theory

In family systems theory, the family is viewed as a relatively closed system of interactions between individual members (Albanese, 2010). The focus is not so much on input variables such as the personalities of individual members, or output variables such as the outcomes of family interactions, but on the *interactions* themselves. As such, it is interested in intra-family processes such as family functioning, conflict, cohesion, and communication. Recent examples include Doherty and Craft (2011) and Garrett-Peters et al. (2011).

Several key assumptions of systems theory are used in family scholarship. Perhaps the most important is that a family system must be understood as a whole. In other words, the behaviour of individual members, when looked at separately, will not provide an accurate picture of what is occurring (Whitechurch and Constantine, 2004). Because of the interconnectedness between different components of the family system and the individuals within, the system looked at holistically is something different than separate individuals co-existing. Given this focus on the whole, pathology within families is not seen as attributable to single individuals but rather as a dysfunction of the entire system (Albanese, 2010).

Social Exchange theory

There are four key assumptions in this theory. The first is that individuals are motivated by self-interest (White and Klein, 2008). To this end, social relationships are often described as a kind of market. In this 'market', each individual behaves in a self-interested way in order to maximize his or her individual gains. Homans for instance, states that "the open secret of human exchange is to get from the other man, a behaviour that is more valuable to you than it is costly to him and to give him behaviour that is more valuable to him than it is costly to you" (1961). The second key assumption is that individuals are constrained by their choices. In the attempt to maximize their gains and limit their losses, individuals use past experiences to make choices about the future (Smith et al., 2009).

Experiential Family Therapy

Virginia Satir was one of the pioneers in the evolving field of family therapy. Her ability in 'disentangling people from mystifying communicational traps' (Hoffman, 1981, p. 221) was famous. Her core concept is the self-esteem of persons, their ability to value their own self and to stay in a loving and friendly relationship with themselves. The way a person is able to do this is responsible for the way she/he communicates. The processes in the family of origin and with significant others intensely affect one's sense of self. Low self-esteem is required to be protected and covered by the individual, who therefore starts to communicate in an incongruent way. Satir saw four main patterns of incongruent communication (Satir et al., 1991):
• Blaming (attack as a way of protection);
• Placating (actively going into 'one-down-position');
• Irrelevant (distracting to avoid confrontation);
• Computing (avoiding any feelings).

Congruence means that the self-esteem of a person does not need to be protected and thus he/she is able to communicate clearly, openly and directly, expressing their own needs and interests without expecting immediate fulfilment, and listening to those of the others.

Context family theory

Every time the theme of kinship is explored, comparisons lead to other contexts. The researchers in Family business should be alert to the likelihood that any set of observations – whether categorized as a matter of politics and law, ritual and ideology, labour needs, ecology, and so on – might prove to be vital for explaining the family (Yanagisako, 1979). Kinship is notoriously inter-connected with other social and cultural topics (Creed, 2000; Godelier, 2011 [2004]; Peletz, 1995). For example, a study of 'the function of kinship in politics' concluded that understanding requires attention to 'other social categories' such as factionalism, religious affiliation, class and ethnicity' (Buessow, 2011, p. 108).

Evolutionary theory

Regardless of environmental constraints and additional difficulties, as the literature on strategy and family businesses demonstrate, family firms often have choices, and some are able to adapt as their environments change (Gallo 1995; Naldi et al. 2007). The family firm literature could benefit from borrowing the concept of co-evolution from the evolutionary approach. Theorists now recognize that families, organizations, populations, communities and their environments can co-evolve (Kepner 1983). For example, at the organizational level, co-evolution occurs within a population of heterogeneous firms when organizations with adaptive learning capabilities are able to interact and mutually influence each other (McKelvey 2002; Volberda and Lewin 2003).

The 'Family' as a Variable and Unit of Analysis

Despite several articles that have focused on clarifying the definition of a family firm and developing scales to measure the extent of family influence in a firm, hardly any efforts have been directed to either define or measure the 'family' variable (Pearson et al., 2014). Even in scales aimed to measure the family climate or family harmony or family influence, the task of defining the 'family' is left to the respondent which leaves open the possibility of multiple interpretations of a key term in the same data set. This is especially important given the significant heterogeneity in family structures in today's society where simultaneously multiple variations of family co-exist (Walsh, 2003; McGoldrick et al., 2010). While the 'family' variable is

important when one has to discuss the firm level of analysis, it becomes even more critical when designing studies using family as a unit of analysis.

As a social institution, family shapes the values of its members. In turn, these values influence the attitudes and behavioural choices of family members. It should be noted that 'family' as the unit of analysis does not imply unified or harmonious families. As with all social settings, conflicts are a natural part of the family (McKee et al., 2014). Yet another reason for considering the 'family' as the unit of analysis is the growing interest in the potential of family as a carrier of different forms of capital, such as social, cultural, financial, and human (Danes, 2014; Sorenson, 2014). In terms of defining and developing measures of the 'family' variable, some green shoots of ideas are emerging in the literature. For example, Sharma and Salvato (2013) propose the adoption of essence and component-based approaches for defining the family variable. They suggest that the components of consanguinity, cohabitation, legal status, generations, gender, and birth-order may be used to operationalize the family variable, which is what we are also doing by bringing in the factors from relations, behaviour, and social context. Data collected on these component dimensions can help to determine the heterogeneity of families within a study.

Given the strong case made by several recent authors in the 25th anniversary issue of FBR of a critical need to focus more attention on the 'family' variable in family business research (Litz et al., 2012; James et al., 2012), great promise lies in devoting efforts to design studies based on the family as the unit of analysis and to define and develop valid and reliable measures for the family variable

Context

Family business scholars have incorporated context in theoretical as well as empirical studies. Examples include understanding survival versus demise of family enterprises in hostile environments (e.g., Hatum, 2007; Orozco and González, 2014), the role of institutions in shaping the values of family enterprises (Parada, Nordqvist, and Gimeno, 2010), portfolio entrepreneurship (Sieger et al., 2011), principal–principal conflicts (Sauerwald and Peng, 2013), succession (Saxena, 2013), and the role of location and number of owners on innovation output (Deng et al., 2013).Over the last few decades, many regions in the world, including Asia Pacific, Europe, and Latin America have experienced rapid contextual changes.'

Foundational Theories with Family Science and Research Gap

Family science is an applied, interdisciplinary field of study. Research in this discipline tends to focus on family relationships, processes, communication, wellbeing and experiences in various socio-cultural, economic, and political contexts. We too are adopting some of the factors from it rather than restricting the work to only one context. For more comprehensive reviews of the theories indicated as foundational theories for this research, one can refer to the handbooks by Bengtson et al. (2005), White and Klein (2008), and Smith et al. (2009). Although numerous other foundational theories used in family science could potentially be drawn upon to inform and enrich understanding of family enterprise, our focus is on systems theory, life course theory, social exchange theory, structural functionalism, and symbolic interactionism.

Table 1: Comparison of 4 theories *				
	Life course	Systems	Symbolic interactionism	Social exchange

Source and Origin	It is derived from family development theory. It arose in the 1960s. The early theorists include Hill and Rodgers (1964) and Elder (1974).	It is derived from general systems theory. It arose in the late 1950s. The early theorists include Kantor and Lehr (1975) and Bateson et al. (1950) and x.	It is derived from American pragmatism. It arose in 1920s; The early theorists include Mead (1934), Cooley (1956) and Blumer (1969).	It is derived from market concept of exchange. It arose in the late 1950s. The early theorists include Thibault and Kelly (1959) and Homans (1961). .
Core Constructs	Life events, transitions, and paths; historical context; linked lives between family members.	Interaction patterns within the whole family system (e.g., conflict, communication); self-reflexivity	Symbols; values; behavioral norms; shared meanings; family identity; self-concept.	Resource exchange; social rewards and costs; self-interest; rationality; reciprocity; interdependence.
Central Premises	The life paths of individuals are influenced by the social ecology and historical context in which they live	The interconnections between family premise members impact the functioning of the family system	Interactions between family members create shared identities through common symbols (e.g., situations and values).	Family relationships entail a rational exchange of resources (social and economic) to maximize rewards and minimize costs.
Key assumptions	The timing and context of key events and transitions during an individual's life course are important	Individual behavior is best assumptions understood through an analysis of the family system	Individuals develop concepts of self and identity mainly through interactions in the family.	Individuals are rational and motivated by self-interest in relationships.
Analysis level Fundamental Question	Individual in social context What are the consequences of an individual's family context, and the timing of key events and transitions,	Family What are the processes or interactions within families that impact the overall functioning of the family system	Individual within family How do a family's symbolic meanings and interaction patterns affect how individual family members	Individual What are the costs and rewards of family relationships for a specific individual in the family and how do these influence his or her behavior?

	on his/her life course		think, feel and act?	
Recent work within the family science	Kennedy et al. (2010); Rauscher (2011)	Doherty and Craft (2011); Garrett-Peters et al. (2011)	Powell (2011); Sykes (2011)	Donnelly and Burgess (2008); Nakonezny and Denton (2008)
Literature Illustrative work within the family business literature	McGoldrick (1992); Kaye (1996); Gersick et al. (1997); Owens (2002); Klein (2008); Hoy and Sharma (2010)	Davis and Stern (1981); Tagiuri and Davis (1982); Beckhard and Dyer (1983); Kepner (1983); Lansberg (1983); Hollander and Elman (1988); McCollom Hampton (1988); Whiteside and Brown (1991); Stafford et al. (1999); Olson et al. (2003); Dyer (2006); Bjonberg and Nicholson (2007); Pieper and Klein (2007); Lumpkin et al. (2008);Distelberg and Blow (2011)	Lansberg (1983); Bertrand and Schoar (2006); Dyer (2006); Hall and Nordqvist (2008); Sharma and Manikutty (2005); Milton (2008); Sharma and Nordqvist (2008); Shepherd and Haynie (2009); Smith (2009)	Kellermanns and Eddleston (2007); Ward et al. (2007); Leaptrott and McDonald (2008)

* Table made based on papers reviewed by researcher.

To neutralize this effect, families exhibit visible conformity in the business practices that are important to these non-family stakeholders. In short, the use of socio emotional wealth preservation as a primary driver of family-controlled firms goes beyond the general proposition of institutional and stakeholder theory – that the presence of influential stakeholders exerts pressures on firms' decisions – to suggest that family owners exert additional pressure on themselves to be held in high regard by internal and external stakeholders. They do so for a combination of instrumental and normative motives that are intrinsically tied to the preservation of the family's SEW.

Table 2: Integrative approach to stakeholder management in family business indicating the need for this nature of research**			
Theory	Main propositions regarding stakeholder Management	The Research Gap	The SEW preservation approach to stakeholder management in family firms

Institutional Theory	Organizations need legitimacy to survive (Scott, 1995). Concern over legitimacy influences firms by pushing them to adopt certain managerial practices that are expected to be socially valuable even if these practices do not yield an economic benefit (Deephouse, 1999; Scott, 1995).	Why are some firms (e.g., family firms) more likely to respond to institutional pressures than others?	Variation in firm's responses to institutional pressures and stakeholders' needs is likely to be a function of who controls the organization and how much the controlling party values achieving social worthiness apart from any economic gains (Berrone et al., 2010)
Stakeholder theory	Firms goals in addition to maximizing shareholders' shareholders' economic value, and should accordingly aims at satisfying the needs of multiple constituents (Donaldson and Preston, 1995; Freeman, 1984). Regardless of the approach (instrumental versus normative), the link between stakeholder management and financial results is unclear (Margolis and Walsh, 2003; Orlizky et al., 2003).	Why are some firms (e.g., family firms) more likely to attend to external stakeholder demands than others?	Family firms that use socioemotional criteria when making choices are more likely to engage in stakeholder management activities because by doing so they protect and enhance their socioemotional endowments (Berrone et al., 2012)

****** Table based on papers reviewed by researcher.

Fundamental to family systems theory is a paradigmatic as well as a programmatic shift from an individualistic to a systemic approach. Stanton (2009) pointed out:

If we want to understand and treat individuals, couples, families, and larger social groups effectively, we need to conceptualize cases within the system in which they exist (the context and the meaning attributed to the context), assess the salient factors in the system, and intervene at identified points across the system.

There are several specific concepts that underlie the theoretical framework of family systems theory (Whitchurch and Constantine, 2009). The more prominent of them are briefly reviewed below:

Holism- This is a core concept in any systems theory: a system must be seen and understood as a whole, and it cannot be understood by dividing the elements into single units (nonsummativity).

Goal orientation- Families organize their common life according to more or less explicit goals, which are supposed to provide their coexistence as a family with a sense of meaning and continuity.

Equifinality and multifinality- Equifinality means that a system might achieve a certain goal via different activities. By contrast, multifinality denotes the idea that a specified family activity might lead to different consequences. So, the question of how the 'ideal' successor might be brought up, for example, turns out not to be answerable (Erdmann, 2010).

Regularity, rules, and patterns- Upon observing families over a longer period of time it becomes evident that the relationships among family members are 'patterned' in such a way that they can be characterized by certain rules and routines (Fiese, 2006).

Circular causality- In contrast to linear causality, which rests on the assumption that an event A causes a consequence B, circular causality conceives of ongoing interactions among family members as a reciprocal influence (Watzlawick et al., 1967).

Positive and negative feedback- Generally speaking, feedback is a process initiated by someone with the intention of either destabilizing or restabilizing an initial situation or set point. With respect to interaction among family members, positive feedback is principally change-oriented, whereas negative feedback aims at returning to the initial situation.

Boundaries- An important aspect when dealing with families is looking at the boundaries among the members within the family. This might be the boundary of each family member related to his or her individual privacy, or the boundaries governing the relations among family subsystems.

Self-organization and self-reference- An important feature of living systems is that – within certain limits – they are endowed with the potential for self-development and self-preservation – a feature that has been called 'autopoiesis' by Maturana and Varela (1980; see also Haken, 1992). An autopoietic system continuously reproduces the elements that it consists of, balancing out top-down and bottom-up processes.

According to Family business literature, family activities constitute the whole gamut of enterprises in which an entrepreneur or next-generation CEO and one or more family members considerably influence the firm. They alter it via various mechanisms like managerial or board participation, their ownership control, the culture and values that family shareholders hold that impacts the enterprise and their strategic preferences of shareholders. In this context, participation denotes the nature of the involvement of household members in the business. In the business, the household members may be involved as part of the management team, as board members, as shareholders, or as supportive members of the family foundation.

The complexity of Family businesses

Two theories can be focused upon to understand the complexities associated with the family business. These theories provide a wide spectrum of understanding of the issues involved. The first theory is based on the work of Johannisson and Huse (2000). Different rationalities include family, business, and owner's points of view. Different ideologies include entrepreneurship, managerialism, and paternalism. This is a model where differences in rationalities and ideologies can easily cause conflicts. Johannisson and Huse (2000) define this theory as "a consistent and permanent way of perceiving and appreciating the world that, accompanied by an emotional commitment, generates a particular mode of conduct" (p. 356). The second theory is Bowen's family systems theory. It brings out the complexities in family businesses. It is rooted in general systems theory (Bowen, 1981; Von Bertalanffy, 1968). According to Bowen, families

are social systems and individuals are parts of the family system. People develop predictable patterns of interaction. These patterns evolve into rules or beliefs about acceptable behaviour (Guttman, 1991; Kerr and Bowen, 1988).

Familiness

Familiness is one of the major factors that distinguish family businesses from nonfamily businesses. Familiness is the unique bundle of resources and capabilities that are exclusive to a business. It is a result of family interaction between the family, its individual members, and the business in the form of knowledge, social capital, and intentional trust (Habbershon & Williams, 1999). Identification of condition of interactions, resources, and capabilities that contribute to competitive advantage are necessary for a successful family business. Familiness enables a business to develop, adopt or implement strategies that enhance its efficiency and effectiveness, potentially beyond that of competitors (Habbershon & Williams, 1999;).

Women in family businesses

The entrepreneurship and family business literature normally supposes that an individual, the 'owner-manager' who is also presupposed to be the entrepreneur, handles the processes of founding and leading a business. That individual is conventionally identified as a male. The focus of research is often on him. The wives/partners/daughters/sisters of the owner-manager are seldom talked about in the literature. This lack of research interest, consequently, revealed the term 'invisible' or 'hidden' which describes the role of women in founding and operating businesses (Baines and Wheelock, 1999; Colli et al., 2003; Dhaliwal, 1998; Marshack, 1994; Mulholland, 1996a,b; Ogbor, 2000; Poza and Messer, 2001). In much of the literature, 'family' and 'business' appear as contradictory, opposite and in tension with respect to each other. Williams and Locke (1999) noticed that these "feminine qualities" can definitely be advantages in contemporary organizations where sharing information, developing relationships, empowering employees, and making collective decisions are important leadership skills.

Drivers of Women in Family Businesses

At the family level, the key drivers of women's entrepreneurial ingress in the family business are spousal support, culture and family tradition. Spousal support strengthens one to venture into the family business. Starting up a family business would be difficult without the support provided by the spouse. Spouse's support may be emotional, psychological, practical, or a combination of both (Blenkinsopp & Owens, 2010). Additionally, women are more likely to undertake and establish an entrepreneurial project with their husband if they both share a common vision, business goals, objectives, and risk approach (Van Auken & Werbel, 2006).

Gender Discrimination

Gender discrimination has got its own implications for the organizational outcomes and the way gender will be perceived at a particular place (especially the female) (Reskin, Barbara F. 2000). Scholars and the lay public alike recognize that women have significantly increased their participation in the labor force over the past half-century. There exists a hegemonic belief that men are supposed to earn, and females are supposed to take care of the needs of the home.

Trusted Advisors

Three types of advisors have been classified in the literature on trusted advisors in family businesses. They are: formal advisors (many of whom are trusted advisors), informal advisors, and family-firm boards (Strike, 2012). The firm or the family businesses hires the formal advisors. They hold either an external or an internal position. They can be either content experts or process-oriented consultants. Content experts are specialists within a specific area such as financing, law, tax, investments, or products. Process-oriented consultants are the ones who focus on implementing new structures or developing the firm. Both, content experts and process-oriented consultants can serve as trusted advisors.

People Associated with Decisions in Family businesses

Organizational researchers are increasingly drawn to family businesses because they reveal so many important models and theories: theory of the firm, family dynamics and human development, agency theory, labor market theory, economic development, social capital theory, and many others. Professional service providers and consultants are also attracted to family enterprises as clients. In fact, over the latest decades, there have been many more professionals working with family firms than studying them. In the field, the best opportunity to maximize our understanding of family businesses is to utilize all of these sources of data on family firms; that is, to integrate formal research and theory with documented implementation experience to build a platform for future development.

Life partner: Irrespective of the situation in family businesses the life partner is an important consultant and decision maker. They have a contribution in not only how the major decisions will be made, but they also help in absorbing the major shocks that the owner faces and that is the reason the familiness component of the family businesses talk about life partners as the ones who help in making decisions.

Venture partner: Venture partner is the one who is involved in a lot of decisions making that takes places on day to day basis for the firm. Some of them are the initial funders of the current firm, or they would have entered in between to join, support and gain out of the opportunity to be with the business and has a strong impact on the decision making part of the family business.

Expert: The expert is associated with the role of the advisors and as indicated in the earlier paragraphs the role of consultants and researchers associated with the family business is that of an expert. Many of the top managers who play roles in business and its decisions are experts in functional fields or domains, and thus, this helps the family business through the means of their direction and advice. These experts may not be full-time employees but they are the people who are associated with decision making in the family business.

Investor: Investors majorly act as the funding source. They may not have the expert functional knowledge, but they play their role in terms of capital support of the family, and they hold the transitional role of sometimes partners, sometimes experts and predominantly the role of investors only.

Community: Within the past literature of family business from 2013 onwards, family heterogeneity is also represented in terms of the community at local and state level, of which the business is a part of in terms of networks, associations. Moreover, they exert an impact on the decision making of a family business. The community is said to impact specifically in terms of movements and the values which the family business is associated with.

Employee: The clear focus of Agency theory has been on the employee who has an impact on decisions regarding family movement, successor, governance, and also towards the family designs.

To address the heterogeneity in family business, the researcher has included all the above-mentioned stakeholders as respondents in the survey conducted for this study.

Role congruence

Assumptions and expectations regarding the attributes and behavior of men and women in social settings are influenced and defined by Gender Roles. (Eagly & Karau, 2002; Heilman, 2001; Heilman, Block, Martell, & Simon, 1989; Rudman & Glick, 2001; Schein, 1973). Thus, it leads to various stereotypes about men and women. As per Western beliefs and societies, women are gender stereotyped as relationship-oriented. To build upon the various theory that are indicative towards role congruity we put them together through various lenses and also each of the components how they are different from role congruity are looked is presented in the table given below. To ensure congruence is difference from other work on gender and its implication is clearly established.

Table 3: Role congruity as per various theoretical lenses*

S. No.	Role congruity Related literature (Name of the paper)	Central theme of the paper on Role congruity	Part taken in for our study
1.	Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.	Popularity of the glass ceiling concept may start from the rare occurance of of women in leadership posts, despite the presence of indicators indicating somethng different for female leaders.	1. Role congruity is different from glass ceiling. 2. Its not just the " Pipeline problem" for female talent 3. Social roles are socially shared expectations that apply to persons who occupy a certain social position or are members of a particular social category
2.	Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371.	Social roles are socially shared expectations that apply to persons who occupy a certain social position or are members of a particular social category	1. Perceivers infer that there is correspondence between the types of actions people engage in and their inner dispositions. 2. The descriptive aspect of gender roles thus originates in perceivers' correspondent inferences from the observed behaviour of men and women to their personal qualities 3. Evidence abounds that gender roles also embrace injunctive norms about male and female behavior.

3	Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353-363.	The overall approval of communal qualities in women and agentic qualities in men has been demonstrated	1 The beliefs that people hold about ideal women and men 2 The beliefs that women and men hold about their ideal selves attitudes and prescriptive 3 Beliefs that people hold about the roles and responsibilities of women and men

*Table made on papers reviewed by researcher.

Succession Planning

The process of transferring and handing over the ownership and leadership from an incumbent to a successor from next generation is broadly termed as ‘Succession’. In a family business’s life cycle succession is considered one of the most important processes as accorded by the researchers. The reason for its importance is said to be its quintessential effect on firm’s culture, strategy, viability, and existence. (Handler, 1994; Ahlers, Hack, & Kellermanns, 2014; Cater III & Kidwell, 2014). Succession-planning process is primary and one of the most crucial parts of the overall succession process.

Affection

Bonds of affection, cooperation, unity, emotional connect, responsibility and loyalty among the family members strengthen the family into a system (Schneider, 1989). Concept and notion of need are the foundation of this system. The paramount social role and purpose of the family is to protect, care and support the family members. Certain behavioral processes are more important than the communication of affection. They are essential for the development, preservation, and continuance of deep, warm and devoted relationships.

Alliances

The words "alliance" and "coalition" are defined and used interchangeably in present-day research, dictionaries, and thesauri. Alliances are described as "a merging of efforts or interests by persons, families, states, or organizations”. “An alliance may apply to any connection entered into for mutual benefit" (Random House Dictionary, 1969). The cause and motive of an alliance may be to accomplish a goal. An alliance may or may not be formed against someone else, which will be called a "target" of the alliance.

Companionship

Increased productivity and cost-effectiveness are the organizational benefits of mentoring (Murray, 1991; Zey, 1984). Other benefits are: improved employee recruitment, development, and retention (Berube, 1996; Horgan & Simeon, 1990; Lawrie, 1987; Ragins & Cotton, 1993; Wright Wright, 1987). The type of the mentoring relationship can be formal or informal. Formal mentoring is managed and sanctioned by the organization. Informal mentoring is the one which spontaneously occurs without any external involvement from the organization. Research

substantiates that mentoring can be helpful for the members of family business (Goldberg, 1996; Simurda, 1995).

Instrumental Aid

Besides, business-related goals which are assessed by business researchers, family businesses are likely to pursue family-related goals. Moreover, they are optimistic to achieve a good performance in both the dimensions (Basco & Perez Rodriguez, 2009; Sorenson, 1999. Various such diverse findings can be explained through a general “socioemotional wealth” model developed by Gomez-Mejia et al. (2007).

Intimate Disclosure

It is essential for family businesses to create an atmosphere of trust and transparency among family members. Keeping the pay undisclosed is a comfortable stand for many family business owners. Nonetheless, nondisclosure does not prevent family members from developing strong emotions, suspicions, and beliefs about one another (Aronoff and Ward, 1993). In dynamic environments, success and failure can be decided by a family firm’s ability to manage its critical knowledge resources.

Nurturance

Very often, families and business have always existed in tandem to a large extent (Morck and Yeung, 2002; Narva, 2001). Generally, the primary motivation for starting and growing a business is the economic necessity of earning a living and supporting a family (Winter et al., 1998). Among other motivators, lifestyle and wealth accumulation goals are heavy influencers in deciding whether a particular family member or members choose to start a business in conjunction with their family. When collaborating with family then the business provides income to the family and the family provides a critical supply of paid and unpaid labor.

Reassurance of worth

Interaction of the firm and family creates a bundle of resources which are idiosyncratic and affect the activities and performance of family businesses. Studies of Fortune 1000 firms discovered that family firms owned by lone founders have higher performance, stronger entrepreneurial orientation, highly intense growth strategies, and greater investments in firm infrastructure as compared to family businesses with multiple owners. Additional family owners add diversity but complexity to the organizational governance and decision processes which is eventually difficult to manage under high levels of uncertainty, complexity, and ambiguity (Lambrecht & Lievens, 2008). These conditions generally exist in firms pursuing and investing in innovation.

Antagonism

Strong conflicts of interest between the incumbent and the following generation can threaten the process of succession. The conflict can be regarding ownership, continuity, control of assets, power distribution within the family, and in attaining legitimacy from the incumbents, family members, and owners, non-family executives and other external stakeholders. Firms are embedded in a social context. The next-generation entrepreneurs within family firms are interjected into a prevailing network structure.

Conflict

There exist two general approaches to the concept of "conflict." According to Schmidt and Kochan (1972), the first approach accentuates upon phenomena linked to competitive intentions, such as deliberate interference with the other's objectives. This definition tends to be closer to the generally acknowledged usage of the term "conflict". It is particularly popular in the industrial relations literature discussing strikes and job actions. According to Pondy (1967), the second approach has been to embrace more general definitions. These definitions move "upstream" in the conflict procedure to involve events (usually a party's perceptions) which occur before the choice of conflict-handling modes.

Relative power

The mode of operation by CEOs depends and varies with the context of their relationships with managers, directors, major customers, suppliers, and others. The aforementioned context is referred as their "social embeddedness." Social embeddedness is a function of formal arrangements, for instance, position(s) and extent of shareholdings. It can also be a function of informal factors like age, length of tenure, and education. Along with it, comes a set of contacts, expectations, obligations, social reciprocity norms, and interrelationships with the other actors.

Research Gaps

Extensive literature review of the area of "Study of Gender Role Congruity in Family businesses" indicate towards lack of study in the area of role and its congruence in family businesses. It is evidently clear that family business studies have not been looked upon from multiple people perspective and there is lack of studies when it comes to contribution of interactions between the people. Multiple discussion and future research sections of various journals discussing family business review directs that the research on role congruity with multiple people perspective should be encouraged. Thus, the research gaps can be summarised as:

1) The presence of role congrity in family businesses is not studied but is suggested as point of study in the literature.
2) Gender's role in Social provisions in family businesses is not studied.
3) Gender of the successor plays a moderating role on certain qualities of the person within the family business. Thus, those certain qualities need to be identified.

Despite several articles that have focused on clarifying the definition of a family firm and developing scales to measure the extent of family influence in a firm, hardly any efforts have been directed to either define or measure the 'family' variable (Pearson et al., 2014). Even in scales aimed to measure the family climate or family harmony or family influence, the task of defining the 'family' is left to the respondent, leaving open the possibility of multiple interpretations of a key term in the same data set. The 'family' variable cannot be restricted to any single aspect. Thus, the below-mentioned table 4 provides a glimpse from literature review regarding the understanding of family units.

Table 4: Literature review indicating family variable and the three provisions

	Life course	Systems	Symbolic interactionism	Social exchange
Source and	It is derived	It is derived from	It is derived	It is derived from

Origin	from family development theory. It arose in the 1960s. The early theorists include Hill and Rodgers (1964) and Elder (1974).	general systems theory. It arose in the late 1950s. The early theorists include Kantor and Lehr (1975) and Bateson et al. (1950) and x.	from American pragmatism. It arose in 1920s; The early theorists include Mead (1934), Cooley (1956) and Blumer (1969).	market concept of exchange. It arose in the late 1950s. The early theorists include Thibault and Kelly (1959) and Homans (1961).
Core Constructs	Life events, transitions, and paths; historical context; linked lives between family members.	Interaction patterns within the whole family system (e.g., conflict, communication); self-reflexivity	Symbols; values; behavioral norms; shared meanings; family identity; self-concept.	Resource exchange; social rewards and costs; self-interest; rationality; reciprocity; interdependence.
Central Premises	The life paths of individuals are influenced by the social ecology and historical context in which they live	The interconnections between family premise members impact the functioning of the family system	Interactions between family members create shared identities through common symbols (e.g., situations and values).	Family relationships entail a rational exchange of resources (social and economic) to maximize rewards and minimize costs.
Key assumptions	The timing and context of key events and transitions during an individual's life course are important	Individual behavior is best assumptions understood through an analysis of the family system	Individuals develop concepts of self and identity mainly through interactions in the family.	Individuals are rational and motivated by self-interest in relationships.

Research Objectives

Synthesis of the research and studies conducted symbolizes that although gender has been shown to impact leadership practice, the true impact of gender on the roles which female may step in still remains somewhat difficult to completely discern. Owing to the complexity of the impact of gender, some have contended that there is no theoretical basis for examining the

issue. On the other hand, few others aver that gender differences must be assessed to determine how role acceptance develops for women as compared to men. Nevertheless, certain challenges existed with regard to determining the role that gender plays in identifying the impact, but evidence suggests that gender influences peer interaction of female leader with colleagues. Role congruity theory provides certain explication for understanding the influence of gender on outcomes for female leaders. In the analysis of dominant theoretical approaches in the research of women in management, Fagenson (1990) acknowledged and combined two predominant gender perspectives to contend for a third perspective namely, gender- organization-systems approach. The two predominant perspectives are a person-or gender- centered perspective, and an organizational perspective which is also what we are trying to focus through our research statement and the research objectives are centered around it.

Problem statement:

1. Does role congruity exist in family businesses where now females are also becoming successors?
2. What is the role of gender while looking at provisions of Society in family businesses?
3. What nature of dynamics comes in interplay when we are talking about heterogeneous responses for roles in family business.

RESEARCH METHODOLOGY

Research Design

For the purpose of this study, we are focusing on Cross-Sectional research design. In a cross-sectional study, the data gathered is from a pool of participants with varied characteristics and demographics known as variables. Age, gender, income, education, geographical locations, and ethnicity are all examples of variables. The variables, or demographics, used in a single study are based on the type of research being conducted and on what the study aims to prove or validate.

Cross-sectional studies are innately clear and assists in examining a large number of variables. The researcher can standardize the methods, and clear-cut definitions can be applied to the exposure and the endpoints.

1. It is comparatively quick and easier to conduct with no long periods of follow-up.
2. The data on all variables is collected only once.
3. The investigator is able to measure prevalence for all factors under investigation.
4. It helps in studying multiple outcomes and exposures.

Sample and Data Collection Procedure/Sampling design

Family businesses are diverse in nature. The simplest form of family business is the one which might be housed and running from one single room and still sustain for a substantial time. Thus, it is important for us to define the criteria for the sample, which will be used as a sample to identify the population for our study. The sample was identified based on the following criteria:

1. It should be family business which has continued for at least 30 years.
2. There should be a minimum two generations which should have worked in the family business, and the family business should have continued for that much time.
3. The business should have a turnover of at least INR 30 lakhs per year.

4. There should be succession which should have happened in the business.
5. There should be minimum 8 nonfamily members who should be working as employees in the organization.

One of the key reason behind this type of sample size was based on the criteria that are suggested in methods of family research. It suggests defining the sample size criterion specifically to avoid any arbitration or error that might be possible in research. If the business fulfilled all the above-mentioned criteria then only they were included in the sample. The nature of sampling design we followed is purposive sampling. Also as indicated in literature to look at heterogenity of the sample, we need to have criteria which will lead to purposive sampling (Bjornberg, A. & Nicholson, N., 2007). Thus, it suits our study.

Data Sources

We have used primary and secondary data sources to complete our study.

Primary Data source: Data was collected from family businesses using the tool and as per the research design. (Sample size of 218 Family businesses of Gujarat)

Secondary Data source: Literature review of journals, Hand book and articles from various sources.

Tool for Data Collection

Based on the above literature and the gaps identified, we were able to understand that for this research we will have to adapt from Furnman's questionnaire on Interaction provisions because we want to observe all nature of interactions towards the family business successor, and also to obtain responses from the people who will be associated with the family successor. The input from literature review and the integrative theory offered a significant answer to the point that whose responses should be considered for the family successor. Therefore, on the basis of review and Furman's work, we identified Life Partner, Expert, Venture Partner, Investor, Community and Employee as the perfect respondents for our work. They represented the multiple people interaction and their responses formed the nested data for the study which made our work much easier to observe and understand.

Pilot study

The Pilot study was conducted in the year 2018 to look at the types of samples emerging for such study and to know if the nature of responses are in line with the expectations. The pilot study was conducted for 30 family businesses and the responses were calculated. The results were approved by Research Guide, Research Defense Committee and the Experts (who had validated the instrument for data collection).

Data collection

A structured questionnaire which looked at social provision was utilized for the study. Each question was prepared for collection of data after conducting a pilot study on Family businesses to pretest the questionnaire. The results and questionnaire were shown to experts for validation and also presented as work in the conference to obtain inputs for further validation and remarks.

The pretesting of the questionnaire was piloted among respondents similar in profile to the final sample of this study. The comments and the feedback from respondents were taken into consideration, and finally, adjustments were incorporated in the questionnaire. The primary data was collected from 354 respondents out of which 136 were rejected after scrutiny, due to incomplete or missing details in the questionnaire.

The primary data for this research study was collected from respondents residing in four cities of Gujarat - Surat, Rajkot, Ahmedabad, and Vadodara. These cities were chosen to geographically represent the state of Gujarat and also, on the basis of the verification of the data on family businesses in Gujarat from CII database, EDI Ahmedabad and Associations of Businesses of the cities. The number of respondents was chosen proportionately from each of the four cities based on the existing number of family businesses in these cities. Based on this proportionate population, the respondents were chosen from the respective cities. The method of sampling used for this research study is Purposive sampling as the sample had to match certain criteria to be considered for the study. The questionnaire method was adopted to collect the primary data from the respondents. The respondents were administered with a structured questionnaire in all four cities of Gujarat and were requested to fill up the questionnaire based on their judgments and beliefs. The reliability of the scale which is a measure of how a scale used in the questionnaire could be relied on to produce similar measurements every time the scale is used, was done through SPSS. Cronbach Alpha was employed to perform Reliability analysis of a scale that consists of various items/variables used in the study. The alpha value is considered a good scale if the value of the scale is 0.7 or more (Nargundkar, 2008; Malhotra, 2004). The value of Cronbach Alpha is shown for each quality included in the research study. The data collected has been from both primary and secondary sources for realizing the objectives of this study. The secondary data for understanding the gaps, current work, and awareness to look at the trends in the research topics, it has been collected and reviewed from various sources such as books, periodicals, journals, research reports, magazines along with searching the internet through Google and Google Scholar. All the sources are appropriately outlined in the bibliography section of the thesis.

Table 5: Cronbach alpha calculation for various factors

Scale	Life partner N=218	Venture Partner N=218	Expert N=218	Community N=218	Investor N=218	Employee N=218
Companionship	0.89	0.89	0.84	0.81	0.76	0.78
Instrumental Aid	0.89	0.89	0.83	0.89	0.86	0.91
Conflict	0.72	0.74	0.73	0.86	0.8	0.75
Antagonism	0.81	0.9	0.8	0.83	0.78	0.83
Intimate Disclosure	0.79	0.74	0.65	0.82	0.76	0.71
Nurturance	0.77	0.8	0.7	0.76	0.72	0.71
Affection	0.72	0.77	0.82	0.73	0.69	0.69
Reassurance of worth	0.71	0.72	0.759	0.69	0.67	0.77
Relative Power	0.93	0.95	0.93	0.91	0.89	0.9
Reliable Alliance	0.84	0.84	0.86	0.88	0.82	0.8

Note: Values are Cronbach Alpha Coefficients (N= Defined for each category). Participants were involved in Family businesses in the defined roles in Gujarat (Surat, Rajkot, Ahmedabad, and Vadodara) between 2015 to 2017 March.

Results and their brief description

The results are mentioned briefly described in table 6. The table shows that first the MANOVA was proved for the particular provision and later it was followed by the quality being tested on each provision for each of the roles. The results are specified in bold where the hypothesis is accepted.

Table 6: Hypothesis testing using MANOVA

		Role-Holders					
S. No	Hypotheses Statements	Life Partner	Venture Partner	Expert	Investor	Community	Employee
	Role Congruence Hypothesis (Through MANOVA)						
1	H1 : Role congruence exist because of gender for the successor	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**
	Quality Based hypotheses under Provision (Through ANOVA)						
1	H1 : Companionship is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
2	H2 : Intimate disclosure is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
3	H3 : Instrumental aid is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
4	H4 : Reassurance of worth is significant for the Role for the successor	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**
5	H5 : Reliable alliance is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
6	H6 : Affection is significant for the Role for the successor	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**
7	H7 : Conflict is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
8	H8 : Antagonism is significant for the Role for the successor	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**
9	H9 : Relative Power is significant for the Role for the successor	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
10	H10: Nurturance is significant for the Role for the successor	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**	**Accepted**

Note: The ‘Role’ denotes the 6 profiles namely, Life Partner, Venture Partner, Employee, Expert and Community of Entrepreneur for whom hypotheses were tested.

First, the main hypothesis was tested for the part that whether role congruence exists for the particular role respondent or not, using MANOVA. It was followed by individual-level ANOVA

test for each quality against the particular role. Wilk's Lambda was identified for the MANOVA test and ANOVA test. M value was observed to point out the significant differences for each of the gender and its significance. Presence of role congruity was proven, and also it was proved that Reassurance of Worth, Affection, Antagonism, and Nurturance are the qualities which show their presence on individual roles and congruity. Thus, using this evidence the model for Social Provisions was created.

MANOVA**:** A MANOVA was conducted with an alpha level of 0.05.

Table 7: MANOVA Values for role-holders	
	MANOVA Values
Life Partner	0.000
Venture Partner	0.001
Expert	0.001
Investor	0.000
Community of Entrepreneur	0.000
Employee	0.020

ANOVA: ANOVA was conducted with an alpha level of 0.025.

Table 8: ANOVA values for various role-holders on Social provisions						
	Life Partner	Venture Partner	Expert	Investor	Community of Entrepreneur	Employee
Companionship	0.525	0.705	0.239	0.616	0.852	0.232
Conflict	0.177	0.636	0.925	0.843	0.741	0.644
Instrumental Aid	0.285	0.802	0.645	0.579	0.770	0.593
Antagonism	**0.006**	**0.006**	**0.003**	**0.002**	**0.008**	**0.021**
Intimate Disclosure	0.242	0.615	0.391	0.630	0.457	0.400
Nurturance	**0.024**	**0.024**	**0.016**	**0.029**	**0.024**	**0.023**
Affection	**0.019**	**0.006**	**0.016**	**0.011**	**0.018**	**0.019**
Reassurance of worth	**0.014**	**0.011**	**0.014**	**0.019**	**0.015**	**0.021**
Relative Power	0.170	0.506	0.772	0.986	0.375	0.850
Reliable Alliance	0.682	0.456	0.310	0.339	0.125	0.296

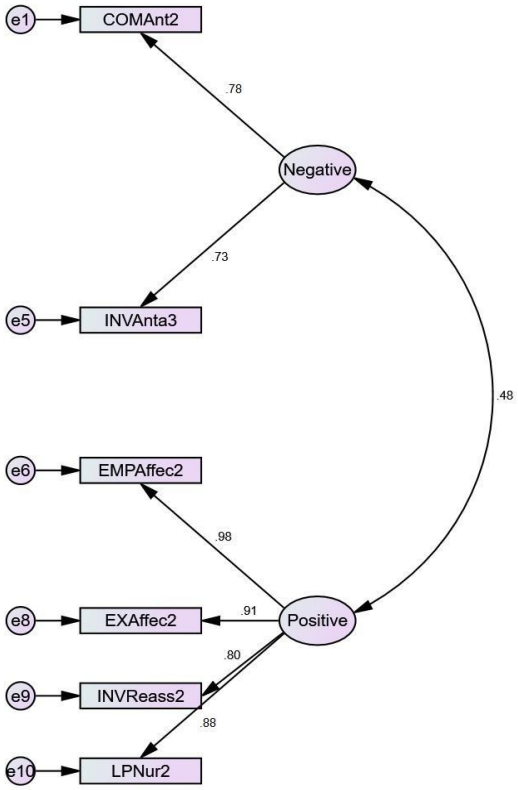
MODEL FOR SOCIAL PROVISION

In reviewing the Model diagram 1 it can be seen that there are 2 unobserved latent factors and 6 observed variables.

The 2 unobserved latent factors are Positive and Negative. The 6 observed variables namely, Community of Entrepreneur Antagonism (COMAnt2) & Investor Antagonism (INVAnta3) function as indicators of latent factor- Negative whereas, Employee Affection (EMPAffec2), Expert Affection (EXAffec2), Investor Reassurance of Worth (INVReass2) and Life Partner Nurturance (LPNur2) function as indicators of latent factor- Positive. Associated with each observed variable is an error term (e1, e5, e6, e8, e9, e10). Errors associated with observed variables represent measurement error, which reflects on their adequacy in measuring the related underlying factors.

Certain symbols are used in path diagrams to denote hypothesized processes involving the entire system of variables. In particular, one-way arrows represent structural regression coefficients and thus indicate the impact of one variable on another. Likewise, the two unidirectional arrows leading from 'Negative' to each of the 2 observed variables (COMAnt2 & INVAnta3); suggest that these score values are each influenced by their respective underlying factors. As such these path coefficients represent the magnitude of expected change in the observed variables for every change in the related latent variable (or factor). The one-way arrows pointing from the enclosed error terms indicate the impact of measurement error on the observed variables.

Diagram 1: Model for Social provision*****



***** Diagram based on AMOS output created using SPSS 22 & AMOS 22

For negative qualities, the model shows that **antagonism from community** and **antagonism from investor** are two strong **negative qualities** under social provisions. For **positive qualities, affection from employees, affection from experts, reassurance of worth by investor** and **nurturance by the life partner** are the qualities which stand out robustly for the successor as the qualities under social provisions.

Moderation

Negative and ComAnt2 (Community's antagonism is moderated by gender)

To test the hypothesis that Community of Entrepreneur's response to negative quality for a successor is moderated by gender. As per the model generated, there can be effect on more qualities under social provisions, but more visible effect is observed on Community of Entrepreneur through Antagonism (COMAnt2) as a Negative quality, for which a hierarchical multiple regression analysis was conducted. In the first step, two variables were included: COMAnt2 and Gender. They account for a significant proportion of variance in Negative factor, $R^2 = .757$, $F(3, 214) = 221.82$, $p < .001$. To avoid potentially problematic high multicollinearity with the interaction term, the variables were centered, and an interaction term between COMAnt2 and Gender was created (Aiken & West, 1991).

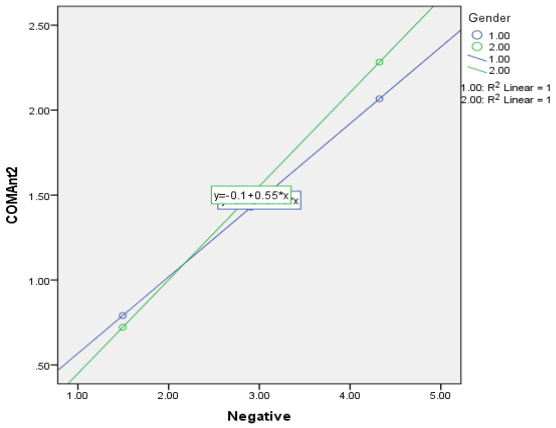
Next, the interaction term between COMAnt2 and Gender was added to the regression model, which accounts for a significant variation in the Negative Factor, $\Delta R^2 = .0076$, $\Delta F(1, 214) = 6.69$, $p = .0104$, $b = .101$, $t(214) = 2.59$, $p < .01$. Examination of the interaction plot showed that for Gender 2, i.e., female as a successor, antagonism which serves as a negative quality is expressed more highly by the Community of Entrepreneurs as compared for Gender 1, i.e., male.

Thus, it is observed that Community of Entrepreneur's response to negative quality for a successor is moderated by gender as it is more obvious in case of females. Therefore, we can say that in case of female successors it is evident that antagonism is expressed against them by the Community of Entrepreneur as compared for male successors.

Table 9: Model Summary for Negative and ComAnt2******

```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .8699      .7567      .1640   221.8243     3.0000   214.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant      .3362      .1975     1.7021      .0902     -.0531      .7256
Gender       -.2191      .1257    -1.7425      .0829     -.4669      .0287
Negative      .3503      .0606     5.7775      .0000      .2308      .4698
int_1         .1007      .0389     2.5862      .0104      .0239      .1774


R-square increase due to interaction(s):
         R2-chng          F        df1        df2          p
int_1      .0076     6.6884     1.0000   214.0000      .0104
```

Diagram 2: Plotting moderation effect of gender on Negative and ComAnt2



********** Diagram ad table based on AMOS output created using SPSS 22 & AMOS 22

It is observed that **Community of Entrepreneur's response to negative quality for a successor is moderated by gender and it is more obvious in case of where it is evident that higher Antagonism** is expressed against them by the Community of Entrepreneur as compared for male successors.

Negative and InvAnta3 (Investor's antagonism is moderated by gender)

To test the hypothesis that Investor's response to negative quality for a successor is moderated by gender. As per the model generated, there can be effect on more qualities under social provisions but more visible effect is observed on Investors through Antagonism (INVAnta3) as a Negative quality, for which a hierarchical multiple regression analysis was conducted. In the first step, two variables were included: INVAnta3 and Gender. These variables account for a significant amount of variance in Negative factor, $R^2 = .759$, $F(3, 214) = 225.14$, $p < .001$. To avoid potentially problematic high multicollinearity with the interaction term, the variables were centered, and an interaction term between INVAnta3 and Gender was created (Aiken & West, 1991). Next, the interaction term between INVAnta3 and Gender was added to the regression model, which accounts for a significant variation in the Negative Factor, $\Delta R^2 = .0075$, $\Delta F(1, 214) = 6.69$, $p = .0104$, $b = .101$, $t(214) = 2.59$, $p < .01$. Examination of the interaction plot showed that for Gender 1, i.e., male as a successor, antagonism which serves as a negative quality is expressed more highly by the Investors as compared for Gender 2, i.e., female.

Thus, it is observed that Investor's response to negative quality for a successor is moderated by gender as it is more obvious in case of males. Therefore, we can say that in case of male successors it is evident that antagonism is expressed against them by the Investors as compared for female successors.

Table 10: Model Summary for Negative and InvAnta3******

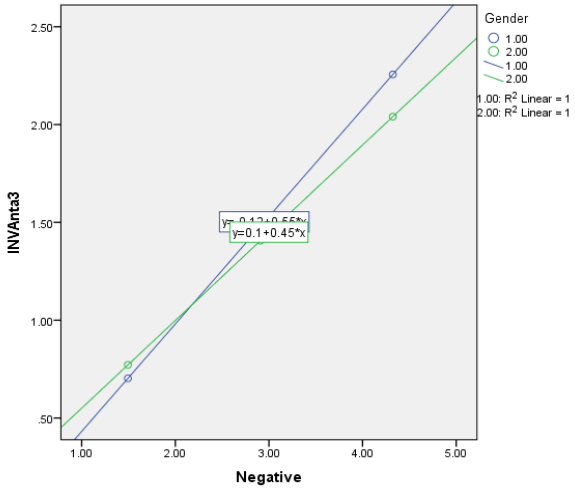
```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .8714      .7594      .1640   225.1441     3.0000   214.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
```

constant	-.3362	.1975	-1.7021	.0902	-.7256	.0531
Gender	.2191	.1257	1.7425	.0829	-.0287	.4669
Negative	.6497	.0606	10.7159	.0000	.5302	.7692
int_1	-.1007	.0389	-2.5862	.0104	-.1774	-.0239

R-square increase due to interaction(s):

	R2-chng	F	df1	df2	p
int_1	.0075	6.6884	1.0000	214.0000	.0104

Diagram 3: Plotting moderation effect of gender on Negative and InvAnta3



********** Diagram ad table based on AMOS output created using SPSS 22 & AMOS 22

It is observed that Investor's response to negative quality for a successor is moderated by gender and it is more obvious in case of males where it is evident that higher Antagonism is expressed against them by the Investor as compared for female successors.

Positive and EmpAffec2 (Employee's affection is moderated by gender)

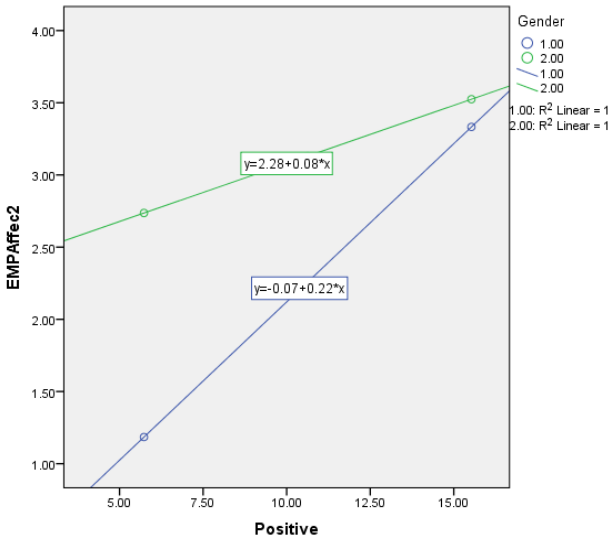
To test the hypothesis that Employee's response to positive quality for a successor is moderated by gender. As per the model generated, there can be effect on more qualities under social provisions, but more visible effect is observed on Employees through Affection (EMPAffec2) as a Positive quality, for which a hierarchical multiple regression analysis was conducted. In the first step, two variables were included: EMPAffec3 and Gender. These variables account for a significant variation in Positive factor, $R^2 = .449$, $F(3, 214) = 57.99$, $p < .001$. To avoid potentially problematic high multicollinearity with the interaction term, the variables were centered, and an interaction term between EMPAffec3 and Gender was created (Aiken & West, 1991). Next, the interaction term between EMPAffec3 and Gender was added to the regression model, which accounts for a significant variation in the Positive Factor, $\Delta R^2 = .0562$, $\Delta F(1, 214) = 21.79$, $p = .000$, $b = .1388$, $t(214) = 4.69$, $p < .01$. Examination of the interaction plot showed that for Gender 2, i.e. female as a successor, affection which serves as a positive quality is expressed more highly by the Employees as compared for Gender 1, i.e., male.

Thus, it is observed that Employee's response to positive quality for a successor is moderated by gender as it is more obvious in case of females. Therefore, we can say that in case of female successors it is evident that affection is expressed for them by the Employees as compared for male successors.

Table 11: Model Summary for Positive and EmpAffec2******

```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .6697      .4485     1.1370    57.9993     3.0000   214.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant    -2.4190      .5322    -4.5453      .0000    -3.4681    -1.3700
Gender       2.3474      .3478     6.7498      .0000     1.6619     3.0329
Positive      .3581      .0455     7.8726      .0000      .2684      .4477
int_1         .1388      .0297     4.6689      .0000      .1974      .0802


R-square increase due to interaction(s):
         R2-chng          F        df1        df2          p
int_1      .0562    21.7991     1.0000   214.0000      .0000
```

Diagram 4: Plotting moderation effect of gender on Positive and EmpAffec2



********** Diagram ad table based on AMOS output created using SPSS 22 & AMOS 22

It is observed that Employee's response to positive quality for a successor is moderated by gender and it is more obvious in case of females where it is evident that higher Affection is expressed for them by the Employees as compared for male successors.

RESULTS

Model for Social Provision

When we look at the structural model path diagram and the values, we get a clear indication for the nature of qualities that affect the decision-makers in his or her role. The model clearly denotes that the part is differentiated into two qualities, i.e., positive and negative qualities. As stated earlier, we are not limiting our self to look solely at the qualities prevalent for a successor, but we are also looking at how these qualities are positive and negative and secondly, which qualities prominently differ due to gender orientation (male or female). Furthermore, we want to present a model which is more concrete than a conceptual model and which will specifically indicate the part of the provisions, specific qualities and the respondents associated which comes into play. It is essential to prove the specificity in provisions. The model evidently indicated that there are two nature of qualities which are affecting the successor and his/her role as perceived by the respondent.

For negative qualities, the model shows that antagonism from community and antagonism from investor are two strong negative qualities under social provisions. For positive qualities, affection from employees, affection from experts, reassurance of worth by investor and nurturance by the life partner are the qualities which stand out robustly for the successor as the qualities under social provisions. Family businesses face a unique set of challenges in prioritizing which stakeholder group matters most and understanding how families give priority to competing stakeholder claims. This has both important theoretical and practical implications. As mentioned before, the model provides some interesting insights in this direction. By building on the individual dimensions of social and emotional wealth, the researcher propose that the different weights that families place on these dimensions may account for family firms' heterogeneity in identifying, prioritizing and responding to different stakeholder groups.

This study broadly distinguishes between primary and secondary stakeholders. Empirical testing of this idea has led to finer categorization of stakeholder groups are needed to capture differences in stakeholder salience or, in other words, 'who really counts' (for family business owners. Moreover, the uniqueness of family firms may alter the bases of stakeholder power, legitimacy and urgency that define stakeholder salience. Compared to non-family businesses, stakeholder power in family firms is more likely to be based on prestige and social symbols such as love or acceptance, legitimacy refers to status derived from birth and close relationships, and urgency is biased toward socioemotional wealth preservation goals.

Moderations Results

This section describes the result of moderation(s). The moderations' result is described factor by factor to emphasize the point that gender strongly impacts roles and qualities which the respondent associates with them.

Negative and ComAnt2 (Community's antagonism is moderated by gender)

It is observed that Community of Entrepreneur's response to negative quality for a successor is moderated by gender and it is more obvious in case of where it is evident that higher Antagonism is expressed against them by the Community of Entrepreneur as compared for male successors, which represents one idea that within the family businesses the males facing antagonism is actually true and more support exist if females take up the same role.

Negative and InvAnta3 (Investor's antagonism is moderated by gender)

It is observed that Investor's response to negative quality for a successor is moderated by gender and it is more obvious in case of males where it is evident that higher Antagonism is expressed against them by the Investor as compared for female successors. Investors are critical and they are as indicated in earlier models are very cautious on their involvement with the role associated and they take care of it by being on a thin line of appreciation or pressure, but the moderations is dependent on the gender which is indicative of the congruency in case of the roles.

Positive and EmpAffec2 (Employee's affection is moderated by gender)

It is observed that Employee's response to positive quality for a successor is moderated by gender and it is more obvious in case of females where it is evident that higher Affection is expressed for them by the Employees as compared for male successors. Employee behaviour is representation of multiple things and this when this is moderated by gender irrespective of the provision it is a good indication of family businesses being more adaptable and open to the newer age of leadership.

Results Discussion

While evaluating a family business successor, it is suitable to obtain a response from multiple people who are associated with the family business because it ensures that the responses are not biased based on any single nature of the response. Secondly, it is important to observe the part that these roles can comprise of multiple profiles which depends on the nature of the business and the associated role. For e.g., the term Life Partner cannot be restricted to a spouse, but it can comprise mother or father. Likewise, Investor can be an institutional investor like banks rather than being limited to an Investor. The Expert can be a mentor or a professor, and their responses and expertise may exist on very different sides. The research is reflecting one important point that role is not just about any particular quality and provision associated with quality.

The role in family business is affected by social provisions. There are multiple qualities which are associated with social provisions which may further define a single quality, for instance, positive or negative as in our study. Thus, social provision has 10 qualities. In this study, we have observed that provisions comprise of both, negative and positive quality rather than looking at each provision only through any one quality. Additionally, the clear-cut segregation of positive and negative qualities lead to a more clear understanding.

One needs to take into account that negatives qualities are not necessarily associated with female successors who are being evaluated and it is quite possible as shown in the results that quality like antagonism can be associated with male successors more strongly as compared to female successors. The idea of having a female successor is acceptable. When we look at the responses of community, expert and investors, their rating towards female successors are higher as compared to their rating for male successors, and this is indicative of the acceptance of female roles in the family business. Previous studies have predominantly solely focused on the point that there is a difference between females or males when it comes to family business. Thus, the contribution of this study is relevant to this particular aspect that there exists a difference but the difference is exactly known in addition to the reason of such a difference which is found to be as role congruity dependent on the gender of the successor.

One major revelation from the study is that the response of various groups/ people associated with family businesses change and does not remain constant. The results denote the same for instance, community's response to the role of family business successor is negative for social provision. This is a strong finding in the study that when it comes to nature of the response, the responses can vary and might not be the same which is clearly shown for a life partner, venture partner, expert, investor, community and employee under all three provisions. Thus, this point in the study provides us with leverage. The leverage is advantageous as it helped us to create models which indicate that the qualities can be segregated as both positive and negative and also, that they act differently for which moderation has been conducted.

The models are symbolic of the heterogeneity of the family business units and how their mechanism cannot be indicated by a single theory or development. It is one unit whose suitability is based on multiple factors when examined from the human resource aspect. The generalizations of gender cannot be applied in family businesses which means that a particular stereotype formed because of gender cannot be true for all family businesses. There are various underlying factors which have been specified in the study to reflect that. Another important milestone that the study has been able to undertake is to look at the moderation effect of gender on the qualities that are identified based on the model that was proven. The moderation for effect of gender on qualities strengthens the purpose of this study that role congruence is affected by gender and the response for the role of decision maker changes as per the successor's gender.

Managerial Discussion

There are a few important points that a manager may like to derive from this particular study. Many studies have indicated that the family businesses need to be examined from the performance point of view and also the task for which the manager has moved into the family business. But family businesses have a softer side which is dependent on various relationships that affect the family business. The manager who will be associated with any nature of work with a family business cannot solely focus on one person and also, limit the focus on any specific task he/she might be associated with. Through the means of this study, the manager will realize how emotions, power, education, time, roles, etc., all are affecting the family business. The context of family business is not just family, but it is also the society, socio-economic aspects, communities and the contextual spaces where the family businesses are working. One may obtain confirmation from this study that expert, investor, employee, life partner, community of entrepreneur- all will have their own role to play when it comes to interaction with the particular role within the family business ecosystems.

Birth order is not a strong factor which is affecting the family business. The variable of gender has a strong role to play within the family businesses, and one important point for the managers in this situation is that it leads to strong emotional ties with certain people. This response will be affecting various alliances and the way they contribute towards the future growth of the family businesses. This is indicative through the results with the life partners where as much they are important for the role of nurturance they contribute in terms of feedback as well.

Behaviors within the family business are not inflexible, and they are not as rigid as what is reflected in multiple studies. Interactions with other people within the family business reduce anxiety, and their behaviors become more acceptable and contributing towards the family business. Multiple people associated with the role of decision maker in family business do not lead to confusion, there are good chances that it may lead to more comfort and support for the

members working in the family businesses. This is clearly indicative of the roles that the expert and investors are playing within the family businesses. They are not only in controlling roles but also in the roles of getting things done. Tomorrow's managers and successors have to plan out how will they be planning out these areas for any system of family business management as they are not acting out in silos, they are interacting and affecting every decision and relation in family businesses.

CONCLUSION

Family business remains an area which has been ordinarily studied within the scope of governance, successions, and leadership. Development has been observed in the last few years which indicates that family business roles are not dependent only on a few factors of demographics. In addition to that, the earlier factors are not completely describing the nature of complexity that runs in the family businesses. There is a much wider scope in this area which can be studied through different methodologies.

1. We have tried to explore the topic on similar grounds by picking four major indicated limitations in the research of family businesses by looking at the family business unit as a variable and studying it using the quantitative method. Thus using social provision to understand various interactions and also been able to propose model based on our analysis.
2. In our study, we have accepted the challenges of heterogeneous responses for the same roles. We have tried to understand the roles not just by one dimension but with multiple dimensions of society, and social provisions.
3. We have examined the role congruence happening because of gender and the various interaction taking place within the organizations due to gender's effect on roles. The difference between family and a group of individuals is similar to the analogy of the difference between a house and a mound of bricks.
4. In the early stages of our field development, researchers have focused on structure (the bricks), but our field has now developed to the point where we have entered stage two of development where we investigate processes (the cement). Thus our study is strong contribution on the sides of the interaction process that takes place.
5. The stage of investigating the richness and dynamics of longitudinal process patterns and trajectories is the future direction of the family business field. Our study has been able to focus on network and the part of interactions which exist between various units.
6. The research contributes on the side where Gender inequality or inequity involves the assumption and/or practice of unequal division of authority and resources on the basis of sex, including ownership, leadership, and participation in the decision-making process. As the role of women changes, there is a reference to the impact of this change on men, and a direct reference to their organizations, which must change as a result and the results of this study proves it on that side.
7. The gender roles refer to the common roles of activity found in business (leader, successor, owner, etc.) in terms of societal expectations about what men can and should do, and what women can and should do. This research extends this further and proves that actually it is the interplay of relations, behavior, and society together. Thus, this strongly contributes to role congruence theory and its presence in family businesses.
8. The other idea it contributes on is gender identity which refers to how individuals define themselves as being a woman or man, according to what it means to them to be a woman

or man; how it influences their selves and their lives. We have proved the embedded concept that: how successors see themselves in reference to their families, particularly their relations (i.e., family identity- social provision).

9. Our study has certain limitations in terms of the respondent's geography (from the state of Gujarat) and from specific nature of family businesses only, but this limitation is essentially the base of future research work, where researchers may test the same provisions on a larger scale and also on a longitudinal basis.
10. Moreover, the study can be used to look at networks within the family businesses and the nature of unique interactions can be proved by using NodeXL. We think that this research has provided a base for further research to work on family businesses systems when data is taken for a single point from multiple sources, and in such a case the nested data will give more results which may provide better insights.

We have been able to prove things on the side of role congruence, heterogeneity and interactions but the idea is that further research work can be done by taking the tenets of expectation states theory, where one can analyze a leading model of family business succession: maybe model of successor commitment to look at the constructed view of gender shifts and opens up points of view, new realizations, and questions related to the attitudes, perspectives, and behaviors of successors within their environments. We are left, we believe, with a stronger model and new research questions to explore. The research illustrates practically how gender can be integrated into family business study either as a focal point or as a secondary but serviceable variable and/or context.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE
Online shopping behaviour: Predicting consumer's purchase intention

Shameem Shagirbasha
Great Lakes Institute of Management
Email: shawaja@gmail.com

ABSTRACT

With the penetration of internet, the online marketplace has seen a boom like never before. E-commerce businesses spend large sums of money to acquire and retain customers by providing tailor made promotions to online customers. Research has been done in this regard, from a customer-oriented and a technology-oriented view. Our research focuses on the latter (technology-oriented view). The approach suggested in the paper analyses the relationship between online shopping experience and purchase conversion with the aid of Google analytics metrics.

KEYWORDS: Online behaviour, Google Analytics, Purchase, Regression, artificial neural network.

INTRODUCTION

Google analytics metrics (Clifton et al., 2012) have been employed to evaluate usability of websites, identify problem areas (Francis et al., 2010; Hasan et al., 2009) and develop critical insights for businesses which can be used for improving customer satisfaction and loyalty (McFadden et al., 2005; Malacinski et al.,2001). These metrics directly translate to factors such as website content, website quality, website usability, waiting time and navigational pattern. These factors have been explored (Hasan et al., 2016; Gao et al., 2010; Hausman et al., 2009; Gao et al., 2006; GG Lee et al., 2005) and the ways to enhance them have also been studied (Wells et al.,2011; Chen et al., 2002). The purchase and reuse intentions have been modelled as a function of features contained in a website. Furthermore, there have been studies that has also emphasized both utilitarian and hedonic characteristics, which are referred to as computer and human factors and the individual perceptions and attitudes toward website use for online shopping (Kukar-Kinney & Close, 2010).

Our approach is based on the functional and visible design factors of websites such as the visual, informational and navigational design characteristics of online portals and how they affect the perception of online consumers. Studies have shown how poorly designed websites impede navigation of the website, searching for products and thus in turn affect the purchase behaviour (Gao et al., 2010; Gao et al., 2006). Thus, sellers should maintain quality websites as consumers rely on website quality as an extrinsic signal and use it to relate to perceived product quality (Wells et al., 2011). Ways for creating a better shopping experience like making users feel comfortable, alluring them to spend more time and revisit, increasing chance of purchase by making websites easy and fun to use have been suggested (Gillenson et al., 2002; Chen et al., 2002).

LITERATURE REVIEW

This paper presents the following as the key factors that drive the purchase intention of online consumers which directly relates to revenue generation from the customer. This brings us to the following hypotheses:
In the context of online shopping, ease of use refers to the customer's perceptions that shopping on the internet will involve minimum of effort. Considering that perceived ease of use attributes to

customer not requiring a lot of mental effort, this influences greatly on the purchase order (Cha et al., 2011)

Hypothesis 1: There is a positive relationship between ease of use of a website and purchase intention of consumer.

There has been evidences that more information on the website (information about the firm, its products and services, promotions etc.) positively affects customer satisfaction (Park et al., 2005). Whether a consumer is looking for specific product or browsing through the website as a form of leisure without intent to buy, there is a preference to information cues for evaluating the quality and performance of products advancing the customer to subsequent stages till the final stage of transaction.

Hypothesis 2: There is a positive relationship between site content and purchase intention of consumer.

When customers interact with an onsite store, they do so via a technical interface and not through any physical guidance. Website design represents the way in which the content is arranged in the website and is therefore acts as the interface that would play an important role in influencing customer satisfaction. Online stores use the implementation of website design factors such as visual, information and navigation design as marketing tools through which trust in the website can be created and subsequently purchase intention can be enhanced.

Hypothesis 3: There is a positive relationship between design of the site and purchase intention of consumer

In addition, as stated earlier, by analysing the site navigational patterns and the content of the pages viewed, the type of visit can be categorized which in turn is used to predict the likelihood of purchase (Carmona, Ramírez-Gallego, Torres, Bernal, del Jesus, & García, 2012; Moe, 2003; Keng Kau et al., 2003; Mobasher et al., 2002). Research has also been done on how to use clickstream data for predicting online purchase intentions (Fernandes et al., 2015). Transforming and aggregating click-stream data at different levels of abstraction (such as traffic, depth, time/date, time duration, depth of a session etc.) have been useful for predicting the purchase behaviour of the users [(B.Hasan, 2016)]. 1) Number of Pages Visited (NPV) in a site session is a measure of stickiness - to indicate depth of visit. 2) Average Session Length (ASL) measures this stickiness for a specified period, 3) Page view session prevalence – to analyse how users' behaviour alters throughout the session, 4) Time duration per conversion (TDPC) reveals the average amount of time elapsed per conversion, 5) The number of pages until conversion. The study concludes that using the page view sequence likelihood can reveal whether a user is about to engage.

Certain other metrics are also discussed in literature (Chi et al., 2002) : 1) most frequently occurring documents which tracks the navigational pattern of the user search- analyse if the users focus predominantly in key pages or do they have broad interests. 2) Law of surfing behaviour – session path length trends and session time trends 3) Session document co-occurrence – How often the most frequent documents appear in the same session for all sessions in the group which is indicative of task success rate and 4) Session document transitions – How often the most frequent documents are followed by another in a session, how often each document is the first or last in a session which characterizes the repeated sequences of forward and backward transitions like product lists to product detail pages.

In the studies described above the focus has been on the technical aspects like the type of page visited, the time spent on each page, the date of visit, the navigational pattern adopted etc. Website performance also has measurable effects on purchase orders. The quicker a webpage loads, the more likely a user will perform the targeted action on the webpage. From these papers we are able to identify certain technical specifications which can be used to model an online user's purchase intention. The following hypotheses have been formulated based on the theoretical backing of these papers:

Hypothesis 4: There is a positive relationship between speed of site and purchase intention of consumer

Hypothesis 5: There is a positive relationship between time spent on site page and purchase intention of consumer.

Remember mad rush during Black Friday? Studies have tried to capture this attribute called special day feature i.e., closeness of the site visiting time to a specific special day to see how it impacts the decision. Results of these studies have concluded that on these days the online activities are more likely to be finalized with transaction.

Hypothesis 6: There is a positive relationship between special day and purchase intention of consumer

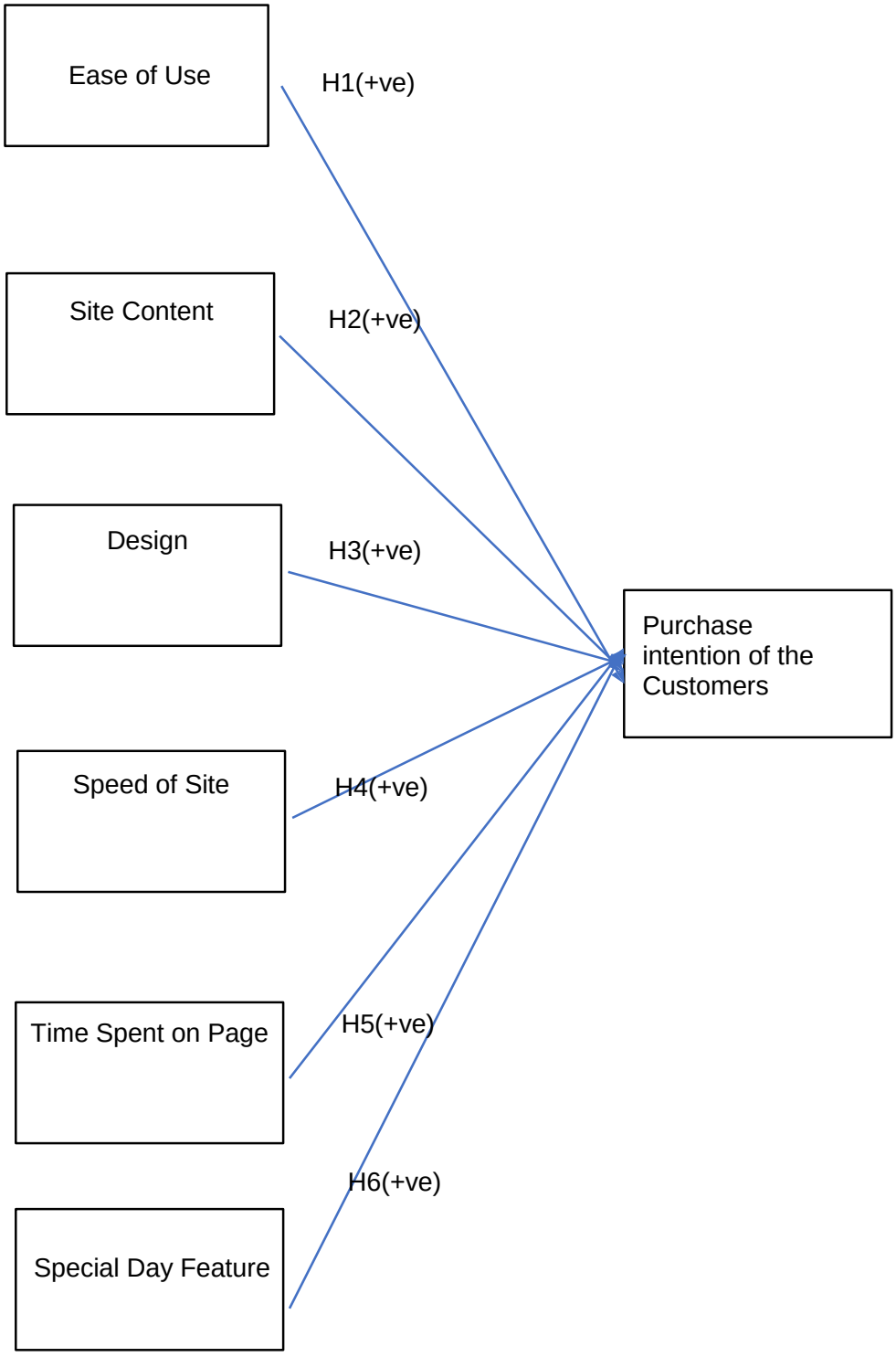
The factors such as website usability, content and design, performance of site, type of pages visited and time spent on each type of page have been quantified with the aid of Google Analytics and the following metrics have been taken into consideration (Sakar et al., 2018).

- Bounce rate – refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session
- Exit rate – for all page views to the page, the percentage that were the last in the session
- Page Value – the average value for a web page that a user visited before completing an e-commerce transaction.
- Number of product related pages visited, and the time spent on product related pages by a user
- Number of pages related to account management visited by a user and the time spent on the same
- Number of informational pages (about Website, address information etc.) visited by a user and the time spent on these pages

(The definitions used have been adopted from (Alpaslan, 2018) and (https://www.lovesdata.com/blog/google-analytics-glossary, n.d.))

The objective of the study is to identify the key factors that drive online consumer's intention to purchase, based on the metrics given and hence predict whether a particular session would translate into revenue for the site and thereby address the problem of shopping cart abandonment.

Figure 1. A conceptual framework of hypotheses: metrics and dependencies.



METHOD

Source

As a part of analysing behaviour patterns of online shoppers, we wished to collect secondary data from the internet. Secondary data are more organized than other methods of raw data collection and provide more precision. Since these datasets have already been used for analysis, they would be reliable and suitable for model building.
Other advantages of using secondary data

- Economical
- Saves efforts and expenses
- More specific
- Gaps can be filled with additional info.

While we did not restrict our search in terms of consumer segment, demography, accessing platform etc., we were typically looking for a dataset such that the it is reasonable, consistent and comparable with all situations of online shopping, where above mentioned parameters such as session time, sequence of visits etc. are quantified and constructed using Google Analytics functions. This is available only from secondary data sources. The dataset would be chosen. Going by this principle, a multivariate training dataset from UCI Machine Learning Repository matched with our requirements: representing consumer activities during online shopping in terms of parameters that were considered in our hypothesis.

Dataset description

The dataset used contains data collected from online user sessions for a Columbia Sportswear company in the year 2018. This dataset contains features belonging to 12330 sessions where each session is unique belonging to a different user. Of the 12330 instances in the dataset and a total of 42261 web hits, 84.5 percentage (10422) were negative class samples that did not end with shopping, and the rest were positive class samples where shopping experience has ended in transaction.

The data can be generalized as it was collected through a one-year period to avoid any tendency to a special day, specific campaign user profile, or period.

Source

1.C. Okan Sakar
Department of Computer Engineering, Faculty of
Engineering and Natural Sciences, Bahcesehir University,

2.34349 Besiktas, Istanbul, TurkeyYomi Kastro
Inveon Information Technologies Consultancy and Trade,
34335 Istanbul, Turkey
Dependent variable: Revenue (Categorical)
Independent variables: The dataset consists of 10 numerical and 7 categorical attributes.

Statistical Analysis of Predictors

Table 1. Numerical Variables (The definitions used in the below table have been adopted from (Alpaslan, 2018) and (https://www.lovesdata.com/blog/google-analytics-glossary, n.d.)

Feature Name	Feature Description	Examples/Comments	Min Value	Max Value
Administrative	Number of pages related to account management accessed by the user.	Login/Logout, Login Register, Referral, Password Recovery	0	2
Administrative Duration	Total amount of time (in seconds) spent by the user on account management related pages		0	3398
Informational	Number of pages related to website, communication and address information of the shopping site visited by the user.	Pop-up, contact us, Catalog, Store, FAQs	0	24
Informational Duration	Total amount of time (in seconds) spent by the user on informational pages		0	2549
Product Related	No. of pages visited by user about product related pages	Search, Product, Cart	0	705
Duration - Product Related	Total time (in seconds) spent by the user on product related pages		0	63973

Bounce Rate	Percentage of the only page user visited before exiting the session.	The dataset contains average bounce rate value of the pages visited by the user	0	0.2
Exit Rate	Percentage of the page that were last in the session before the user exited	The dataset contains average exit rate value of the pages visited by the user	0	0.2
Page Value	Average value of the pages visited by the user before he/she made a transaction or landing on goal page.	This is to get an idea of which page contributed more to the revenue.	0	361
Special Day	Closeness of the site visiting time to a special day	Does not include weekend	0	1

Table 2. Categorical Variables

Feature Name	**Feature Description**	**Number of Categories**
Operating Systems	Operating system of the visitor	8
Browser	Browser of the visitor	13
Region	Geographic region from which the session has been started by the visitor	9
Traffic Type	Traffic source by which the user has arrived at the website (e.g. banner, SMS, direct)	20

Visitor Type	Visitor type as “New Visitor”, “Returning Visitor” and “Other”	3
Weekend	Boolean value indicating whether the date of the visit is weekend	2
Month	Month value of the visit date	12
Revenue	Class label indicating whether the visit has been finalized with a transaction	2

The categorical variable ‘Revenue’ can take two values namely 0 and 1. A positive response translates to 1. This indicates that the customer’s visit to the site yielded in revenue generation. By building a predictive model with ‘Revenue’ as the dependent variable and the other variables mainly the google analytics metrics such as Bounce rate, Exit rate, Page Value etc. as the independent ones we can predict whether a prospective customer’s purchase intention materializes into something quantifiable as ‘Revenue’ in our case. Thus, we try to map a positive purchase intention with the customer actually buying something and hence revenue generation. The factors such as website usability, content and design, slowness of site, type of pages visited and time spent on each type of page and special day feature i.e, closeness of the site visiting time to a specific special day, which are studied in the hypotheses have been quantified with the aid of the above mentioned Google Analytics metrics.

The dataset consists of 10 numerical and 7 categorical variables that can be used as independent variables as mentioned in the previous section. From this available set 11 variables (10 numerical and 1 categorical) have been chosen as required by the study.

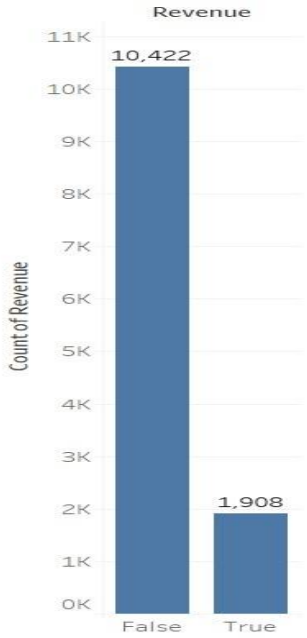
The chosen independent variables are Administrative, Administrative Duration, Informational, Informational Duration, Product Related, Product Related Duration, Bounce Rate, Exit Rate, Page Value, Special Day and Weekend.

The dependent variable as mentioned earlier is Revenue.

Exploratory analysis

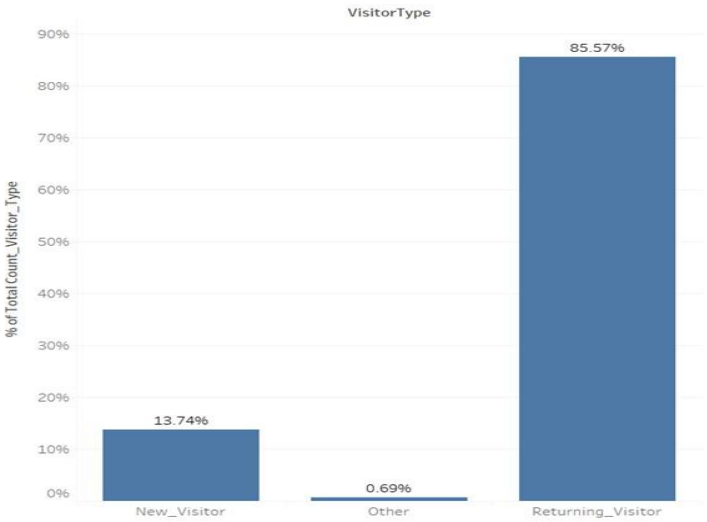
As seen around 84.53% (=10422/ (10422+1908)) of the sample is negative sample, i.e customers whose visits to site did not lead to any revenue generation. This implies that the conversion rate for the site under consideration is around 15.47% (1908/ (10422+1908)). The dataset is found to be an imbalanced dataset, with the True class (=1) found to be significantly lesser than the False class (=0), so further analysis would be done based on taking this into consideration.

Figure 2. Representation of Revenue Conversion



Most of the visitors to the site are returning visitors (85.57%), while around 13.74% are new visitors.

Figure 3. Percentage share of Visitor Type in the considered dataset



We also see that around 14% of the returning visitor belong to the True class i.e. the ones who generate revenue, while around 25% of new visitors contribute to the revenue. The company is thus able to attract newer customers more than the existing ones. This might be indicative of a limited product portfolio because of which returning customers are unable to find what they are looking for and hence leading to a lesser conversion rate.

	VisitorType		
Revenue	New_Visitor	Other	Returning_Visitor
False	75.09%	81.18%	86.07%
True	24.91%	18.82%	13.93%
Grand Total	100.00%	100.00%	100.00%

Figure 4. Visitor Type vs Revenue conversion

Interestingly most of the user sessions are in the weekdays-around 76.74% compared to weekends-23.23%. This might be due to the fact that people tend go out shopping during the weekends.

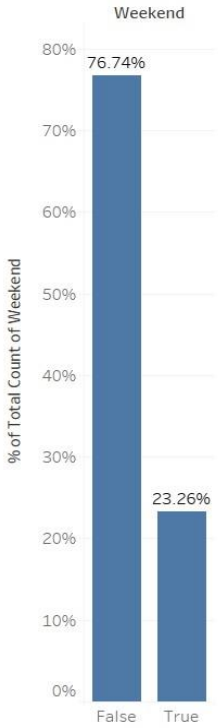


Figure 5. Representation of conversion during weekend/weekdays

However, the conversion rate during weekends (17.40%) is found to be slightly greater than weekdays (14.89%).

	Weekend	
Revenue	False	True
False	85.11%	82.60%
True	14.89%	17.40%

Figure 6. Weekend vs Revenue conversion

The pie chart shows the distribution of special days in each month. It helps to find whether there is any rise/fall in revenue on these special days.

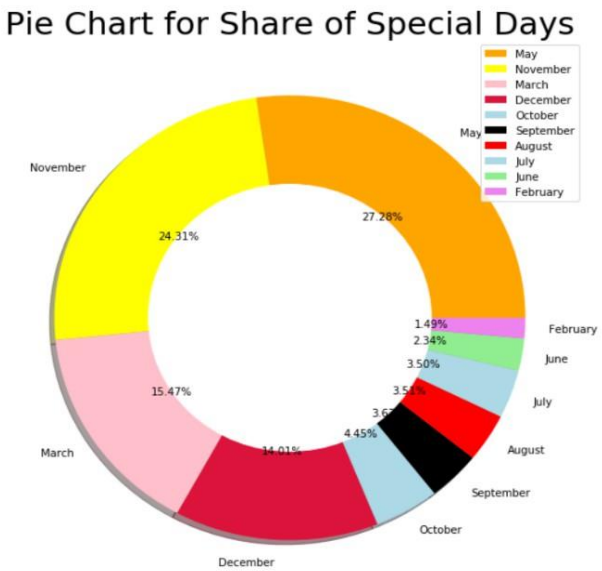


Figure 7. Distribution of special days across months.

We see that however there is no expected rise in revenue conversion during the special days. This might be due to the fact that the data is of an online sporting company and hence there doesn't exist a relation between special days and buying of sporting goods.

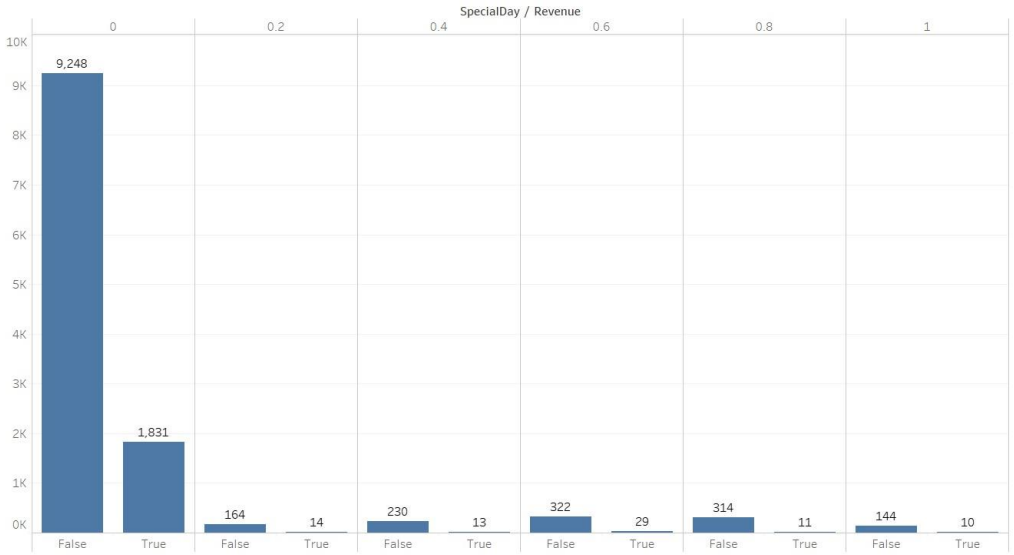


Figure 8. Special day vs Revenue conversion.

The number of visitors by month and the conversion rates are shown. We see more users in the month of March, May, November and December compared to other months.

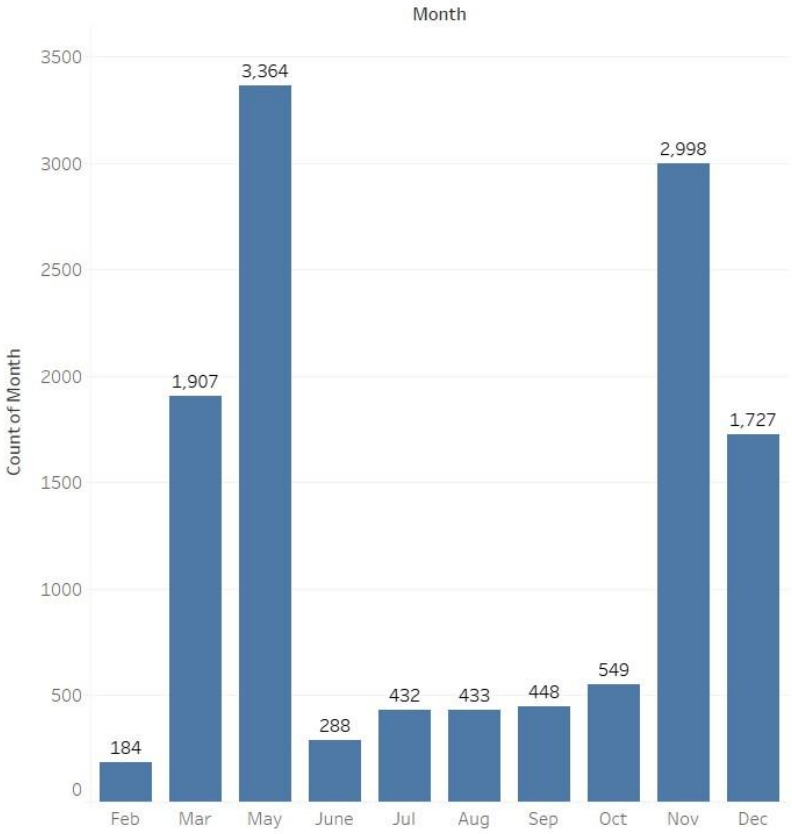


Figure 9. Visitors across months.

Also, the conversion rates are found to be higher in the months of November and October.

	Month									
Revenue	Feb	Mar	May	June	Jul	Aug	Sep	Oct	Nov	Dec
False	98.37%	89.93%	89.15%	89.93%	84.72%	82.45%	80.80%	79.05%	74.65%	87.49%
True	1.63%	10.07%	10.85%	10.07%	15.28%	17.55%	19.20%	20.95%	25.35%	12.51%

Figure 10. Month vs Revenue conversion.

RESULTS

The correlation between the independent variables is examined and the results are presented below.

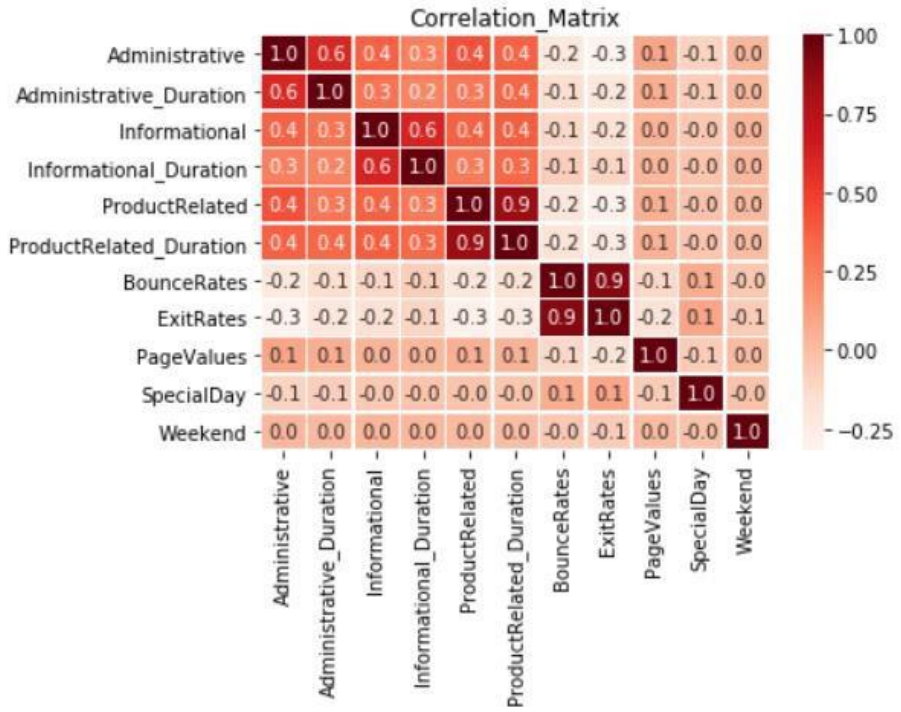


Figure 11. Correlation matrix on metrics

It is seen that there is a high positive correlation between the bounce rate and exit rate metrics. Also, there is a negative correlation between the Page value and the bounce rate and exit rate metrics, i.e. pages which contribute more to revenue generation has less exit and bounce rates. However, this negative correlation is found to be very less which indicates that in actuality users bounce off these high value pages thus leading to loss in revenue.

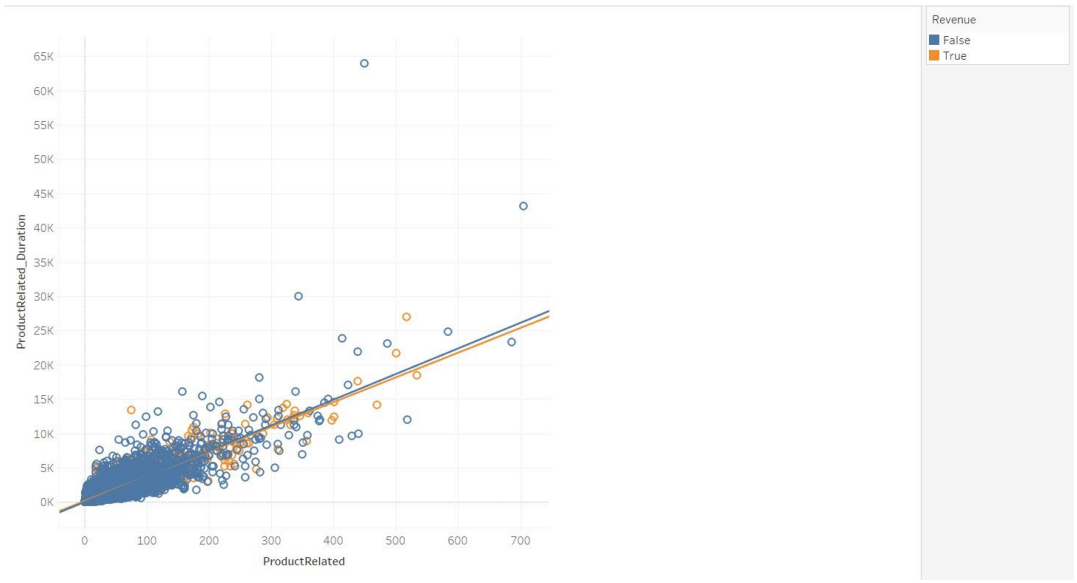


Figure 12. Relation between Product Related page and Product Related duration

We see that high correlation exists between the type of page say for eg. Product Related and the time spend on the particular page which is in this case Product Related_Duration. This is expected as the content of the page directly influences the time spent on it.

Data preparation:

Label encoding- The categorical variables are label encoded.
Oversampling – Since the dataset is imbalanced, we have used the SMOTE technique to address the issue.

Models:

To study the relationship between the dependent and independent variables we have employed three models namely:

1. Logistic Regression
2. XGBoost Algorithm
3. Neural Networks

All models, predict the dependent variable (Revenue), given the independent variables. While the Logistic Regression model focusses on the interpretability aspect of the coefficients, the focus of XGBoost and Neural networks are on the predictive ability.

3.1 Logistic Regression model:

Using the logistic regression model, we study the effects of the independent variables namely the google metrics such as PageValues, type of pages (ProductRelated, Administrative etc.) time spent on these pages (ProductRelated_Duration, Administrative_Duration etc), ExitRates, BounceRates, Special Day etc on the dependent variable namely Revenue.
The results of the Logistic model are shown below.

```
Optimization terminated successfully.
         Current function value: 0.394845
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:                Revenue   No. Observations:                16756
Model:                          Logit   Df Residuals:                    16739
Method:                           MLE   Df Model:                           16
Date:                Fri, 10 Jan 2020   Pseudo R-squ.:                  0.4304
Time:                        20:47:22   Log-Likelihood:                -6616.0
converged:                       True   LL-Null:                       -11614.
                                        LLR p-value:                     0.000
============================================================================================
                               coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------
Administrative              -0.0065      0.009     -0.735      0.462      -0.024       0.011
Administrative_Duration     -0.0003      0.000     -1.781      0.075      -0.001    2.83e-05
Informational               -0.0929      0.023     -3.983      0.000      -0.139      -0.047
Informational_Duration       0.0006      0.000      2.951      0.003       0.000       0.001
ProductRelated               0.0057      0.001      5.532      0.000       0.004       0.008
ProductRelated_Duration      0.0001   2.46e-05      4.196      0.000    5.51e-05       0.000
BounceRates                 -2.4470      1.766     -1.386      0.166      -5.908       1.014
ExitRates                   -9.6322      1.506     -6.396      0.000     -12.584      -6.681
PageValues                   0.1396      0.003     47.348      0.000       0.134       0.145
SpecialDay                  -1.8781      0.177    -10.595      0.000      -2.226      -1.531
Month                        0.1132      0.009     12.781      0.000       0.096       0.131
OperatingSystems            -0.3447      0.027    -12.724      0.000      -0.398      -0.292
Browser                     -0.0209      0.014     -1.440      0.150      -0.049       0.008
Region                      -0.0427      0.009     -4.515      0.000      -0.061      -0.024
TrafficType               8.703e-05      0.006      0.014      0.989      -0.012       0.012
VisitorType                 -0.3955      0.029    -13.626      0.000      -0.452      -0.339
Weekend                     -1.0258      0.063    -16.169      0.000      -1.150      -0.901
============================================================================================
```

Figure 13. LR results-Overall significance and R square

We see from the p-value, the overall logistic regression is significant. The pseudo R square for the model is .4304, i.e. the independent variables are able to explain 43.04% of the variability in the dependent variable. Further analysis of the variables indicates that the variables except 'Administrative', Administrative_Duration', 'BounceRates', Browser' and 'TrafficType' variables all other variables are significant.

Using the coefficients of the independent variables in the LR model, the odds ratio and probability have been calculated as follows:

Odds ratio = exp(coef(LR model))
Probability = Odds/(1+Odds)

Table 3. Odds and Probabilities for Logistic Regression

Variables	Coefficients	Odds Ratio	Probability
Administrative	-0.007	0.993521079	49.84%
Administrative_Duration	0.0003	0.999700045	49.99%
Informational	-0.093	0.911284624	47.68%
Informational_Duration	0.001	1.00060018	50.01%
ProductRelated	0.006	1.005716276	50.14%
ProductRelated_Duration	0.0001	1.000100005	50.00%
BounceRates	-2.447	0.086552856	7.97%
ExitRates	-9.632	6.55826E-05	0.01%
PageValues	0.140	1.149813781	53.48%
SpecialDay	-1.878	0.152880303	13.26%
Month	0.113	1.119855882	52.83%
OperatingSystems	-0.345	0.708432851	41.47%
Browser	-0.021	0.979316891	49.48%
Region	-0.043	0.958198807	48.93%
TrafficType	0.00008	1.000087034	50.00%
VisitorType	-0.396	0.673343283	40.24%
Weekend	-1.026	0.358509543	26.39%

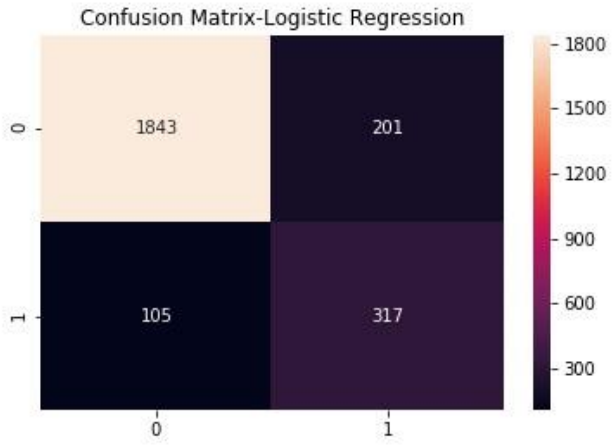
We see from the above a table that the major variables that influence the purchase intention are the PageValues and the type of pages and time spend on these.

The Logistic Regression model focussed on the interpretability of the coefficients. The focus of the next two models namely XGBoost and neural networks is on predictive ability, i.e. to predict whether a customer visit to a site will lead to revenue generation.

XGBoost Algorithm

The XGBoost algorithm outperforms Logistic regression in terms of predictive ability. The results are shown below.

Figure 14. Confusion matrix XGBoost algorithm



We see that the accuracy of the model is 87.59%. The metric that we are more interested in is the recall, which has improved considerably to 75%. The recall metric tells us that out of 100 cases where a customer bought something, 75 of these are identified by the algorithm. Since our objective is to predict whether a customer buys, this recall metric tends to be what our focus should be on.

	precision	recall	f1-score	support
0	0.95	0.90	0.92	2044
1	0.61	0.75	0.67	422
accuracy			0.88	2466
macro avg	0.78	0.83	0.80	2466
weighted avg	0.89	0.88	0.88	2466

Figure 15. Evaluation parameters for AG boost algorithm

The area under the AUC curve has also gone up to 83%, which indicates that the model does a better job in classification.

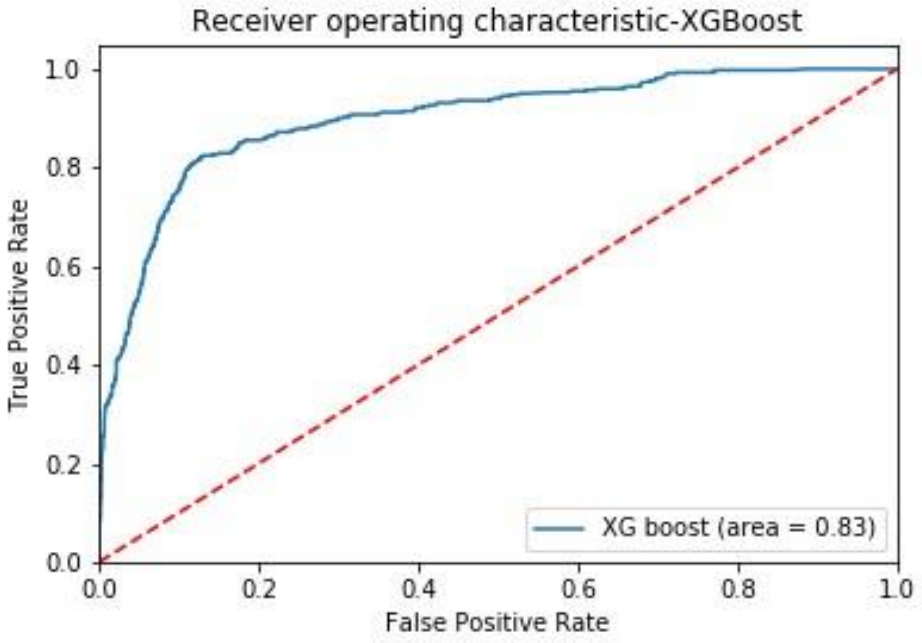


Figure 16. AUC graph for AG

Neural Networks

The neural networks outperform XG boost in terms of predictive ability as shown below. We see that the accuracy has improved to 89.10% along with recall which has gone up to 79%.

	precision	recall	f1-score	support
0	0.95	0.88	0.92	2044
1	0.58	0.79	0.67	422
accuracy			0.87	2466
macro avg	0.77	0.84	0.79	2466
weighted avg	0.89	0.87	0.88	2466

Figure 17. Evaluation parameters for ANN

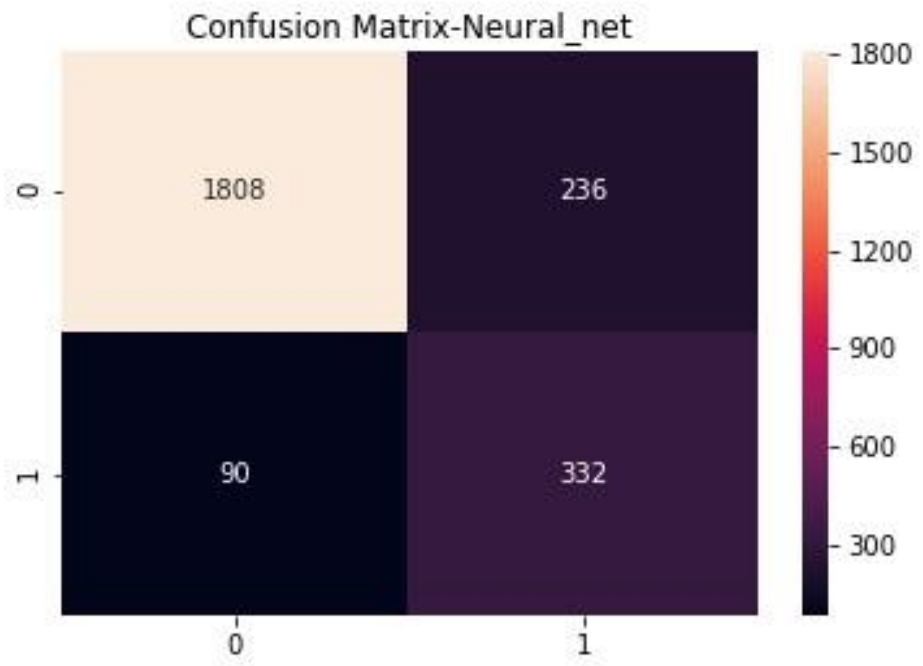


Figure 18. Confusion matrix for ANN.

DISCUSSION

The first part of the research tries to map the purchase intentions of customers with the characteristics of the website such as ease of use, site content, design of the site, speed of the site, special day feature and time spent on the site page. The google analytics metrics such as Page value, number of pages visited during a session, time spent on these pages, exit and bounce rates etc. have been used to quantify these characteristics. We see that as the page value increases, i.e. the average value for a page that a user visited before landing on the goal page or completing an ecommerce transaction increases, chances for revenue generation go up. Similarly, the more time spent on product related pages increases chances of revenue generation. So, the focus should be on improving these metrics. The Page value metric is directly related to the site content and design of the site. Better the content and design, the more chance for the page to contribute towards revenue creation. Hence, we see that site content and design of the site to have a positive relation with consumers purchase intention.

Another metric under consideration is the type of page and time spent on it. We see that the content, ease of use and speed directly influences the number of pages visited and the time spend on each. Better content coupled with ease of use and speed leads to the user visiting more pages as well as increasing the time spend on the page. These factors directly lead to better revenue generation prospects, which is quantified in our study by the purchase intention of the customers.

Since, the data chosen was that of an ecommerce site dealing in sports goods, the special day feature which is representative of the closeness to a special day, is not found to be a significant factor. However, for other products this might turn out to be important.

The focus on the second part of the research is to show two models which can predict whether a visitor's session would lead to revenue generation. In this regard the XGBoost algorithm and Neural networks were explored. Based on the values of google analytics metrics, these models are able to predict the purchase intention of customers with considerable effectiveness.

Businesses can analyse these google metrics real time to drive up revenue conversion. Say for example by studying the time spend on different pages and the number of pages visited, businesses can see which pages are more appealing to the customer and thereby provide customized content. Also, being able to predict a customer's intention helps the company in deciding how much to spend on a particular customer and helps in deciding the marketing campaign and funds to be utilized on a detailed level. More offers can be directed towards customers who have a higher probability of purchasing. From the company's perspective, what content appeals to the customers will help in pushing tailor made marketing campaigns. This level of real time customization combined with appropriate marketing will boost revenues for the companies.

Limitations and future research directions

This paper presents an analysis on the strength of the factors that increases conversion rate. However, this research is based on a specific sports apparel. Efforts can be made to analyse these factors across other e-commerce platforms with diversifying range of products to get the overall picture that helps us understand how it affects trust in B2C ecommerce.

The study takes a technology-oriented view to predicting customers purchase intention, i.e. it explores only the technological aspects of the website. The customer side of the equation i.e. the background of the customer, his previous transaction history etc, has not been explored. This analysis can also take Hofstede's approach and be extended to multiple nations to measure significant difference in perception of the considered attributes across cultures.

REFERENCES

Carmona, C. J., Ramírez-Gallego, S., Torres, F., Bernal, E., del Jesus, M. J., & García, S. (2012). Web usage mining to improve the design of an e-commerce website: OrOliveSur. com. *Expert Systems with Applications*, *39*(12), 11243-11249.

Cha, J. (2011). Exploring the internet as a unique shopping channel to sell both real and virtual items: A comparison of factors affecting purchase intention and consumer characteristics. Journal of Electronic Commerce Research, 12(2), 115.

Fernandes, R. F., & Teixeira, C. M. (2015). Using clickstream data to analyze online purchase intentions.Masters thesis, University of Porto.

Gao, Y., & Wu, X. (2010). A cognitive model of trust in e-commerce: Evidence from a field study in China. *Journal of Applied Business Research (JABR)*, *26*(1).

Kau, A. K., Tang, Y. E., & Ghose, S. (2003). Typology of online shoppers. *Journal of consumer marketing*, *20*(2), 139-156.

Kukar-Kinney, M., & Close, A. G. (2010). The determinants of consumers' online shopping cart abandonment. *Journal of the Academy of Marketing Science*, *38*(2), 240-250.

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, *13*(1-2), 29-39.

Park, J., & Stoel, L. (2005). Effect of brand familiarity, experience and information on online apparel purchase. *International Journal of Retail & Distribution Management*. Vol. 33, No. 2, pp.148–160.

Wells, J. D., Valacich, J. S., & Hess, T. J. (2011). What signal are you sending? How website quality influences perceptions of product quality and purchase intentions. *MIS quarterly*, 373-396.

DECISION SCIENCES INSTITUTE

Do Incentives for Government School Students Improve their Learning Outcomes? An Integrated Analysis using Machine Learning Algorithm and Econometric Modelling

Nitin Jain
IIM Udaipur, India
nitin.2020phd@iimu.ac.in

Mohit Jain
VIT Bhopal, India
jain12mohitk@gmail.com

ABSTRACT

Even more than seventy years after gaining independence and with a population reaching the 1.38 billion-mark, unsatisfactory learning outcomes in school education remain a glaring issue for India. The extant research shows an opportunity to improve the methodology, which this paper addresses using the random forest machine-learning algorithm in the first stage of the instrumental variable econometric technique to enhance the accuracy of the estimates. The findings reveal that government incentives do not have a positive spillover on students' learning outcomes in government schools, so government policy measures should focus on the learning outcomes and not just school enrollment.

KEYWORDS: Educational Inequality, Learning Outcomes, Endogeneity, Education Policy, Machine Learning

INTRODUCTION

"*The destiny of India is now being shaped in her classrooms.*"
- India Education Commission, 1964–66

The above quotation aptly highlights that education is a crucial driver in the socio-economic growth of a nation. A solid education contributes to individuals' well-being *instrumentally* by elevating income and living standards and *intrinsically* by raising individual capabilities and the welfare of future generations through intergenerational transmissions (Dreze & Sen, 2013). Economists argue that the "great divergence" between the developed and under-developed nations is reinforced – if not caused – by rapid developments in schooling witnessed in the developed countries (Chaudhary, Musacchio, Nafziger, & Yan, 2012). Extant research has shown a positive association between education and earnings, health, and other economic and well-being outcomes. Measures of psychosocial competencies such as self-efficacy, sense of inclusion, self-esteem, and educational aspirations are correlated with caregiver's education and school participation. These reflect life skills, which play an essential role in future socio-economic status (Dercon & Krishnan, 2009).

India has taken initiatives to improve education at the grass-root level. As per the "National Policy for Children," which the Government of India adopted in 1974, children are considered an important asset and need to be provided adequate services throughout the period of their growth (Apte, 1979). After that, multiple initiatives were taken to uplift the standard of education

in the country. These include the Non-Formal Education Program (1979-1980); Operation Blackboard for Small Rural Schools (1986) (Tilak, 1995); Total Literacy Campaigns (1988); and the District Primary School Education Program (1994-2002) (Asadullah & Yalonetzky, 2012). The post-National Policy for Education (NPE) era witnessed a spate of educational initiatives, both national and foreign-funded (De, Noronha, & Samson, 2003). In 2001-02, India initiated one of the biggest global educational initiatives – the Sarva Shiksha Abhiyaan (SSA) (Banerji & Mukherjee, 2008). With a national-level scope and scale, it had massive financing. Very recently the National Education Policy (NEP)-2020 has been announced with a gamut of reform measures and aims to transform India's education landscape by 2030.

It is unfortunate that despite governmental thrust on equal access to quality education, India is home to the highest number of illiterate people, and every third illiterate in the world is an Indian (Jha, 2007). According to Kurien (1981), the most conspicuous failure of the Indian education system had been the inability after three decades post-independence to implement Article 45 of the constitution that mandated free and compulsory education for all children until the age of fourteen years. The contribution of primary schooling to economic development is much higher than has been perceived by Indian society (Colclough, 1982). Still, the progress on increasing access to elementary education in India has been slow, although steady (Little, 2010). There seems to be a persistence of the stereotypical view that poverty is the main reason for low school enrollment (Ota & Moffatt, 2006). In Indian society, there could also be an intersection of factors, such as gender, caste, religion, poverty, etc. that might hinder school enrollment more than just poverty by itself. The issues of limited access to, as well as poor quality of education and teachers, and high drop-out, and repetition rates, have continued to be other vital concerns (Mishra, 1999). However, the issues that have received greater public attention in the past were more related to educational power, such as regulating access to educational system and decisions about which social strata, actors, or institutions would benefit from the proposed changes (Rosenthal, 1974).

Another important debate has been the significant rise of private schools in India over the past couple of decades. This rise has been attributed to the demand for education and lack of adequate facilities (Huisman, Rani, & Smits, 2010) in government schools and their poor performance (Gouda, Chandra Das, Goli, & Maikho Apollo Pou, 2013). Private schools provide an English medium of instruction (Wadhwa, 2009); they show more accountability (G. Kingdon, 1996); they have a better infrastructure (Reddy & Galab, 2018); they possess a greater market orientation and foster literacy (Pal & Kingdon, 2010). Households that send their children to private schools are systematically different from those that do not (Borooah, 2012; Chudgar & Creed, 2014; Chudgar & Quin, 2012). The mushrooming of private schools has raised concerns around the deepening of inequity in society by them. The rapid migration of children from the government to private schools has rendered many government schools economically unviable (Kingdon, 2017). Hence, it is critical to analyze if there is indeed a difference in the students' learning outcomes in these two types of schools and whether government initiatives such as free books and uniforms, fee-waivers, etc. improve the learning outcomes.

We consider that an essential aspect of the empirical analysis on the impact of the type of schooling (private versus government) on learning outcomes is the decision to enroll in either a government or a private school. This decision in Indian society is driven by various social, economic, cultural, and demographic factors, besides schools' infrastructure and soft such as discipline, uniforms, etc. Therefore, a key challenge in comparing the private and government schools lies in addressing this lack of randomness in school enrollment decisions. The two

dimensions, i.e., private school students belonging to higher socio-economic strata and their parents placing a greater value on education, pose a challenge to conclusions on the comparison between learning outcomes of students in private and government schools (Desai, Dubey, Banerji, & Vanneman, 2009).

While government interventions and initiatives may increase school enrollment, they may not improve learning outcomes (Banerjee, Cole, Duflo, & Linden, 2007). As opposed to mere enrollment years in school, cognitive skills are powerfully related to income earnings and hence, economic growth. Therefore, policymakers must understand how learning outcomes are impacted (Azam, Kingdon, & Wu, 2016). A crucial gap in the extant research about the impact of government incentives on school education enrollment and learning outcomes in India is the lack of methodological rigor. Extant studies have not addressed the issue of endogeneity very well, which raises a question on the accuracy of their estimates and conclusions. Secondly, to the best of our knowledge, there is no extant research, both in India and abroad, which has applied machine learning techniques in conjunction with econometric models to improve the accuracy of the estimates in the context of education-related empirical research.

All these reasons make it compelling to examine India's school education structure, where the landscape of education has its unique features. Here household and demographic characteristics such as caste, gender, religion, etc. can play a pivotal role in a child's educational trajectory. This analysis extends the work of Jain (2020) by estimating the endogenous variable from the instrument using the random forest algorithm and considering a broader scope encompassing India as a whole and not just its rural areas. The instruments considered to address endogeneity are government incentives, i.e., free books, free uniforms, and fees paid by the Government. The remaining paper is organized as follows. The next section outlines the literature, which is succeeded by a section on the need to integrate the random forest algorithm with instrumental variables (IV) analysis. Thereafter, data, analysis, discussion, and conclusion follow in that order.

LITERATURE REVIEW

Educational Reforms and Government Incentives

According to the Education for All (EFA) program by India, the universalization of elementary education (UEE) was introduced as a composite program of access to education for children up to 14 years of age through formal or non-formal education programs (Acharya, 1994b). UEE includes the provision of facilities, universal enrollment, and universal retention (Acharya, 1994a; Bordoloi, 2011). However, facilities' provision may not imply universal enrollment, and the latter may not ensure universal retention. Unfortunately, earlier, more emphasis was given on enrollment than on the quality of education, which led to high dropout rates (Chauhan, 2009). Also, universal access to schools should not be confused with the universal provision of facilities. There is a difference in physical access to schools and equality in access to education.

With the enactment of RTE in 2009, India joined 130 other nations in providing free and compulsory education to children between the ages of 6 and 14 years (Kaushal, 2012; Rai, 2014). The importance of primary education is enshrined in the Millennium Development Goals (MDG), which called for universal primary education by 2015 (Jayaraman & Simroth, 2015). Because of this, the government started a wide range of policies targeted at increasing school enrollment. The Indian Government introduced the National Program of Nutritional Support to

Primary Education in 1995. It mandated free meals in all public (government) schools wherein students from grade 1 to grade 5 were served wheat porridge cooked from 100 grams of rice or raw wheat on the school premises to provide 413.80 kcal and 8.2 grams of proteins (Afridi, 2005, 2011).

In 2001, the Supreme Court mandated the provision of cooked meals in all targeted schools within six months. An implicit food subsidy to needy families, it endeavored to alter parents' decision-making regarding their children's schooling. It can also play an important role in overcoming traditional social prejudices prevalent caste- and religion-ridden Indian society (Khera, 2006). Cooked meals may encourage parents to enroll their children if they consider it to improve health and learning outcomes. Thus, these meals are akin to a subsidy to the parents, and they improve child nutrition, which in turn, fosters better learning (Jayaraman & Simroth, 2015). Besides, the lure of a cooked meal could motivate students to attend school more regularly and not just enroll without being regular in attendance. This point is important as mere enrollment does not translate into attendance due to several reasons. The provision of cooked meals at school also ensures that its benefit is received wholeheartedly by the students and not leaked to other household members as in food grains. To identify the causal impact of these meals, Jayaraman and Simroth (2015) consider the staggered implementation of these by the state governments and found an almost 6.6 percent increase in primary school enrollment due to the introduction of meals.

Afridi (2011) concludes that enrollment rates remained unaffected. However, girls' attendance rates did increase, especially in lower grades, due to the transition from the provision of food grains to cooked meals. According to Afridi (2010), extant research shows a positive impact of food transfers on children's school participation. Still, the impact on cognitive ability is ambiguous. For example, A. Singh (2008) finds non-trivial gains of school meals on verbal learning outcomes. However, while analyzing the "cycle program" in the state of Bihar, Muralidharan and Prakash (2017) find that exposure to this program increased the secondary school enrollment of girls by 30 percent and reduced the gender-gap in the age-appropriate category by 40 percent. The impact is more in villages where secondary schools were farther away. The coordinated provision of bicycles to girls seems to have generated externalities beyond the program's cash value, such as improved safety while cycling groups and changes in patriarchal social norms that had inhibited female secondary school participation (Muralidharan & Prakash, 2017).

Private versus Government (Public) Schools

Parents consider factors like their income, access to information on schooling, etc. (Chudgar, 2012) to decide their child's schooling. Caste identity and the number of siblings influence parents' decision to send their children to private schools (Bhattacharya, Dasgupta, Mandal, & Mukherjee, 2015). Males might be given more preference for expensive private schools (Chudgar & Creed, 2014). There is a negative perception of teaching and infrastructural support in government schools, although teachers in government schools are paid a higher salary (Gouda et al., 2013; Nanjunda & Ramesh, 2010). Private schools have better performance (Desai et al., 2009; Gouda et al., 2013; Goyal, 2007; Kingdon, 1996; Tooley & Dixon, 2006; Wadhwa, 2009) as they do not have an "all pass" policy. Thus, in these schools children are promoted to the next grade based on their performance in regular evaluation, and their parents seek remedial action outside of the school if they do not perform well (Banerji, 2000). Private schools impart teaching in English, have better infrastructure, and a low pupil-to-teacher ratio (De et al., 2003; Gouda et al., 2013; Muralidharan & Kremer, 2006).

The degree of the gap between private and government schools in rural areas may vary due to factors like the village within which the private school is situated (Chudgar, 2012). Even in the presence of greater private school supply, traditionally disadvantaged children may not be able to avail themselves of these schools, especially at the lower primary level (Chudgar & Creed, 2014). Even though private schools' growth is massive, it has not made patterns of their enrollment more equitable in rural areas (Chudgar & Creed, 2016; Woodhead, Frost, & James, 2013). Though, in urban areas, there is a declining caste gap in enrolment in private schools (Chudgar & Creed, 2016); the private-government school gaps in secondary school mathematics vary by urban/rural areas and states (Azam et al., 2016).

Parents from the upper socio-economic group send their children to private schools (Hill, Samson, & Dasgupta, 2011). The mediation of schooling by socio-economic factors raises important questions about equity (Hill et al., 2011; Mehrotra & Panchamukhi, 2006) and discrimination in access (Azam, 2017; Hill et al., 2011). To address this inequity, the RTE co-opts private school for the delivery of education. At least 25 percent of these schools' seats are now being given to children from disadvantaged backgrounds, for which the state governments would reimburse them (Geeta G. Kingdon, 2017). Hence, voucher-based reforms which provide access to private schools could, therefore, be a very important policy reform (Tooley, 2016).

Learning Outcomes and Psychosocial Competencies

A fundamental crisis in India's primary education relates to learning. In the absence of standardized tests until secondary level board exams at grades X and XII, this endemic problem of inadequate learning remains hidden (Banerji, 2000). Government schools are unable to provide even the most elementary skills (Abhijit V Banerjee, Banerji, Duflo, Glennerster, & Khemani, 2010). The syllabus assumes linear learning, i.e., children have mastered the prior grades' competency and skills before moving to the current grade (Banerji, 2000). However, for about a decade, until Jan 2019, government schools had an "all pass" policy for primary grades, and mere attendance in a school would suffice for promotion to the next grade. Further, there is little by way of serious or consistent remedial measures for those who lag academically. Therefore, children, who lag, stay behind as the rest of the grade moves ahead, and most teachers in these government schools resort to the less risky route of focusing on children who can cope with the syllabus (Banerji, 2000).

Multiple factors impact learning outcomes – demographics such as gender (Das & Singhal, 2017), rural (Galab, Reddy, & Reddy, 2014); socio-economic (Alcott & Rose, 2015, 2017; Dev, 2016); household, such as physical amenities (U. Desai, 1991); school-level, such as school-type, the medium of instruction (Murugesan, 2015; Nair, 2015), performance-based pay (Geeta Gandhi Kingdon & Teal, 2007), union membership (G. Kingdon & Teal, 2010), student-teacher interaction (Galab et al., 2014); village or district infrastructure, such as employment opportunities (Oster & Steinberg, 2013), household electrification (Dave, 2013); impact of weather (Zimmermann, 2020); education level of politicians, such as chief minister (Clots-Figueras, 2012), etc.

The extant literature has focused on one or more of these factors while analyzing the learning outcomes. Most of the researchers find a superior performance of private schools over their government counterparts on literacy (Goyal, 2007, 2009; Goyal & Pandey, 2009) and numeracy skills (French & Kingdon, 2010; Goyal, 2007, 2009; Goyal & Pandey, 2009; R. Singh & Sarkar,

2015). However, the degree of superiority may vary across these findings, and some researchers did not find any significant difference in their performance. Kumar (2017) makes an interesting observation that may cast doubt over private schools' superior performance. According to Kumar (2017), private schooling does not improve the ability to solve problems where mathematical operations have not been specified. However, it does improve the ability where mathematical operations have been specified.

Random Forest Machine Learning Algorithm

"The ongoing revolution in data science and machine learning has not gone unnoticed in economics and social science."

- Ahrens, Hansen, and Schaffer (2018)

In much of cross-sectional econometrics and empirical work in economics, the tradition has been that the researcher specifies one model, estimates the model on the full dataset and relies on statistical theory to estimate confidence intervals for estimated parameters. Thus, the focus is on the estimated effects, controlling for heterogeneity or other confounding factors, often using a large set of fixed effects (Einav & Levin, 2014) rather than the model's goodness of fit. The goal of accurate prediction of outcomes is abandoned to pursue an unbiased estimate of a causal parameter of interest (Athey, 2018). Over the past 5-6 years, there has been a steadily growing body of literature advocating the use of machine learning (ML) algorithms in conjunction with econometric techniques (see for example, (Athey, 2015, 2018; Athey & Imbens, 2015; Mullainathan & Spiess, 2017)).

Athey (2018) defines ML as a field that develops algorithms designed to be applied to datasets, with the key focus being prediction (regression), classification, and clustering or grouping tasks. ML is beneficial in empirical economics as it provides a data-driven way to find similar restaurant reviews, newspaper articles, etc. and create variables useful for economic analyses (Athey, 2018). There are non-parametric techniques that are more flexible in terms of model form and distributional assumptions. These include classification and regression trees (CARTs) and random forest algorithm (Buskirk, Kirchner, Eck, & Signorino, 2018).

Breiman (2001) defines random forests as a classifier consisting of a collection of tree-structured classifiers as $\{h(x, \varphi_k), k = 1,2 \dots\}$ where $\{\varphi_k\}$ are independently and identically distributed random vectors, and each tree casts a unit vote for the most popular class at input x. The application of random forest models in survey research has steadily increased over the past few years. Few examples include Caiola and Reiter (2010), Earp, Mitchell, McCarthy, and Kreuter (2014), and Buskirk and Kolenikov (2015). Advocating the need to use a fitting procedure other than ordinary least squares (OLS), Ahrens et al. (2018) argue that the OLS estimator has zero bias but not necessarily the best out-of-sample predictive accuracy. The prediction error (PE_0) for y_0 given x_0 can be decomposed as below.

$$PE_0 = E[(y_0 - \hat{y}_0)^2] = \sigma_\epsilon^2 + Bias\,(\hat{y}_0)^2 + Var\,(\hat{y}_0) \quad (1)$$

The bias and variance need to be low to minimize the prediction error, but the bias need not necessarily be zero. Another issue with OLS is overfitting badly and leading to false positives in high-dimensional data (Ahrens et al., 2018). By combining results across an ensemble of trees, random forests avoid the overfitting issue due to a single tree and generate predictions with lower variance than a single tree (Buskirk, 2018). To avoid over-fitting, most ML procedures for

"supervised learning" involve two features: regularized estimation and data-driven choice of regularization parameters (Abadie & Kasy, 2017). Significant improvements in classification accuracy have resulted from growing an ensemble of trees and having them vote for the most popular class. For growing these ensembles, random vectors are often generated that govern each tree's growth in an ensemble, such as bagging (Breiman, 2001).

Alike consulting multiple experts and making a consensual decision based on their advice, repeated statistical analyses on the same data can be combined to form a single result called an ensemble. This is very useful when the original analysis outcome is sensitive to small changes in the sample (Altman & Krzywinski, 2017). In many prediction problems, it is found that combinations of prediction methods ("ensembles") perform better than individual methods, for example, random forests, which combines predictions from many regression trees (Athey, Bayati, Imbens, & Qu, 2019). Breiman (2001) argues that substantial gains in classification and regression accuracy can be achieved using trees' ensemble, where each tree in the ensemble is grown following a random parameter. The steps in constructing random forest models are summarized below (Breiman, 2001; Buskirk, 2018).

Step 1: Generate n_{tree} bootstrap sub-samples, with replacement, from the raw dataset with each bootstrap subsample having the same size as the input dataset.

Step 2: Grow a classification/regression tree without pruning for each of the bootstrap samples in step 1. At each node of each of the trees, randomly select the subset (m_{try}) of the predictor variables, which would be used as the basis for branching.

Step 3: Once all trees are grown, form a prediction for new data as in a usual decision tree. For regression trees, an average of these n_{tree} predicted values are computed. In contrast, the final predicted values for classification trees are calculated as the class with the majority of votes across the n trees.

It is also important to note that the random forest does not need an external cross-validation procedure to estimate the model accuracy (Čeh, Kilibarda, Lisec, & Bajat, 2018). An advantage of decision tree-based methods such as random forests is their ability to natively handle categorical predictors without transforming them (Au, 2018). According to Breiman (2001), since an additive combination of variables is to be defined, each time a categorical variable is chosen to split on at a node, the algorithm selects a random subset of categories of that variable. It then defines a substitute variable which equals one if the variable's categorical value is in the subset and zero, otherwise. So, for a categorical variable with n categories, n-1 dummy variables are generated as numeric variables for node splitting. For numeric predictors, data having values greater than the splitting point go to the right daughter node. However, for categorical predictors, the splitting point is represented by an integer whose binary expansion gives the identities of the categories that go towards left or right (Liaw & Wiener, 2015). For example, if a predictor has three categories and the splitting point is 4, then the binary expansion of 4, i.e., 100 $(0 * 2^0 + 0 * 2^1 + 1 * 2^2)$ indicates that cases with category 3 in this predictor get assigned to the left and the ones in categories 1, 2 get sent to the right.

Table 1: Pros and Cons of Random Forests

Pros	Cons
They can handle categorical, continuous, and skewed predictors. They can be used for scenarios involving a large number of predictors and small sample sizes. They are useful when the outcomes are a complex function of predictors involving interactions.	Missing data should be treated before applying random forest models. Correlated predictors can bias the measures of variable importance. They can be computationally very intensive. They are hard to interpret – while they give excellent predictions, they yield little insight into the data generation mechanism.

Source: Adapted from Buskirk (2018) and Duncan (2014)

DATA

This paper analyzes the India Human Development Survey (IHDS) data for the year 2011-12 which covers 42,152 households, 204,569 individuals, 4,267 schools across 1,501 villages, and 34 states and union territories in India (S. Desai & Vanneman, 2015). IHDS compares well with other national surveys. It captures diverse demographic, socio-economic, and religious characteristics, and has been used in the extant literature.

There are different categories of schools in India, as described in Table 4. For the current analysis, the government schools comprise ESG, government, and government-aided schools, as summarized in Table 5. The learning outcome is computed based on writing assessment, wherein if the student can write a paragraph with up to 2 mistakes, the outcome is 1 and 0, otherwise. The students are from age-groups 8-11 and attend school up to grade 9^{th} in urban or rural areas. A more comprehensive set of control variables have been considered based on the extant literature. These include age, gender, grade, medium of instruction, years of completed education, ever repeated a grade, distance from school, test writing language, school hours per week, homework hours per week, tuition hours per week, number of children per adult in household, state, caste category, religion, household income, and per capita consumption, and household highest education. Control variables like the medium of instruction and household income capture, to an extent, the impact of school infrastructure. Age is considered as a discrete variable to facilitate a more detailed analysis. This consideration of age as a discrete variable is also one of the methodological improvements over the extant research, which mostly examines age as a continuous variable. We also wanted to include control variables such as time spent on household chores such as collecting water and fuel. However, due to the missing value issue, we could not use these controls.

The final dataset had 9,937 observations, and probability weights (pweights) were used for the IV-2SLS and the random forest analysis. Table 2 below gives the descriptive statistics for critical continuous variables, Table 3 shows the correlations for the continuous variables, and Tables 4 through 17 provide the frequency distribution for the key categorical variables.

Table 2: Descriptive Statistics

Variable	Description	Mean	Minimum	Maximum
RO5	Age (years)	9.503	8.000	11.000
ED6	Completed years of education	3.358	0.000	9.000
CS5	School distance (km)	1.824	1.000	99.000
CS10	School hours/week	32.721	0.000	96.000
CS11	Homework hours/week	7.714	0.000	80.000
CS12	Private Tuition hours/week	2.167	0.000	70.000
COPC (‘000)	House per capita expenditure	19.712	2.211	269.420
Nchild_adult	No. of children per adult in a household	0.843	0.111	8.000
Income_person (‘000)	Household monthly income per person	29.906	-45.366	809.566

Source: Authors’ computations, N = 9,937

Table 3: Correlations (N = 9,937)

	RO5	ED6	CS5	CS10	CS11	CS12	COPC	Nchild_ adult	Income_per_ person
RO5	1.000								
ED6	0.564	1.000							
CS5	0.020	0.057	1.000						
CS10	0.018	0.029	0.006	1.000					
CS11	0.071	0.155	0.074	0.126	1.000				
CS12	0.035	0.090	0.033	-0.037	0.220	1.000			
COPC	0.030	0.100	0.153	0.027	0.173	0.130	1.000		
Nchild_adult	0.007	-0.130	-0.090	0.008	-0.138	-0.124	-0.219	1.000	
Income_per_ person	0.008	0.072	0.133	0.004	0.127	0.089	0.472	-0.195	1.000

Source: Authors’ computations

Table 4: Frequency Distribution for School type (CS4) versus Location (Urban or Rural)

School	Urban	Rural	Total
EGS	9 (0.30%)	30 (0.43%)	39 (0.39%)
Government	1093 (36.46%)	4797 (69.13%)	5890 (59.27%)
Government Aided	156 (5.20%)	125 (1.80%)	281 (2.83%)
Private	1643 (54.80%)	1866 (26.89%)	3509 (35.31%)
Convent	68 (2.27%)	45 (0.65%)	113 (1.14%)
Madrassa	24 (0.80%)	76 (1.10%)	100 (1.01%)
Other/ Open school	5 (0.17%)	0 (0.00%)	5 (0.05%)
Total N	2,998	6,939	9,937

Source: Authors’ computations, the column percentages are shown in parenthesis

Table 5: Frequency Distribution for School type (CS4_govt) versus Location (Urban or Rural)

	Urban	Rural	Total
Government Schools	1,258 (41.96%)	4,952 (71.36%)	6,939 (62.49%)
Other Schools	1,740 (58.04%)	1,987 (28.64%)	2,998 (37.51%)
Total N	2,998	6,939	9,937

Source: Authors' computations, the column percentages are shown in parenthesis
Note: EGS, Government, and Government Aided are categorized as government schools and all others as other schools.

Table 6: Frequency Distribution for Writing a Paragraph (TA10B_flag) versus School type (CS4_govt)

TA10B_flag	Government Schools	Other Schools	Total
Yes	4,319 (69.55%)	3,154 (84.63%)	7,473 (75.20%)
No	1,891 (30.45%)	573 (15.37%)	2,464 (24.80%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 7: Frequency Distribution for Gender (RO3) versus School type (CS4_govt)

Gender	Government Schools	Other Schools	Total
Male	2,982 (48.02%)	2,194 (58.87%)	5,176 (52.09%)
Female	3,228 (51.98%)	1,533 (41.13%)	4,761 (47.91%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 8: Frequency Distribution for Free Books (CS21) versus School type (CS4_govt)

	Government Schools	Other Schools	Total
No	485 (7.81%)	3,425 (91.90%)	3,910 (39.35%)
Yes	5,725 (92.19%)	302 (8.10%)	6,027 (60.65%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 9: Frequency Distribution for Fee paid by Government (CS22) versus School type (CS4_govt)

Fee paid by Government	Government Schools	Other Schools	Total
No	2,287 (36.83%)	3,469 (93.08%)	5,756 (57.92%)
Yes	3,923 (63.17%)	258 (6.92%)	4,181 (42.08%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 10: Frequency Distribution for Free Uniform (CS23) versus School type (CS4_govt)

Free Uniform	Government Schools	Other Schools	Total
No	2,642 (42.54%)	3,638 (97.61%)	6,280 (63.20%)
Yes	3,568 (57.46%)	89 (2.39%)	3,657 (36.80%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 11: Frequency Distribution for Religion (ED11) versus School type (CS4_govt)

Religion	Government Schools	Other Schools	Total
Hindu	5115 (82.37%)	2868 (76.95%)	7,983 (80.34%)
Muslim	807 (13.00%)	670 (17.98%)	1,477 (14.86%)
Christian	86 (1.38%)	66 (1.77%)	152 (1.53%)
Sikh	127 (2.05%)	97 (2.60%)	224 (2.25%)
Buddhist	36 (0.58%)	10 (0.27%)	46 (0.46%)
Jain	3 (0.05%)	8 (0.21%)	11 (0.11%)
Tribal	31 (0.50%)	6 (0.16%)	37 (0.37%)
Others	4 (0.06%)	1 (0.03%)	5 (0.05%)
None	1 (0.02%)	1 (0.03%)	2 (0.02%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 12: Frequency Distribution for completed years of education (ED6) versus School type (CS4_govt)

	Government Schools	Other Schools	Total
None – 0	152 (2.45%)	314 (8.43%)	466 (4.69%)
1st Class	424 (6.83%)	408 (10.95%)	832 (8.37%)
2nd Class	1,043 (16.80%)	658 (17.65%)	1,701 (17.12%)
3rd Class	1,415 (22.79%)	790 (21.20%)	2,205 (22.19%)
4th Class	1,448 (23.32%)	788 (21.14%)	2,236 (22.50%)
5th Class	1,103 (17.76%)	524 (14.06%)	1,627 (16.37%)
6th Class	484 (7.79%)	188 (5.04%)	672 (6.76%)
7th Class	124 (2.00%)	49 (1.31%)	173 (1.74%)
8th Class	16 (0.26%)	8 (0.21%)	24 (0.24%)
9th Class	1 (0.02%)	0 (0.00%)	1 (0.01%)
Total N	6,210	3,727	9,937

Table 13: Frequency Distribution for Caste (ED13) versus School type (CS4_govt)

Caste	Government Schools	Other Schools	Total
Brahmin	189 (3.04%)	267 (7.16%)	456 (4.59%)
Forward/General (Except Brahmin)	1,054 (16.97%)	996 (26.72%)	2,050 (20.63%)
Other Backward Classes (OBC)	2,416 (38.90%)	1,748 (46.90%)	4,164 (41.90%)
Scheduled Castes (SC)	1,772 (28.53%)	549 (14.73%)	2,321 (23.36%)
Scheduled Tribes (ST)	724 (11.66%)	135 (3.62%)	859 (8.64%)
Others	55 (0.89%)	32 (0.86%)	87 (0.88%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 14: Frequency Distribution for Completed Years of Education (ED6) versus School type (CS4_govt)

Education Level	Government Schools	Other Schools	Total
None – 0	152 (2.45%)	314 (8.43%)	466 (4.69%)
1st Class	424 (6.83%)	408 (10.95%)	832 (8.37%)
2nd Class	1,043 (16.80%)	658 (17.65%)	1,701 (17.12%)
3rd Class	1,415 (22.79%)	790 (21.20%)	2,205 (22.19%)
4th Class	1,448 (23.32%)	788 (21.14%)	2,236 (22.50%)
5th Class	1,103 (17.76%)	524 (14.06%)	1,627 (16.37%)
6th Class	484 (7.79%)	188 (5.04%)	672 (6.76%)
7th Class	124 (2.00%)	49 (1.31%)	173 (1.74%)
8th Class	16 (0.26%)	8 (0.21%)	24 (0.24%)
9th Class	1 (0.02%)	0 (0.00%)	1 (0.01%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 15: Frequency Distribution for Highest Education in Household (HHEDUC) versus School type (CS4_govt)

Education Level	Government Schools	Other Schools	Total
None – 0	1,519 (24.46%)	402 (10.79%)	1,921 (19.33%)
1st Class	26 (0.42%)	9 (0.24%)	35 (0.35%)
2nd Class	117 (1.88%)	26 (0.70%)	143 (1.44%)
3rd Class	146 (2.35%)	50 (1.34%)	196 (1.97%)
4th Class	210 (3.38%)	37 (0.99%)	247 (2.49%)
5th Class	694 (11.18%)	236 (6.33%)	930 (9.36%)
6th Class	206 (3.32%)	88 (2.36%)	294 (2.96%)
7th Class	400 (6.44%)	125 (3.35%)	525 (5.28%)
8th Class	593 (9.55%)	365 (9.79%)	958 (9.64%)
9th Class	693 (11.16%)	323 (8.67%)	1,016 (10.22%)
10th Class	606 (9.76%)	544 (14.60%)	1,150 (11.57%)
11th Class	159 (2.56%)	92 (2.47%)	251 (2.53%)
12th Class	424 (6.83%)	505 (13.55%)	929 (9.35%)
First-year – post 12th	28 (0.45%)	44 (1.18%)	72 (0.72%)
Second-year – post 12th	44 (0.71%)	57 (1.53%)	101 (1.02%)
Bachelor's	258 (4.15%)	477 (12.80%)	735 (7.40%)
Above Bachelor's	87 (1.40%)	347 (9.31%)	434 (4.37%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 16: Frequency Distribution for Medium of Instruction (CS8) versus School type (CS4_govt)

Language	Government Schools	Other Schools	Total
Hindi	3394 (54.65%)	1,662 (44.59%)	5,056 (50.88%)
Assamese	80 (1.29%)	23 (0.62%)	103 (1.04%)
Bangla	399 (6.43%)	39 (1.05%)	438 (4.41%)
Gujarati	291 (4.69%)	90 (2.41%)	381 (3.83%)
Marathi	484 (7.79%)	87 (2.33%)	571 (5.75%)
Oriya	446 (7.18%)	39 (1.05%)	485 (4.88%)
Kannada	279 (4.49%)	42 (1.13%)	321 (3.23%)
Malayalam	93 (1.50%)	14 (0.38%)	107 (1.08%)
Tamil	125 (2.01%)	20 (0.54%)	145 (1.46%)
Telegu	119 (1.92%)	32 (0.86%)	151 (1.52%)
English	217 (3.49%)	1,587 (42.58%)	1,804 (18.15%)
Punjabi	204 (3.29%)	41 (1.10%)	245 (2.47%)
Urdu	62 (1.00%)	46 (1.23%)	108 (1.09%)
Others	17 (0.27%)	5 (0.13%)	22 (0.22%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

Table 17: Frequency Distribution for Test Language (TA10A) versus School type (CS4_govt)

Language	Government Schools	Other Schools	Total
Hindi	3,346 (53.88%)	2,191 (58.79%)	5,537 (55.72%)
Assamese	81 (1.30%)	30 (0.80%)	111 (1.12%)
Bangla	355 (5.72%)	42 (1.13%)	397 (4.00%)
Gujarati	297 (4.78%)	88 (2.36%)	385 (3.87%)
Marathi	485 (7.81%)	109 (2.92%)	594 (5.98%)
Oriya	424 (6.83%)	32 (0.86%)	456 (4.59%)
Kannada	270 (4.35%)	56 (1.50%)	326 (3.28%)
Malayalam	87 (1.40%)	29 (0.78%)	116 (1.17%)
Tamil	117 (1.88%)	40 (1.07%)	157 (1.58%)
Telegu	116 (1.87%)	55 (1.48%)	171 (1.72%)
English	511 (8.23%)	998 (26.78%)	1,509 (15.19%)
Punjabi	89 (1.43%)	33 (0.89%)	122 (1.23%)
Urdu	32 (0.52%)	24 (0.64%)	56 (0.56%)
Total N	6,210	3,727	9,937

Source: Authors' computations, the column percentages are shown in parenthesis

METHOD

Methodological Contribution: Integrating Machine Learning with Econometric Modelling

An analysis – be it quantitative or qualitative – can only be as good as the data being analyzed and its method. For quantitative studies, concerns have been raised on the quality of data, their accuracy (G. Kingdon, 2008; Kurien, 1981), comprehensiveness, and availability in the desired form at the right time (Aggarwal, 2000). Lack of adequate data is a key limitation cited by most papers (see, for example, Agrawal and Agrawal (2018); S. Desai et al. (2009)). According to Aggarwal (2000), there is no data on a large number of unrecognized schools. There is a lack of information on educational financing, institutional infrastructure, mobility, and unit costs'

estimation. On a similar note, G. G. Kingdon (1996) argues that enrollments in government-funded elementary schools are greatly exaggerated in official statistics. The level of data, i.e., householder state, can also be influential in choosing the appropriate methodology and hence, the outcome.

In terms of empirical techniques, both parametric and non-parametric methods are applicable depending upon the situation. The gold standard for drawing inferences about a policy's effect is a randomized controlled experiment (RCT) (Athey & Imbens, 2017). However, in many cases, RCTs are difficult to be conducted, which is where a large portion of the empirical analysis on policy-related questions in economics relies on observational data – where policies are determined in a way other than RCTs (Athey & Imbens, 2017). Causal inference is drawn from observational data via empirical strategies. It is said to be identified if it can be learned when the data is sufficiently large (Athey & Imbens, 2017).

The most widely used technique, ordinary least squares (OLS), although very simple to apply, has challenges. A fundamental problem with OLS models is that they do not treat for endogeneity, which is a key issue with non-random selection of samples. For example, a critical econometric pointer in the analysis of private schools is that the supply/provision of these schools is not random (Chudgar & Creed, 2014). Therefore, scholastic outcome analysis of private school students or schools with better infrastructure involves endogeneity. Children who perform well may belong to relatively more educated and affluent families than those from government schools or schools with poor infrastructure.

Therefore, the need is to move from the traditional OLS approach to technically more appropriate methods such as RCTs and Instrumental variables (IVs). The use of inappropriate methods or data could yield misguiding results. The next step in this direction would be to explore the feasibility of leveraging machine learning techniques in conjunction with econometric methods. In predictive models, the key concern is between expressiveness and overfitting, and this tradeoff can be looked at the goodness of fit in an independent test. On the other hand, causal models are concerned whether the parameter estimates from a particular sample are spurious, and if the assumptions needed to identify a causal effect are satisfied (Athey, 2018). IV refers to an estimation technique used to address multiple violations, which are collectively tagged under endogeneity, of assumptions that are needed for OLS to generate consistent estimates. Thus, an IV approach aims to allow causal inference in a non-experimental setting.

Hence, it would be interesting how these approaches could be integrated and the impact of such an initiative. We think that adopting a machine learning approach can improve the first step's accuracy in the IV analysis. We did not find any peer-reviewed publication that has empirically explored this. Though we did find suggestions to employ ML algorithms and econometric techniques together in publications such as Duncan (2014), Mullainathan and Spiess (2017), besides a series of papers by Athey and co-authors (Athey, 2015, 2018; Athey et al., 2019; Athey & Imbens, 2015, 2019). According to Duncan (2014), a fundamental problem of instrumental variable analysis is weak instruments without over-fitting. This problem can be overcome by using the random forest technique to predict the endogenous variable from the instrument. We have explored it for our analysis, as explained below.

An IV-2SLS model is represented as below (see Angrist and Pischke (2008)).

$$Y_i = X_i\beta + Controls_i + \varepsilon_i \quad (2)$$

$$X_i = Z_i\pi + Controls_i + \vartheta_i \quad (3)$$

$$Y_i = X_i\delta + Controls_i + \mu_i \quad (4)$$

Here $\delta = \pi\beta$, $\varepsilon = \mu - \beta\vartheta$. Equations (4.1), (4.2), and (4.3) are, respectively, called the structural, the first stage, and the reduced form representations. For the current scenario, $\mathrm{Y_i}$, $\mathrm{X_i}$, and $\mathrm{Z_i}$, respectively, denote the learning outcome (ability to write a paragraph), type of school (government versus others), and the instruments (free uniform and school fee paid by the government). The errors are clustered at the household level.

The conditions for an instrument Z's validity are strength, i.e., $\mathrm{Cov(Z, X)} \neq 0$; exclusion restriction, i.e., $\mathrm{Cov(Z, \varepsilon)} = 0$; and over-identification, the latter being done for multiple instruments. An instrument variable is weak if its correlation with the included endogenous regressor is small. With weak instruments, the IV-2SLS approach is biased towards OLS, and the 2SLS tests have the wrong size. To detect the strength of instruments, the F-statistic in the first stage should at least be 10. We also checked the standard errors between the OLS and IV models. If the errors are inflated by 5-10 times, then it could indicate weak instruments. The exclusion restriction criterion mandates that the IV should not be directly correlated with the main dependent variable. It can be justified on sound theoretical arguments. To assess it empirically, we regress the learning outcome against (a) only the instruments and controls, and (b) the endogenous variable, instruments, and controls. The instruments should be statistically significant in the case of (a) and non-significant (b) to validate the exclusion restriction criterion. To perform the over-identification test, we consider the Sargan-Hansen test with the following hypothesis. Rejection of the null hypothesis could imply that the instruments are correlated with residuals, or there are omitted variables in the model.

$$H0: E(Instruments, IV\ model\ residuals) = 0 \quad (5)$$

$$H1: E(Instruments, IV\ model\ residuals) \neq 0 \quad (6)$$

Fine-tuning of parameters to minimize the out-of-bag (OOB) error is an important feature in random forests. An OOB error is similar to n-fold cross-validation, so random forests can fit in one sequence, unlike many other non-linear operators (Hastie, Tibshirani, & Friedman, 2017). For tuning, two key parameters are the number of trees and the number of variables sampled at each stage. The number of trees should not be set too small to ensure that every input row gets predicted a few times. The default values for the number of variables sampled are $\sqrt{m}$ and $m/3$ for classification and regression, respectively, where m is the total number of variables. It is important to highlight the challenges faced in executing random forest algorithm with categorical variables. Presently, the R package can handle categorical variables with at most 53 categories. Therefore, variables like the district could not be used in our case. Instead, we used the state as a control variable for geographical location. This limit on the number of categories is one of the drawbacks, which is expected to be resolved as the machine learning execution packages develop further.

The current problem is a case of forbidden regression wherein the outcome and endogenous variables are binary. Forbidden regressions produce consistent estimates only under very restricted assumptions, which rarely hold in practice (Wooldridge, 2010). Because neither the conditional expectations operator nor the linear projection carries through non-linear functions making only an OLS regression in the first stage to produce fitted values uncorrelated with the

residuals (Greene, 2018), for this reason, the two-step probit needs to be changed (Wooldridge, 2010). This paper adopts the following approach:

Step1: Estimate the endogenous variable using the instruments and other explanatory variables.

Step2: Apply the two-step approach with the estimated values from step 1 as an instrument.

The estimates so obtained would be consistent, although less efficient than they would be if we were to take the endogenous variable's non-linear nature. However, the objective here is to get the marginal effects at the mean, so the linear probability model (IV-2SLS) used here suffices. Step 1 is executed with the random forest algorithm to integrate machine learning. This step helps in improving the estimation accuracy. Classification results from step 1 are compared – with and without using the random forest algorithm in the first step – via classification (confusion) matrix and area under the reverse operating characteristics (AUROC) curve.

RESULTS AND ANALYSIS

The first model is an OLS model without the instruments, the second one is the IV-2SLS with the three instruments, and this one suffers from the forbidden regression problem. The endogenous variable is estimated using a probit model with the instruments and control variables as the explanatory variables in the third model. For the last model, this estimation of the endogenous variable is done via the random forests model.

Due to space constraints, the estimates for only key variables have been tabulated here. Here, CS4_govt indicates the binary endogenous variable on enrollment in a government school while CS21 (free books), CS22 (free uniform), and CS23 (fee paid by the government) are the three instruments. The overall dependent variable (outcome) is the learning outcome, i.e., whether the student can write a paragraph with up to two mistakes or not. Table 17 shows that the estimate for enrollment in a government school deteriorated when OLS without instruments was replaced with an IV-2SLS. This deterioration indicates that government school students have an inferior learning outcome than their private school counterparts and government incentives – free books and uniforms, and the fee paid by the government – do not have a positive spillover effect on learning outcomes. The compliers here include those students who cannot afford to school but would join a government school if these incentives are provided.

To determine which of model 3 and model 4 yields more accurate estimates, we perform the AUC (see Figure 1) and classification matrix (see Table 19) analyses. They both show that the random forest-based IV-2SLS model is superior even though for the random forest model, only 70% of randomly-chosen data was used for model building and fine-tuning. In contrast, for the probit model, the full dataset was used for model building. For the random forest approach, the number of trees chosen was 15,000, and the number of variables randomly chosen was 5 for each of the trees. This configuration yielded the lowest OOB error of 6.64%. Contrary to the concerns around random forest being a black-box approach, we contend that it isn't entirely a black box as it does rank-order the variables based on their importance using the decrease in Gini as the metric. From Figure 2, we find that provision of free books (CS21) is the most important predictor for the decision to enroll in a government school, followed by free uniforms (CS22) and fees paid by the government (CS23).

Table 18: Second Stage Estimates

Variables	OLS without instruments (Model 1)	IV-2sls (Model 2)	IV-2sls-Modified (Model 3)	IV-2sls-RF Modified (Model 4)
CS4_govt – (Whether studying in a government school)				
Yes	-0.1555*** (0.0158)	-0.1883*** (0.0217)	-0.2147*** (0.0225)	-0.1815*** (0.0176)
RO3 (Gender) – Base is Male				
Female	-0.0034 (0.0116)	-0.0020 (0.0118)	-0.0003 (0.0117)	-0.0021 (0.0116)
RO5 (Age) – Base is 8 years				
9 years	0.0553*** (0.0176)	0.0532*** (0.0176)	0.0532*** (0.0176)	0.0543*** (0.0175)
10 years	0.0534*** (0.0188)	0.0486** (0.0188)	0.0487** (0.0189)	0.0512*** (0.0188)
11 years	0.0680*** (0.0209)	0.0625*** (0.0209)	0.0624*** (0.021)	0.0655*** (0.0208)
Urban2011 – Base is Rural				
Urban	0. 0179 (0.1248)	0.0107 (0.0127)	0.0104 (0.0127)	0.0146 (0.0125)
HHEDUC – Base is Never attended				
1st grade	0.0691 (0.0947)	0.0638 (0.0939)	0.0637 (0.0939)	0.0667 (0.094)
2nd grade	0.0623 (0.0452)	0.0614 (0.0449)	0.0613 (0.0449)	0.0619 (0.0449)
3rd grade	0.0321 (0.0422)	0.0267 (0.0419)	0.0264 (0.0419)	0.0296 (0.0419)
4th grade	0.0635* (0.0374)	0.0624* (0.0372)	0.0623* (0.0372)	0.063* (0.0372)
5th grade	0.0328 (0.0283)	0.0286 (0.0281)	0.0284 (0.028)	0.0309 (0.0281)
6th grade	0.1117*** (0.0307)	0.1072*** (0.0303)	0.107*** (0.0303)	0.1096*** (0.0304)
7th grade	0.0915*** (0.0302)	0.0884*** (0.0298)	0.0883*** (0.0297)	0.09*** (0.0299)
8th grade	0.076*** (0.0259)	0.0697*** (0.0255)	0.0694*** (0.0255)	0.0731*** (0.0256)
9th grade	0.1424*** (0.0235)	0.1361*** (0.0231)	0.1358*** (0.0231)	0.1395*** (0.0232)
10th grade	0.0763** (0.0296)	0.066** (0.0293)	0.0655** (0.0292)	0.0716** (0.0293)
11th grade	0.1307*** (0.0316)	0.1253*** (0.0311)	0.1251*** (0.0311)	0.1283*** (0.0312)
12th grade	0.1314*** (0.0276)	0.1213*** (0.0275)	0.1206*** (0.0275)	0.1268*** (0.0275)
Post 12th – Year 1	0.1384*** (0.0513)	0.1202** (0.0494)	0.1194** (0.0493)	0.1301** (0.0503)
Post 12th – Year 2	0.1421*** (0.0322)	0.1288*** (0.0326)	0.1282*** (0.0326)	0.136*** (0.0322)
Bachelor's	0.1673*** (0.0241)	0.1552*** (0.024)	0.155*** (0.0239)	0.1618*** (0.0239)

Above Bachelor's	0.1437*** (0.0266)	0.1306*** (0.0267)	0.13*** (0.0267)	0.1377*** (0.0265)
ID11 (Religion) – Base is Hindu				
Muslim	-0.0389** (0.0189)	-0.0391** (0.0188)	-0.0390** (0.0188)	-0.039** (0.0188)
Christian	0.0559 (0.0421)	0.0603 (0.0418)	0.0605 (0.0418)	0.0579 (0.0418)
Sikh	0.0298 (0.0396)	0.0283 (0.04)	0.0284 (0.04)	0.0291 (0.0396)
Buddhist	0.073 (0.0476)	0.075 (0.0478)	0.0751 (0.0479)	0.0739 (0.0475)
Jain	0.0895*** (0.0298)	0.0956*** (0.0314)	0.096*** (0.0315)	0.0923*** (0.0303)
Tribal	0.0396 (0.0925)	0.04 (0.0924)	0.0401 (0.0924)	0.0398 (0.0922)
Others	0.1407*** (0.052)	0.1461*** (0.0562)	0.1465*** (0.0563)	0.1432*** (0.0537)
None	0.1961* (0.112)	0.1863 (0.1282)	0.1860 (0.129)	0.1916 (0.119)
ID13 (Caste) – Base is Brahmin				
General (Except Brahmin)	-0.0384 (0.0256)	-0.0405 (0.0256)	-0.0412 (0.0257)	-0.0394 (0.0255)
OBC	-0.0328 (0.0256)	-0.034 (0.0256)	-0.0345 (0.0256)	-0.0334 (0.0255)
SC	-0.0947*** (0.0271)	-0.0913*** (0.027)	-0.0916*** (0.027)	-0.0932*** (0.0269)
ST	-0.1148*** (0.0318)	-0.1122*** (0.0317)	-0.1126*** (0.0317)	-0.1136*** (0.0316)
Others	0.036 (0.0584)	0.0397 (0.0583)	0.0394 (0.0583)	0.0377 (0.0582)

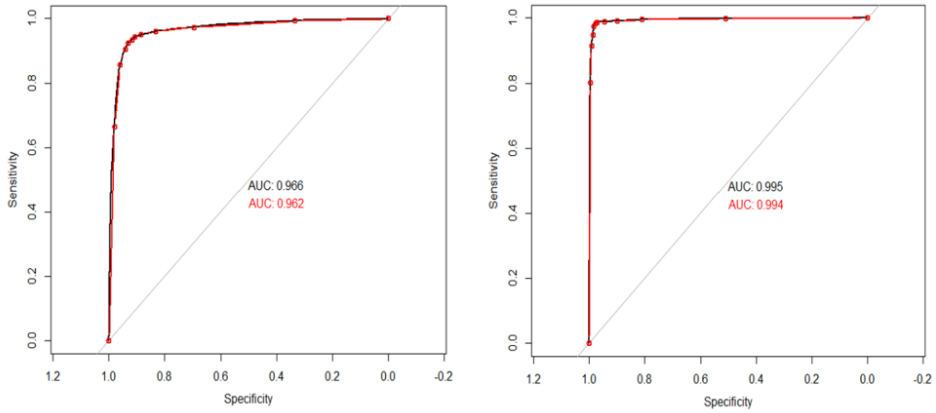
Source: Authors' computations, only the estimates for key variables have been tabulated due to space constraints

***Significant at 1% level **Significant at 5% level *Significant at 10% level

Table 19: Classification Matrix for Endogenous Variable Prediction using Probit model and Random Forest Algorithm

	Probit Model		Random Forest	
	Reference		Reference	
Prediction	CS4_govt = No	CS4_govt = Yes	CS4_govt = No	CS4_govt = Yes
CS4_govt = No	3396	375	3629	88
CS4_govt = Yes	327	5822	94	6109
Accuracy	92.92%		98.17%	
Sensitivity	91.22%		97.48%	
Specificity	93.95%		98.58%	
AUROC	0.96		0.99	

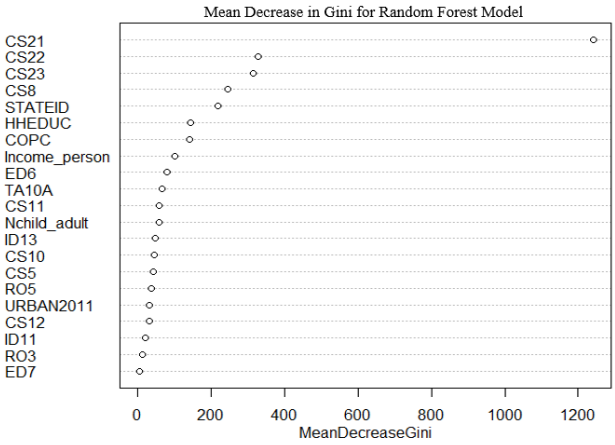
Figure 1: AUROC for Endogenous Variable Prediction using Probit Model and Random Forest Algorithm



Endogenous variable prediction using probit model (Model 3)

Endogenous variable prediction using random forest algorithm (Model 4)

Figure 2: Rank-Ordering of Variables by Mean Decrease in Gini using Random Forest Algorithm



We also performed the endogeneity test for the three IV-2SLS models. All of them rejected the null hypothesis of exogeneity. The instruments are strong as determined from the F-test of the first stage regression for models 2 through 4, shown in Table 20. The null hypothesis for Bassman and Sargam overidentifying restrictions is satisfied for model 2.

Table 20: First Stage Summary from Instrumental Variable Analysis

Model	Variable	R-square	Adjusted R-sq	Partial R-sq	Robust F (2, 5036)	Prob > F
Model 2	CS4_govt	0.7156	0.7124	0.4998	1164.24	0.0000
Model 3	CS4_govt	0.7417	0.7389	0.5459	3964.68	0.0000
Model 4	CS4_govt	0.9131	0.9121	0.8471	32819.5	0.0000

Source: Authors' computations, F statistic is adjusted for clusters in IDHH (Household id)

Table 21 is another way to test the exclusion restriction criteria empirically, as discussed earlier. When we include the instruments along with the endogenous variable, the instruments should ideally be non-significant at the 5% significance level. In this case, one instrument, CS21 (free books), is significant. However, we don't think this is a big challenge; theoretically, the

instruments are not correlated with the outcome variable. These government incentives are provided to encourage enrollment in schools, and merely enrolling does not translate into higher learning outcomes. The over-identification test has an insignificant chi-square statistic, which augurs well. Also, IV models' standard errors are not inflating much over the OLS values, as shown in Table 18, indicating the instruments are not weak.

Table 21: Estimates from regression with endogenous and instruments along with control variables

Variables	Coefficient	Robust Std. Error	T Statistic	P-value	95% Confidence Interval	
CS4_govt	-0.0987	0.0197	-5.0100	0.0000	-0.1374	-0.0601
CS21	-0.0661	0.0218	-3.0300	0.0020	-0.1089	-0.0233
CS22	-0.0167	0.0167	-1.0000	0.3170	-0.0494	0.0160
CS23	-0.0043	0.0176	-0.2500	0.8060	-0.0388	0.0302

Source: Authors' computations, estimates for only key variables have been tabulated due to space constraints

Therefore, we can conclude that the random forest-based IV-2SLS model works better than other versions of IV-2SLS and OLS. The government incentives for enrollment do not have a positive spillover effect on the learning outcomes of government school children aged 8-11. Having the first step of IV-2SLS via a machine learning algorithm such as random forest yields estimates closer to the population parameters and should be used for causation analysis.

DISCUSSION

There is a growing focus in the extant research and policy on the quality of schooling since this determines the future trajectory in terms of educational, intellectual, and economic attainment of not just an individual but also society as a whole. Education is considered one of the core functions to converge towards an egalitarian society (Gouda et al., 2013). As Chatterji (2008) asserts, the externalities generated by primary education may render it more important than what its relatively low private rate of return may show. One of the many changes in the post-liberalization era (1991 onwards) in India is the rise in private schooling (French & Kingdon, 2010), which are perceived to be much better and regarded as the first step towards soaring in life. However, a key challenge is to isolate the actual effect of private schooling on students' performances from other factors that may influence the outcome. The heterogeneity in the private and government schools' quality is reflected in the students' performance, which should ease policymakers to make appropriate decisions and raise government schools' teaching and infrastructural standards. The findings indicate that the government incentives to improve enrollment in government schools do not lead to an improved learning outcome, which is a key takeaway for policymakers to create policies towards the latter.

Machine learning methods have been productive in applications ranging from medicine to allocating fire and health inspectors in cities. However, several gaps exist between making a prediction and making a decision, and underlying assumptions need to be understood to optimize data-driven decision-making (Athey, 2017). The past decade's data revolution is likely to have a further and profound effect on economic research (Einav & Levin, 2014). The current analysis is an initial step to explore an integrated approach consisting of econometric modeling and machine learning algorithms. This also opens up vistas of diverse research opportunities. On the methodological front, as discussed earlier, forbidden regression is being addressed by

new methods such as the special regressor method. A future step could be to explore the integration of machine learning with such econometric techniques.

An important limitation of this study is the restricted sample of age 8-11 years for whom the test scores were available. It is being assumed that writing ability is a good indicator of one's cognitive ability. However, there are other ways to gauge one's cognitive ability. These need to be analyzed to understand more deeply if there is a positive spillover impact of government incentives on the learning outcomes in government schools. Future analysis could also explore techniques such as regression discontinuity design to identify the income band for the compliers of this analysis. All of these are critical to understanding because selecting the school type will have a long-term impact on the earnings of an individual, which would affect the overall growth and development when aggregated at the state and country-level. Therefore, this inquiry is of considerable importance from a policy perspective.

CONCLUSION

"*I believe that machine learning (ML) will have a dramatic impact on the field of economics within a short timeframe. Indeed, the impact of ML is already well underway, and so it is perhaps not too difficult to predict some of the effects.*"

- Athey (2018)

The current analysis has been done with this aim and is expected to be a guiding tool for academic researchers, practitioners, and educational policymakers. The inquiry becomes overly critical, given the NEP's focus is to elevate India's education system on a holistic level. The paper shows how integrating the random forest algorithm improves the causal estimates obtained from an IV-2SLS approach. The findings reveal no positive spillover effect of government incentives – free books, free uniforms, and fees paid by the government – on the learning outcomes of students aged 8-11 years in government schools. This analysis is pertinent as education is a quintessential element for human beings to evolve and contribute to society. It is both the quality of education and its accessibility to everyone that translate into improved learning outcomes for the students. Elementary education is the stepping stone to attain secondary and higher education. Therefore, its foundation needs to be strong. The relevant question is not just whether children are attending school, but also how well they learn and hone their skills during the formative years. The focus should be on improving the learning outcomes as these would help in annulling the impact of the socio-economic disparities on primary education.

REFERENCES

Abadie, A., & Kasy, M. (2017). The risk of machine learning. *arXiv preprint arXiv:1703.10935*.

Acharya, P. (1994a). Problems of Universal Elementary Education. *Economic and Political Weekly, 29*(49), 3098-3105.

Acharya, P. (1994b). Universal Elementary Education: Receding Goal. *Economic and Political Weekly, 29*(1/2), 27-30.

Afridi, F. (2005). Midday Meals in Two States: Comparing the Financial and Institutional Organisation of the Programme. *Economic and Political Weekly, 40*(15).

Afridi, F. (2010). Child welfare programs and child nutrition: Evidence from a mandated school meal program in India. *Journal of Development Economics, 92*, 152-165.

Afridi, F. (2011). The Impact of School Meals on School Participation: Evidence from Rural India. *Journal of Development Studies, 47*(11), 1636-1656. doi:10.1080/00220388.2010.514330

Aggarwal, Y. (2000). *Public and Private Partnership in Primary Education in India*. Operations Research and Systems Management Unit. National Institute of Educational Planning and Administration. New Delhi.

Agrawal, T., & Agrawal, A. (2018). Who Gains More from Education? A Comparative Analysis of Business, Farm and Wage Workers in India. *The Journal of Development Studies, 55*(6), 1081-1098. doi:10.1080/00220388.2018.1443209

Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2018). *LASSOPACK and PDSLASSO: Prediction, model selection and causal inference with regularized regression.* Paper presented at the London Stata Conference 2018.

Alcott, B., & Rose, P. (2015). Schools and learning in rural India and Pakistan: Who goes where, and how much are they learning? *Prospects, 45*(3), 345-363. doi:10.1007/s11125-015-9350-5

Alcott, B., & Rose, P. (2017). Learning in India's primary schools: How do disparities widen across the grades? *International Journal of Educational Development, 56*, 42-51. doi:10.1016/j.ijedudev.2017.05.002

Altman, N., & Krzywinski, M. (2017). Points of Significance: Ensemble methods: bagging and random forests. In: Nature Publishing Group.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*: Princeton university press.

Apte, M. (1979). Priority Needs of Children in India. *Indian Journal of Public Administration, 25*(3), 549-568.

Asadullah, M. N., & Yalonetzky, G. (2012). Inequality of Educational Opportunity in India: Changes Over Time and Across States. *World Development, 40*(6), 1151-1163. doi:10.1016/j.worlddev.2011.11.008

Athey, S. (2015). *Machine learning and causal inference for policy evaluation.* Paper presented at the Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science, 355*(6324), 483-485. Retrieved from https://science.sciencemag.org/content/sci/355/6324/483.full.pdf

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507-547): University of Chicago Press.

Athey, S., Bayati, M., Imbens, G., & Qu, Z. (2019). *Ensemble methods for causal effects in panel data settings.* Paper presented at the AEA Papers and Proceedings.

Athey, S., & Imbens, G. W. (2015). Machine learning methods for estimating heterogeneous causal effects. *stat, 1050*(5), 1-26.

Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments* (Vol. 1, pp. 73-140): Elsevier.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*, 685-725.

Au, T. C. (2018). Random forests, decision trees, and categorical predictors: the" Absent levels" problem. *The Journal of Machine Learning Research, 19*(1), 1737-1766.

Azam, M. (2017). Explaining Caste Differences in Private School Attendance. *Review of Development Economics, 21*(4), 1191-1204. doi:10.1111/rode.12305

Azam, M., Kingdon, G., & Wu, K. B. (2016). Impact of Private Secondary Schooling on Cognitive Skills: Evidence from India. *Education Economics, 24*(5), 465-480.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy, 2*(1), 1-30.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*.

Banerji, R. (2000). Poverty and primary schooling: field studies from Mumbai and Delhi. *Economic and political weekly*, 795-802.

Banerji, R., & Mukherjee, A. N. (2008). Achieving Universal Elementary Education in India. *Margin: The Journal of Applied Economic Research, 2*(2), 213-228. doi:10.1177/097380100800200204

Bhattacharya, S., Dasgupta, A., Mandal, K., & Mukherjee, A. (2015). Understanding the “Sorting Hat”: The Role of Family and Caste Network in School Choice Decision. *ESP Working Paper Series*(69).

Bordoloi, R. (2011). Challenges in Elementary Education in India: Various Approaches. *Journal of Education and Practice, 2*(7).

Borooah, V. K. (2012). Social Identity and Educational Attainment: The Role of Caste and Religion in Explaining Differences between Children in India. *Journal of Development Studies, 48*(7), 887-903. doi:10.1080/00220388.2011.621945

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Buskirk, T. D. (2018). Surveying the forests and sampling the trees: an overview of classification and regression trees and random forests with applications in survey research. *Surv Pract, 11*, 2709.

Buskirk, T. D., Kirchner, A., Eck, A., & Signorino, C. S. (2018). An introduction to machine learning methods for survey researchers. *Survey Practice, 11*(1), 2718.

Buskirk, T. D., & Kolenikov, S. (2015). Finding respondents in the forest: A comparison of logistic regression and random forest models for response propensity weighting and stratification. *Survey Methods: Insights from the Field*, 1-17.

Caiola, G., & Reiter, J. P. (2010). Random Forests for Generating Partially Synthetic, Categorical Data. *Trans. Data Privacy, 3*(1), 27-42.

Čeh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information, 7*(5), 168.

Chaudhary, L., Musacchio, A., Nafziger, S., & Yan, S. (2012). Big BRICs, weak foundations: The beginning of public elementary education in Brazil, Russia, India, and China. *Explorations in Economic History, 49*(2), 221-240. doi:10.1016/j.eeh.2012.02.001

Chauhan, C. P. S. (2009). Education for all in India: A second look. *International Journal of Lifelong Education, 28*(2), 227-240. doi:10.1080/02601370902757091

Chudgar, A. (2012). Variation in Private School Performance: The Importance of Village Context. *Economic and Political Weekly, 47*(11), 52-59.

Chudgar, A., & Creed, B. (2014). *How does Demand for Private Schooling Vary Across Locations with Different Private School Supply? Analysis of Data from Rural India*. Michigan State University.

Chudgar, A., & Creed, B. (2016). How are private school enrolment patterns changing across Indian districts with a growth in private school availability? *Oxford Review of Education, 42*(5), 543-560. doi:10.1080/03054985.2016.1217688

Chudgar, A., & Quin, E. (2012). Relationship between private schooling and achievement: Results from rural and urban India. *Economics of Education Review, 31*(4), 376-390. doi:10.1016/j.econedurev.2011.12.003

Clots-Figueras, I. (2012). Are female leaders good for education? Evidence from India. *American Economic Journal: Applied Economics, 4*(1), 212-244.

Colclough, C. (1982). The Impact of Primary Schooling on Economic Development: A Review of the Evidence. *World Development, 10*(3), 167-185.

Das, U., & Singhal, K. (2017). *Gender Differences in Mathematics Performance: Evidence from Rural India*. Paper presented at the IARIW-ICIER, New Delhi.

Dave, M. (2013). *When the Lights Go on: Household Electrification and Educational Outcomes in Rural India.* (Master of Public Policy). Georgetown University,

De, A., Noronha, C., & Samson, M. (2003). Private Schools for Less Privileged: Some Insights from a Case Study. *Economic and Political Weekly, 37*(52), 5230-5236.

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments, 19*(2), 138-163.

Desai, S., Dubey, A., Banerji, R., & Vanneman, R. (2009). Private schooling in India: A new educational landscape. In S. Bery, B. Bosworth, & A. Panagariya (Eds.), *India Policy Forum 2008-09* (Vol. 5, pp. 1): SAGE Publications India.

Desai, S., & Vanneman, R. (2015). India human development survey-ii (ihds-ii), 2011-12. *ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 31*.

Desai, U. (1991). Determinants of educational performance in India: role of home and family. *International Review of Education, 37*(2), 245-265.

Dev, M. (2016). Factors Affecting the Academic Achievement: A Study of Elementary School Students of NCR Delhi, India. *Journal of Education and Practice, 7*(4).

Dreze, J., & Sen, A. (2013). *India: Development and Participation* (2 ed.): Oxford University Press.

Duncan, G. M. (2014). Causal random forests. In.

Earp, M., Mitchell, M., McCarthy, J., & Kreuter, F. (2014). Modeling nonresponse in establishment surveys: Using an ensemble tree model to create nonresponse propensity scores and detect potential bias in an agricultural survey. *Journal of Official Statistics, 30*(4), 701-719.

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science, 346*(6210), 1243089. Retrieved from https://science.sciencemag.org/content/346/6210/1243089.long

French, R., & Kingdon, G. (2010). *The Relative Effectiveness of Private and Government Schools in Rural India: Evidence from ASER data*. DoQSS Working Papers. UCL Institute of Education, University College London.

Galab, S., Reddy, P. P., & Reddy, V. N. (2014). *Classroom Process Teacher Ability and Student Performance: Evidence from School based Component of Young Lives in Undivided Andhra Pradesh*. Working Paper No. 134. Centre For Economic and Social Studies.

Gouda, J., Chandra Das, K., Goli, S., & Maikho Apollo Pou, L. (2013). Government versus private primary schools in India. *International Journal of Sociology and Social Policy, 33*(11/12), 708-724. doi:10.1108/ijssp-12-2012-0105

Goyal, S. (2007). Learning achievements in India: A study of primary education in Rajasthan. *Human Development Unit, South Asia Region, The World Bank, Washington DC*.

Goyal, S. (2009). Inside the house of learning: the relative performance of public and private schools in Orissa. *Education Economics, 17*(3), 315-327. doi:10.1080/09645290903142577

Goyal, S., & Pandey, P. (2009). *How do Government and Private Schools Differ? Findings from two large Indian states*. Retrieved from

Greene, W. H. (2018). Econometric Analysis. In: Prentice Hall, Englewood Cliffs, NJ.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning. In (12 ed.): Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

Hill, E., Samson, M., & Dasgupta, S. (2011). Expanding the school market in India: Parental choice and the reproduction of social inequality. *Economic and political weekly*, 98-105.

Huisman, J., Rani, U., & Smits, J. (2010). School characteristics, socio-economic status and culture as determinants of primary school enrolment in India. *Nijmegen Center for Economics (NiCE) Working Paper 10-109*.

Jain, N. (2020). Do School Ownership and Incentive Mechanisms Affect Learning Outcomes? An Analysis from Rural India. *International Journal for Infonomics, 13*, 1962-1968.

Jayaraman, R., & Simroth, D. (2015). The Impact of School Lunches on Primary School Enrollment: Evidence from India's Midday Meal Scheme Impact of school lunches on primary school enrollment. *The Scandinavian Journal of Economics, 117*(4), 1176-1203.

Jha, P. (2007). Guaranteeing Elementary Education: A Note on Policy and Provisioning in Contemporary India. *Journal of South Asian Development, 2*(1), 75-105. doi:10.1177/097317410600200104

Kaushal, M. (2012). Implementation of Right to Education in India: Issues and Concerns. *Journal of Management and Public Policy, 4*(1), 42-48.

Khera, R. (2006). Mid-day meals in primary schools: Achievements and challenges. *Economic and political weekly*, 4742-4750.

Kingdon, G. (1996). The Quality and Efficiency of Private and Public Education: A Case Study of Urban India. *Oxford Bulletin of Economics and Statistics, 58*(1).

Kingdon, G. (2008). School-sector Effects on Student Achievement in India. In R. Chakrabarti & P. E. Peterson (Eds.), *School Choice International: Exploring Public-Private Partnerships* (pp. 111-139). Cambridge: MIT Press.

Kingdon, G., & Teal, F. (2010). Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach. *Journal of Development Economics, 91*(2), 278-288.

Kingdon, G. G. (1996). Private Schooling in India: Size, Nature, and Equity-Effects. *Economic and Political Weekly, 31*(51), 3306-3314.

Kingdon, G. G. (2017). *The Private Schooling Phenomenon in India: A Review*. IZA Discussion Papers. Institute of Labor Economics (IZA). Bonn.

Kingdon, G. G., & Teal, F. (2007). Does performance related pay for teachers improve student performance? Some evidence from India. *Economics of Education Review, 26*(4), 473-486.

Kumar, S. M. (2017). Comparing private and government schools in India: the devil is in the maths. *Applied Economics Letters, 25*(6), 409-414. doi:10.1080/13504851.2017.1327118

Kurien, J. (1981). Towards Universal Elementary Education: Promise and Performance. *Economic and Political Weekly, 16*(40), 1608-1618.

Liaw, A., & Wiener, M. (2015). randomForest: Breiman and Cutler's random forests for classification and regression. *R package version, 4*, 6-10.

Little, A. W. (2010). *Access to Elementary Education in India: Politics, Policies and Progress*. Retrieved from Brighton, UK.:

Mehrotra, S., & Panchamukhi, P. R. (2006). Private provision of elementary education in India: findings of a survey in eight states. *Compare: A Journal of Comparative and International Education, 36*(4), 421-442. doi:10.1080/03057920601024883

Mishra, R. (1999). Research on education in India. *Prospects, 29*(3), 335-347.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106.

Muralidharan, K., & Kremer, M. (2006). *Public and private schools in rural India*. Department of Economics. Harvard University. Cambridge, MA.

Muralidharan, K., & Prakash, N. (2017). Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics, 9*(3), 321-350. doi:10.1257/app.20160004

Murugesan, S. S. (2015). *Does Medium of Instruction matter? A propensity score analysis of the effect of English Medium Education on students learning outcomes in India.* (Master of Public Policy). Oregon State University,

Nair, P. S. K. (2015). *Does Medium of Instruction Affect Learning Outcomes?—Evidence Using Young Lives Longitudinal Data of Andhra Pradesh, India*. ESP Working Paper Series. Privatisation in Education Research Initiative (PERI).

Nanjunda, D. C., & Ramesh, C. (2010). Elementary Education in Rural Karnataka - An Analysis. *Artha Journal of Social Sciences, 9*(2), 1-11.

Oster, E., & Steinberg, B. M. (2013). Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics, 104*, 123-135. doi:https://doi.org/10.1016/j.jdeveco.2013.05.006

Ota, M., & Moffatt, P. G. (2006). The within-household schooling decision: a study of children in rural Andhra Pradesh. *Journal of Population Economics, 20*(1), 223-239. doi:10.1007/s00148-005-0033-z

Pal, S., & Kingdon, G. G. (2010). *Can Private School Growth Foster Universal Literacy? Panel Evidence from Indian Districts*. IZA Discussion Papers. Institute for the Study of Labor (IZA). Bonn.

Rai, A. (2014). Implementation of the RTE Act: RTE Forum's Stocktaking Report. *Social Change, 44*(3), 439-449. doi:10.1177/0049085714536806

Reddy, P. P., & Galab, S. (2018). Schooling and Learning Outcomes in Andhra Pradesh: An Evidence from Young Lives – A Longitudinal Study. *IASSI Quarterly: Contributions to Indian Social Science, 37*(2).

Rosenthal, D. B. (1974). Educational Politics and Public Policymaking in Maharashtra, India. *Comparative Education Review, 18*(1), 79-95.

Singh, A. (2008). Do school meals work? Treatment evaluation of the midday meal scheme in India. *Young Lives*.

Singh, R., & Sarkar, S. (2015). Does teaching quality matter? Students learning outcome related to teaching quality in public and private primary schools in India. *International Journal of Educational Development, 41*, 153-163. doi:10.1016/j.ijedudev.2015.02.009

Tilak, J. B. G. (1995). Elementary Education in India in the 1990s: Problems and Perspectives. *MARGIN*.

Tooley, J. (2016). Extending access to low-cost private schools through vouchers: an alternative interpretation of a two-stage 'School Choice' experiment in India. *Oxford Review of Education, 42*(5), 579-593. doi:10.1080/03054985.2016.1217689

Tooley, J., & Dixon, P. (2006). 'De facto' privatisation of education and the poor: implications of a study from sub-Saharan Africa and India. *Compare: A Journal of Comparative and International Education, 36*(4), 443-462. doi:10.1080/03057920601024891

Wadhwa, W. (2009). Are Private Schools Really Performing Better than Goverment Schools? *ASER*.

Woodhead, M., Frost, M., & James, Z. (2013). Does growth in private schooling contribute to Education for All? Evidence from a longitudinal, two cohort study in Andhra Pradesh, India. *International Journal of Educational Development, 33*(1), 65-73. doi:10.1016/j.ijedudev.2012.02.005

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*: MIT press.

Zimmermann, L. (2020). Remember when it rained–Schooling responses to shocks in India. *World Development, 126*, 104705.

DECISION SCIENCES INSTITUTE

The influence of gaining employment on individual prescription drug utilization in the U.S., 2007-2016

Xiaohui You
Metropolitan State University of Denver
Email: brightdawn99@gmail.com

Cyril Chang
The University of Memphis
Email: cchang@memphis.edu

ABSTRACT

This paper examines the relationships among gaining employment, individual health status, and changes in prescription drug utilization in the US. Statistical models based on the Grossman framework using three-stage least squares techniques and the Medical Expenditure Panel Survey data, 2007-2016. We find that an individual's drug utilization is positively associated with the gaining employment process and improved health status, controlling for age, wages, education, and uninsured status. Results suggest a pattern in prescription utilization: the longer an individual gets a job, the lower its rate of drug utilization. Most importantly, results confirm the benefit of insurance coverage expansion under ACA.

KEYWORDS: prescription drug utilization, employment, health status, Grossman model, three stage least squares

INTRODUCTION

More recent, rising prescription drug prices and diminished innovation by drug companies are calling for federal action. Consumers across America and the health care system cannot continue to bear the burden of high and rising prescription drug costs. Trump Administration announced several rounds of drug pricing proposals, and it took steps toward future reforms in the hope to significantly reduce the prescription drug burden for both consumers and the overall health care system.

Several significant factors could influence prescription drug utilization and spending, such as health status, employment status, insurance status, income, etc. The interrelationships among employment, health status, and healthcare care utilization are complex. An individual's employment status, especially the loss of a job, can reduce healthcare use (including prescription drug utilization). On the other hand, the loss of a job might be the result of poor health, which could lead to increased use of healthcare services and prescription drugs. Further complicating the interrelationships is the feedback of healthcare and prescription drug uses that can improve health and the likelihood of getting a job.

Opportunistically, the severe recession triggered by the 2008 financial crisis and the subsequent slow recovery over the next decade provides a rare opportunity to empirically investigate the complex relationships among employment, individual health, and prescription drug utilization. With the country in a deep recession, 2007-2009, the American economy experienced lower growth and higher unemployment for a long time. According to a 2010 Commonwealth Fund survey, about 9 million working-age adults lost their health insurance in 2008-2010; besides, 75 million adults did not get needed healthcare with many skipping prescription drugs, doctor visits,

and recommended tests and treatments (Theodorou et al. 2011). Some argue that there might be a fundamental change in Americans' preferences for healthcare and utilization patterns of healthcare because many people are extremely budgeting conscious (Fodman and Book 2010).

The great recession coincided with the beginning of the Obama Administration, which launched the Patient Protection and Affordable Care Act (ACA) in 2010. ACA's ambitious goals are to increase the quality of health care and affordability of health insurance while simultaneously lowering the uninsured rates and costs of healthcare. It is of policy significance to determine whether ACA has achieved the intended effects of reducing the number of uninsured and increasing the use of health care such as prescription drugs.

The purpose of this study is to investigate the association between the interlocking influences of a change in an individual's employment status, personal health (individual perceived health status, less healthy if < Good and healthy if Good, or > Good), and his/her drug utilization. To capture this complicated relationship for an empirical investigation, we developed two-stage least squares and three-stage least squares statistical models based on the Grossman theory of healthcare utilization. The data are from the Medical Expenditure Panel Survey (MEPS), a national representative sample of individual utilization of health services and prescription drugs, for 2007-2016, a period beginning from the economic recession and recovery to the implementation of ACA.

The rest of the paper proceeds as follows. Section II reviews the earlier research about the effects of employment factors on drug utilization. Section III derives the theoretical framework and presents the empirical models. Section IV describes the data and variables used in this study, and Section V focuses on the empirical results. Finally, Section VI concludes with reflections on some policy implications of the findings.

LITERATURE REVIEW

The research on employment, health insurance, and prescription drug utilization is not new. For example, Bass and Woodward (1978) conducted a drug treatment experiment to demonstrate the efficacy of pooling community employment-related services to assist clients in drug addiction treatment. In the same vein, Silverman and Robles (1999) suggested that unemployment and prescription drug use were firmly related to observational data from various sources. The Grossman (1972) model of health demand proposed that the changing of employment and health status affecting healthcare utilization (including prescription drug utilization) when outlined the interaction between a demand function for health and a production function for health.

Using logistic regression techniques and controls for differences in socio-economic characteristics, Stronks et al. (1997) found a close association between income and health of individuals. The association was particularly strong due to employment status, especially among men. For women, however, the income was still found to be related to health controlling for employment status. Other researchers have focused on the relationships between unemployment and unhealthy health behaviors. Henkel (2011) summarized the results of a comprehensive review of research findings on the effect of unemployment on such preventive services as alcohol and drug addiction treatment and smoking cessation and found that unemployment increased the risk of relapse after alcohol and drug addiction treatment.

Recently, researchers have investigated the effects of insurance coverage and healthcare access and utilization. Using Census Bureau's Current Population Survey (CPS) data, Holahan (2011) found that the number of uninsured nonelderly Americans increased by 5.6 million, from 45 to 50 million during 2007-2009 employer-sponsored insurance plans experiencing a more considerable drop in the rate of coverage than public insurance programs. Researchers such as Truffer et al. (2010) found out that the rising unemployment rates and the loss of insurance

subsidies were responsible for the observed slowing down of the growth of health care spending during the Great Recession.

At the individual level, Lusardi et al. (2010) and Lusardi et al. (2015) pointed out that more than a quarter of Americans had reported reducing their routine healthcare use during the great recession, not only among the unemployed but also the insured were not immune to the harmful effects of the recession. The pressure to cut business expenses had forced employers to raise copays to shift healthcare costs to their employees, thus negatively affecting the prescription drug use of those who remain employed. Based on the patient-level data from a large firm that had increased employee copayments on prescription drugs, Gibson et al. (2005) found that higher copayments were a significant barrier to prescription drug use, especially among employees with an existing chronic condition or a newly diagnosed chronic illness.

Our brief review of the literature shows that a large body of research on prescript drug utilization exists. However, the existing literature suffers from two major interrelated methodological weaknesses. First, most previous studies were empirical did not base on a broad-based conceptual framework (for example, the Grossman model we use in this study) describing the interlocking relationships among prescription drug use, employment, and individual health status. Second, most previous studies used a single-equation approach intuitively to determine the effects of employment and health on drug utilization, thus ignored the likelihood of the mutually dependent nature of the relationships. A single equation model can adequately investigate the relationship of association between two variables if the underlying relationship is one way; that is, the independent variable is genuinely independent, not dependent on other variables. In the interrelationship between drug use, employment, and individual health, the single-equation approach violates the underlying assumptions of regression analysis. It gives rise to inefficient and biased estimates.

Therefore, in this study, using high-quality Medical Expenditure Panel Survey (MEPS) individual data, 2007-2016, we estimate the simultaneous equation system with the Grossman theoretical models to address the effects of the individual's health and employment status on prescription drug uses. Using Two Least Squares (2SLS) and Three Least Squares (3SLS) approaches and controlling for gender, insurance status, wages, and initial incomes, we investigate the association among prescription drug utilization, the gaining of health status, and the gaining of employment for the working-age individuals (age 18-65).

THE EMPIRICAL MODEL

We analyze a theoretical framework based on the Grossman model of health care demand and production (1972; 1999), sketching the theory and describing the empirical model. The conceptual framework begins with a health gross investment function in the Cobb-Douglas form:

$$I_i = EM_i^{1-\alpha} TH_i^{\alpha} X_i^{1-\beta} T_i^{\beta} \quad (1)$$

where I_i is the gross investment in health (using the prescription drug), where M_i and X_i are the inputs purchased in the market (M_i is the person investments in health, prescription medicines in our study), TH_i and T_i is the time input. E_i is a stock of human capital, education, which is assumed exogenous.

The time constraint implies that total time available (Ω) allocated across working time (TW), time lost from the market and nonmarket activities (TL), the time investment in the production of health (TH), and time spent in producing other goods or leisure time (T). Since $\Omega = TW_i + TL_i + TH_i + T_i$ in the Grossman framework, we replace T_i of equation (1) by $(\Omega - TW_i - TL_i - TH_i)$ and obtain:

$$I_i = EM_i^{1-\alpha} X_i^{1-\beta} TH_i^{\alpha} (\Omega - TW_i - TL_i - TH_i)^{\beta} \quad (2)$$

With the elasticity, $\varepsilon = \frac{1}{1+C}$, the basic structural equations with the production function of health, equation (1) becomes,

$$lnH_i = \varepsilon \ln W_i - \varepsilon \ln \pi_i - \varepsilon ln\delta_i \quad (3)$$

Where W_i represents the wage while employed and π_i is the marginal cost of gross investment in prescription drug spending (RXS) in period i with the rate of depreciation (δ_i). Grossman's notion that health stock as a capital good, enhanced by investment in health and diminishes with age, accidents, and risky behaviors (1972; 1999). To be specific, H_i, the ith individual's health stock is a function of wage income, the individual's health expenditures (include spending on prescription drugs), and the rate of health depreciation. H_i varies with age, a variable with direct or indirect effects (through the changing of health status) on health care services expenditures.

With $\Omega = TW_i + TL_i + TH_i + T_i$, the health gross investment function becomes,

$$\ln I_i \equiv lnH_{i+1} + ln(1 + H_i/\delta_i) = \ln E_i + (1-\alpha)\ln M_i + \alpha \ln TH_i + (1-\beta)\ln X_i + \beta \ln(\Omega - TW_i - TL_i - TH_i) \quad (4)$$

where E_i, the human capital-related factors, such as *le* (education), gender, and race; and other inputs commodity-related factors, X_i, such as initial income or discounted property, insurance status…etc., those exogenous variables on healthcare expenditures.
Subtracting (3) from (4), then

$$lnH_{i+1} + ln(1 + H_i/\delta_i) - lnH_i = \ln E_i + (1-\alpha)\ln M_i + \alpha \ln TH_i + (1-\beta)\ln X_i + \beta \ln(\Omega - TW_i - TL_i - TH_i) - \varepsilon \ln W_i + \varepsilon \ln \pi_i + ln\delta_i \quad (5)$$

After rearranging, the equation becomes,

$$\underbrace{lnH_{i+1} + ln(1 + H_i/\delta_i) - lnH_i}_{\Delta HG} = \ln E_i + (1-\alpha)\ln M_i + (1-\beta)\ln X_i + \varepsilon \ln \pi_i - \varepsilon \ln W_i + ln\delta_i + \underbrace{\alpha \ln TH_i + \beta \ln(\Omega - TW_i - TL_i - TH_i)}_{\Delta UG} \quad (6)$$

where $lnH_{i+1} + ln(1 + H_i/\delta_i) - lnH_i$ in equation (6), is the stock of the health of a particular individual produces a flow of healthy status relative to his age depreciation and previous health status in a given two-year period (Δ*HG,* label as *hg*), which will be the equation (9) below. $\alpha \ln TH_i + \beta \ln(\Omega - TW_i - TL_i - TH_i)$ in equation (6) is the individual's working time distribution and allocation, which is the gaining of employment process during the five rounds in the two-years (Δ *UG,* label as *eg*) period. This equation precisely describes the changing status from unemployment to employment. In other words, the leisure time relates to the time with employment and the time under unemployment (Equation (10) below).

The Grossman model's main point is that it is the changing of employment and health status affecting healthcare utilization (including prescription drug utilization) rather than the employment and health status. Specifically, how long it takes an individual from unemployment to employment (gaining employment process) and the gaining health status from unhealthy. To date, no study addressing this point, previous literature either focus on employment or

unemployment status or unhealthy conditions effects on healthcare utilization. One important reason might be that it is difficult to find such data to address this scenario. This framework's complex nature makes it difficult to empirically model the effects of those changes in employment and health status on health care utilization.

The MEPS Longitudinal Dataset has five rounds every two years, which provided us the necessary information to calculate the gaining of employment (∆UG, or *eg*) and health status (∆HG, or *hg*). We could then capture how the marginal benefits (employment and better health status) equal the marginal cost of capital (prescription drug) in terms of the gross investment. The complicated relationship between drug utilization (quantity, *rxq*) and spending (*rxs*) is also a problem. To be specific, drug utilization and expenditure (RX) can affect each other (Seemingly unrelated), while *eg* and *hg* also affect RX. mutually (Endogeneity). The empirical models of structural prescription drug utilization equations (investment), health and employment production function corresponding to the empirical model in this study are as follow,

$$rxq = \beta_0 + \beta_1 hg + \beta_2 eg + \beta_3 age + \beta_4 lw + \beta_5 le + \beta_6 rxs + \beta_7 lnX + \varepsilon \quad (7)$$

$$rxs = \alpha_0 + \alpha_1 hg + \alpha_2 eg + \alpha_3 age + \alpha_4 lw + \alpha_5 le + \alpha_6 lnX + \epsilon \quad (8)$$

$$hg = \gamma_0 + \gamma_1 rx + \gamma_2 le + \gamma_3 age + \gamma_4 X + \mu \quad (9)$$

$$eg = \theta_0 + \theta_1 rx + \theta_2 erg + \theta_3 lw + \theta_4 X + \nu \quad (10)$$

where the lowercase variables are in natural logarithm, and all variables defined as above. If we replace $\ln M_i$ with *rxq*, after rearranging the terms, the empirical model equation (6) becomes to equation (7), the demand for prescription drug equation; if we replace the $\ln \pi_i$ with *rxs*, then the equation (6) becomes equation (8), the prescription drug spending equation. Equation (9), the gaining of the healthy status equation, derived from the empirical model equation (3) and Δ*HG* term in equation (6); and equation (10), the gaining of employment status equation, is derived from Equation (6), the corresponding MEC schedule in the Grossman model associated with the changing status from unemployment to employment, Δ*UG* term, and *erg* (regional employment-population ratio), *lw* (wages), and other variables, such as *linc* (non-wage income). Equation (9) and (10) here are to address the endogeneity problem with *rxq* and *rxs*, the equation (7) and (8).

Since the goal of this study is to address the effects of the changing employment status on prescription drug utilization, estimations are all restricted to individuals age ranging from 18 (at the end of the first year in the two years) to 65 (at the end of the two years). By clustering the dwelling unit ID (DUID) to correct the potential problem of heteroscedasticity, we estimate the empirical models of structural prescription drug utilization and spending equations (7) to (8). Beginning with the ordinary least squares (OLS) regression, we can get a general picture of the relationship between the dependent variable (*rxq*, *rxs*) and the independent variables. The simultaneous equation models, Two-Stage Least Squares (2SLS) are employed to obtain unbiased and efficient estimates of structural parameters. Based on the 2SLS results, we finally estimate the Three-Stage Least Squares (3SLS) with equations (7, 8, 9 and 10) to address the endogeneity of *eg* and *hg* on both *rxq* and *rxs* and correct the duplicates problem from the 2SLS. To be specific, *age*, affects RX. directly, and also affects RX. indirectly through the changing of health status, where *age* is an instrument variable in equation (9). Same for the *wage*, it affects RX. directly, and also affects RX. indirectly through the changing of employment status, where *wage* is an instrument variable in equation (10).

In the 2SLS approaches, we high a set of exogenous variables as instruments (wages, *lw*; age, *age*; non-wage income, *linc*; regional employment-population ratio, *erg*, and education, *le*) to handle the endogeneity of the gaining employment and health status variables (*eg* and *hg*). The time that the individual takes to gain employment from unemployment status is relevant to prescription drugs uses; however, the effect is very complicated. On the one hand, the switch from unemployment to employment status might attribute to the consumption of prescription

medicines and the increase of human capital. On the other hand, the longer it takes an individual to get a job, the less the health services the individual might consume because healthcare is a normal good. Besides, people tend to delay health services during the job-seeking process. For the first case, *eg* might have a positive relationship with drug utilization due to *hg*. Since *hg* is also an independent variable here, we expected *eg* should negatively affect on drug utilization as the second case. In contrast, the third case would add complexity to the effect on RX (See Figure 3 The Hypothesized Modelling Diagram).

Since there are multiple endogenous variables and numerous instruments, we need to handle the identification problem. Fortunately, with the instrument variables (*lw*, *linc*, *erg*, and *le*) and two endogenous variables (*eg* and *hg*), we are still able to identify the structural equation's coefficients as the instruments are more than endogenous variables. The problem here will be the test of over-identification to the endogenous variable models, i.e., whether at least one of the instruments is wrong. Besides, we employ the Hansen J statistic for the overidentification test of all instrument variables. The Underidentification test (Kleibergen-Paap rk LM statistic) is precise, along with the test of endogeneity.

DATA AND VARIABLES

Data

The data source is the Medical Expenditure Panel Survey (MEPS), a nationally representative survey with rich individual-level data. A longitudinal data 2008-2016, and a two-year specific period with five rounds in each dataset, 2007-2008, 2009-10, 2010-11, and 2011-12 MEPS data used for analysis in depth. Individual patient characteristics include Employment status, Perceived Health Status, Income, Age, Gender, Race, Health insurance status (Uninsured), and the individual's education years. Observations for most of the variables are over 10,000 for each of the three datasets. Variable labels and definitions are in Table 1.

Table 1. Variable definition

Label	Definition
RXQ (rxq)	A count of all prescribed medications purchased during a specific year.
RXS (rxs)	Sums all amounts paid out-of-pocket and by third party payers for each prescription purchased in a specific year.
hg	If the individual has a change of perceived health status from less healthy level to healthier level in 1 to 4 rounds in the specific two-year period, then, hg equals to 1 to 4, respectfully.
eg	If the individual has a change from unemployment to employment status in 1 to 4 rounds in the specific two-year period, then, *eg* equals to 1 to 4, respectfully.
INC (linc)	Person's non-wage income.
UNINSURED	Uninsured=1, otherwise, 0.
FEMALE	Female=1, Male=0.
AGE (age)	Individual age.
E (*le*)	Individual's education years.
W (*lw*)	Individual's wage.
white	Individual's race, White=1, otherwise 0.
asian	Individual's race, Asian=1, otherwise 0.
hisp	Individual's race, Hispanic=1, otherwise 0.
erg	The regional employment to population ratio.

Note: Lowercase variable label, rxq, rxs, linc, age, lw, and le, are in natural logarithm.

The dependent variables in this study are RXQ and RXS (See Table 1). RXQ (RXTOT in MEPS data) is the natural log of Prescribed Medicines, including Refills in 2007-2016, the utilization quantity variable, which is a count (how many prescripts) of all prescribed medications purchased during a specific year (includes initial purchases and refills). RXS (RXEXP in MEPS data) is the natural log of the spending of Prescribed Medicines, including Refills 2007-2016. The expenditure variable sums all amounts paid out-of-pocket and by third-party payers for each prescription purchased in a specific year.

Why we have the two dependent variables in this study is mainly because of the distinctive feature of prescription drug consumption. Most of the medicines in the market have two names, brand name and generics. Unless those insensitive to money, individuals would have the chance to ask the doctor for the generic version instead of the brand name instead of generic drugs and save the amount of money. Therefore, individuals can adjust their consumption of prescription drugs, with a fixed amount of money consuming more with generics or with a fixed count of prescripts consuming less with the brand name. This feature of the consumption gives the individual consumer more choices but makes the prescription drug utilization more complicated. To fully capture the drug utilization, a simultaneous equation of seemly-unrelated regression with two dependent variables should be an excellent method to deal with this problem.

To capture the changes from unemployment to employment and from the less healthy to better status, we use the MEPS Panel sample with detailed information within five rounds in a specific two-year period. The individuals (with person ID: DUPERSID) on this dataset represent those who were in the MEPS population for all or part of 2007-2008, 2009-2010, 2011-12, 2013-14, and 2015-16 (we also have 2010-11 data estimation), which enable me to estimate the switching of employment and Perceived health status in full five rounds in the two years for the same individual. And the Dwelling unit ID, DUID, allows me to correct the potential heteroskedasticity in estimation by clustering the group effects.

To identify how the employment status can influence drug utilization by the individual, we need to figure out the definition and measure of drug utilization. The factors can drive up drug utilization first. Drug utilization research can involve different aspects of drug use, and drug prescribing, such as pattern, quality, determinants, and outcomes of use, the marketing, distribution, prescription drug uses in society, with particular emphasis on the resulting medical, social and economic consequences (Sjoqvist and Birkett, 2003).

According to different types of drug utilization research, the existing literature highlighted the importance of the demographics structure, insurance, race, and gender effects (Svarstad, *et al.*, 1987; Roe, et al., 2002; Simoni-Wastila, *et al.*, 2004; Sznitman, 2007; Han and Rizzo, 2012; Skoog, *et al.*, 2014) on the prescription drugs utilization. The prevalence of prescription drug use varies significantly by sex, age, and region of the country and varies across race race (Fillenbaum *et al.*, 1993). As for stressing the insurance influences on prescription drug spending, Leibowitz *et al*. (1985) and Thomas *et al*. (2002) suggested that individuals with more generous insurance buy more prescription drugs, and with more aggressive cost-sharing have lower total prescription drug spending.

In general, there is a positive relationship between income and healthcare utilization and expenditures (You and Okunade, 2017; Okunade *et al*., 2018); as the increase of the individuals' income (including wages and other income), people have the propensity to use and spend more on healthcare services, including prescription medications. Topically, females and older people use more prescription drugs than males and age groups for every drug benefiter. Besides, insurance status is related to the quantity of prescriptions individuals purchased. Therefore, the independent variables in this study include individual's demographic characteristics, age in years, gender (Female), non-wage income or discounted property income, wages, insurance status

(Uninsured), Employment status (Unemployed, Employed), and Perceived health status (Poor, Fair, and Better than Good).

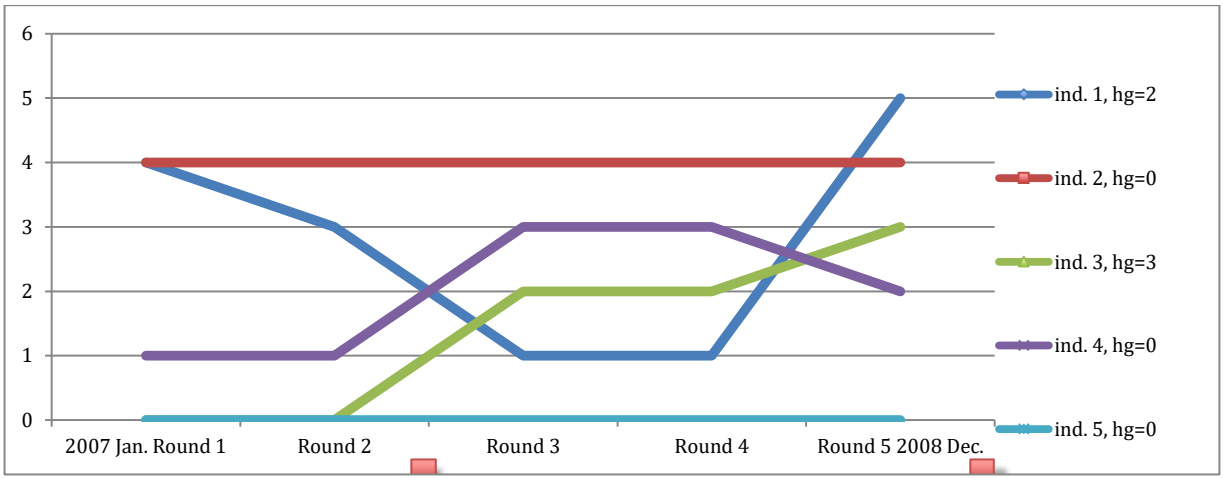
Variables selection and construction

The gaining of health status (ΔHG, or hg)

The two-year longitudinal data with five rounds of Perceived Health Status variables were applied to address the changing and recovery process of an individuals' health status. The variable of changing health status is a process from less healthy (< Good) to healthy (Good, or > Good, such as Very Good and Excellent) status. The Perceived Health Status "Good" is the threshold made to distinguish healthy to unhealthy.

For example, Individual 1 with hg=2, means gaining healthy status (> =Good) in 2 rounds from a less healthy level (< Good). As shown in Graph 1, Individual 1's Perceived Health Status is Very Good (or even better, > Good) at Round 1, falls to Good at Round 2; it does not matter because Good is the threshold made in this study to distinguish healthy to unhealthy. This individual's Perceived Healthy status falls to Fair at Round 3 and 4, then increases to Excellent in Round 5, which means there are two rounds below "Good" before gaining "Good" or better health status; therefore, hg=2. Individual 2 and Individual 5 with hg=0 because their health status with no change though they have entirely different health status, one has been "Excellent" and the other has been "Poor" all the time. Individual 4 also has hg=0; this is because he falls less healthy after gaining a healthy status in 2 rounds. Since the goal in this research is to estimate the changing effects of health status, the hg=0 is not included. The gaining of healthy status will be hg_i, i=1, 2, 3, 4 (See Figure 1).

Figure 1: The gaining health status



Note: 1 to 5 stands for Poor, Fair, Good, Very Good, and Excellent, respectively.

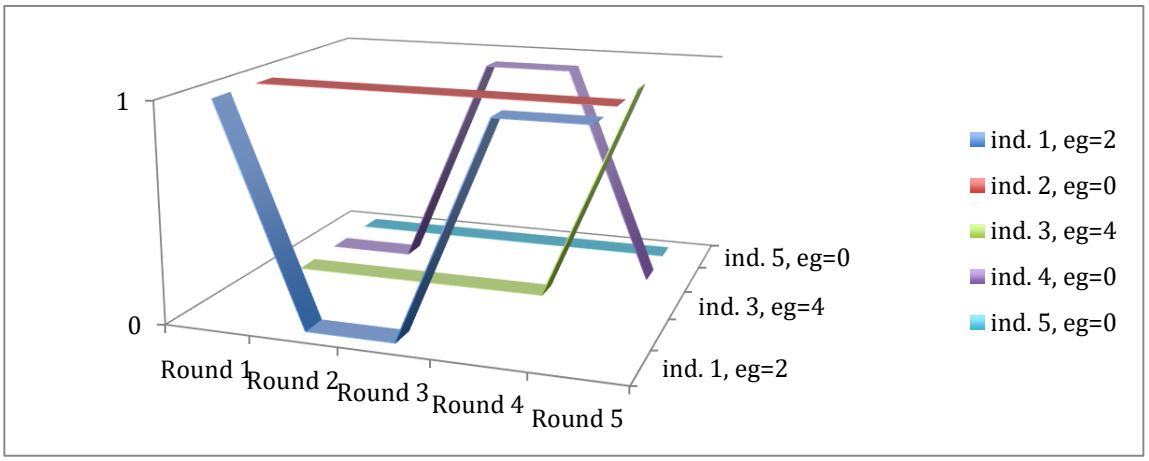
The gaining employment from unemployment (ΔUG, or eg)

The average length of unemployment changed dramatically about 40 weeks with the 2007-2009 recession from 10 weeks typically. Hence, it would be meaningful to detect if the period from unemployment to employment status for an individual would affect his (her) prescription drug utilization. For the two-year MEPS Longitudinal Data, Δ*UG (eg*) is gaining employment from

employment status for working-age individuals age 18-65 in 1 to 4 rounds within the 5-round in the two-year longitudinal period.

Given there are five rounds in each two-year longitudinal data file, the switch of employment status will have four different types, switching in one round, two rounds, three rounds, and four rounds. If the individual only uses one round to gain employment from the unemployment status, then eg=1, if he or she uses two rounds, the eg=2, etc. Again, since no changing on the employment status regardless of being unemployed or employed all the time, eg=0, which is not included in the changing of employment status, hence, eg_i, i=1, 2, 3, 4 (See Figure 2).

Figure 2: The gaining employment status



Note: 1 stands for employed and 0 for unemployed.

Variables with direct or indirect effects on prescription drug utilization

Independent variables in this study include individual's demographic characteristics, age in years, gender (Female), non-wage income, wages, insurance status (Uninsured), except the Employment and Perceived health status variables. Thornton and Rice (2011) suggests that insurance has a direct effect on healthcare spending, income, education, and age. It has an indirect impact because these variables are significant determinants of health status, which is an essential factor affecting healthcare spending. Meanwhile, the socio-economic variables, income, and education can have direct and indirect effects (via health status).

In general, the insurance coverage and availability can directly affect healthcare spending by affecting the individual patient's demand for healthcare services and influencing the cost of providing healthcare services to the supply side. The focus in this study is the demand for prescription medicine rather than the supply side given the dataset is the individual level. The insurance factor included here is an individual's Uninsured status to capture the insurance effects on prescription drug utilization, especially targeting the effect of the insurance beneficiary associated with the 2010 Affordable Care Act. Regarding the potential occurrence of adverse selection problems, some healthy people may choose *UNINSURED* under ACA insurance beneficiary because they have good health status. In contrast, the less healthy people may switch to insured (having any insurance), then the effect of *UNINSURED* on drug utilization will be more significant than some other times.

Typically, aging people use more prescription drugs than other age groups for every drug benefiter (Long 1994). Besides, female influences the prescription drug more related to the

number of prescriptions purchased than the spending. Hence, female (*FEMALE*) should have been an exogenous variable with a direct effect on the prescription drug utilization from the individual perspective based on the individual-level data. Age (*age*), on the other hand, according to the theoretical model I derived, age (*age*) may have both direct and indirect effects (age depreciation effects, δ_i, on the changing of health status) on prescription drug utilization. Since this research focuses on the changing of employment effects on prescription drug utilization, I restrict the individuals aging from 18 to 65, the number of working-age in the US labor market.

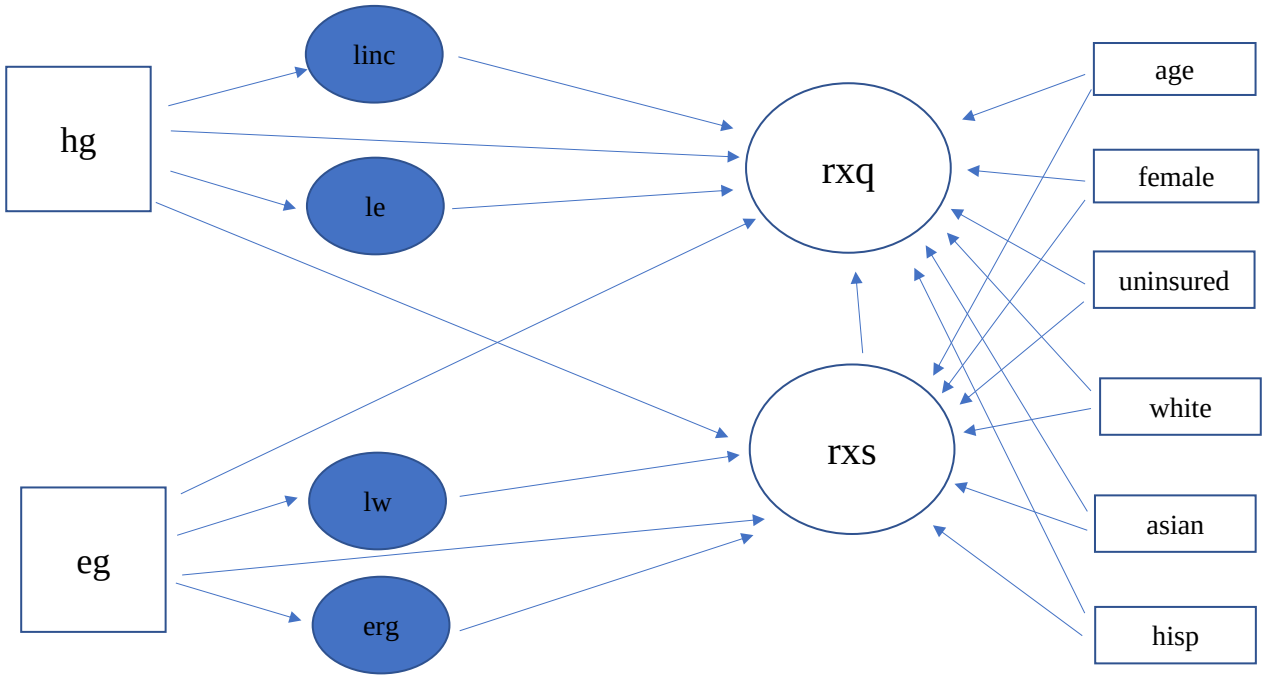
Previous research has mostly stressed the positive relationship between income (total income) and healthcare spending while ignoring the potential indirect effect of income on healthcare spending through the individual patient's employment or health status. Income has a positive relationship with employment status, mainly due to the wage effect. To correct the high correlation between the total income the wage, we separate the individual's total income into the non-wage income (*linc*) and the wage earnings (*lw*). As a result of this variable selection, *linc* becomes an exogenous variable with an indirect effect on prescription drugs. At the same time, *lw* may have direct and indirect effects on drug utilization, which is precisely consistent with the theoretical Grossman framework we derived.

It is well known that years of education are the most critical correlation factors of health, robust to different health health (Grossman, 2000). Rosen and Taubman (1982) and Berger and Leigh (1989) suggests that education directly affects individuals' health status by increasing the efficiency that individuals produce health. Thus, in this study, we use years of education as the instrument variable to measure the indirect effect of the gaining of healthy status on prescription drug utilization. The regional employment-population ratio, erg, may directly influence the individual's attitude towards actively seeking employment, while may not directly affect the individual's drug utilization, but through the changing process of gaining employment somehow. Therefore, the *erg* is used as an instrumental variable to *eg*. We also include some demographic characteristics, except *age* and *female* we have discussed, there is a group of the race variables, *white* (White), *asian* (Asian), and *hisp* (Hispanics).

Table 2. Descriptive statistics for 2007-2016

Variable	Obs.	Mean	SD	Min.	Max.
RXQ	41855	2.207	1.188	5.866	0.405
RXS	55467	5.052	2.275	12.821	0.405
HG	83666	0.508	0.764	4.000	0.000
EG	83666	0.147	0.580	4.000	0.000
AGE	81256	3.260	0.943	4.443	0.000
FEMALE	83666	0.522	0.500	1.000	0.000
UNINSURED	83666	0.124	0.330	1.000	0.000
LE	75579	1.832	0.882	2.833	0.000
LW	44842	9.880	1.168	12.556	1.504
LINC	32804	7.961	2.326	12.386	-0.693
WHITE	83666	0.613	0.487	1.000	0.000
ASIAN	83666	0.073	0.260	1.000	0.000
HISP	83666	0.317	0.465	1.000	0.000
ERG	83666	0.584	0.082	0.638	0.000

Figure 3: The Hypothesized Modelling Diagram



EMPIRICAL RESULTS

Using Medical Expenditure Panel Survey (MEPS) data for 2007-2016, and specifically check for 2007-08, 2009-10, 2010-11, and 2011-12, we examine how the gaining employment and health status affect the prescription drug utilization (spending and quantity) for working-age individuals (18-65 years old).

Results from OLS estimations

Beginning with the OLS estimation to capture the effectiveness of the variables selected in this study, Table 3 and Table 4 present the results of OLS estimation for the prescription drug utilization and spending equation, respectively, for the 2007-2016 period; Table 5 and Table 6 present the results of OLS estimation for the prescription drug utilization and spending equation, respectively, for 2007-08, 2009-10, 2010-11, and 2011-12 four separate periods. While most parameter estimates have expected signs, apparent magnitudes, and reasonable significance levels, the term variance, R^2 statistics range, is 0.14 -0.72, indicating that the models do not fit the dataset very well. The correlation test shows no high correlation among these variables. The joint effects F-tests (with P-value) of these seven variables are significant for the total dataset, 2007-2016, and separate four datasets as well, indicating that they jointly explain the variation in the drug utilization. Hence, we can use these variables for the simultaneous equation estimations.

Table 3: OLS Prescription Drug Utilization (rxq) estimation results (t-statistics in parentheses)

	2007-2016		
Variables	Coef.	t	p
rxs	0.454963	178.0323	0.0000
hg	0.001534	0.259645	0.7951
eg	-0.004904	-0.799938	0.4238
age	0.541477	45.3719	0.0000
female	0.079445	9.198272	0.0000
UNINSURED	0.008612	0.565233	0.5719
lw	-0.077286	-20.67367	0.0000
white	-0.015998	-1.674077	0.0941
asian	-0.160569	-8.475633	0.0000
hisp	-0.035107	-3.369512	0.0008
_cons	-1.800636	-32.69811	0.0000
R-squared	0.659798		
Adjusted R-squared	0.659653		

Table 4: OLS Prescription Drug Utilization (RXS) estimation results (t-statistics in parentheses)

	2007-2016		
Variables	Coef.	t	p
hg	-0.025515	-1.596439	0.1104
eg	-0.045249	-2.803192	0.0051
age	2.024408	67.62211	0.0000
female	0.267644	11.63113	0.0000
UNINSURED	-0.890245	-23.68299	0.0000
lw	-0.182675	-17.85705	0.0000
white	0.285762	11.13844	0.0000
asian	-0.311537	-6.303835	0.0000
hisp	-0.575594	-21.1289	0.0000
_cons	-0.719338	-5.08957	0.0000
R-squared	0.178544		
Adjusted R-squared	0.178299		

Table 5: OLS Prescription Drug Utilization (rxq) estimation results (t-statistics in parentheses)

	2007-08		2009-10		2010-11		2011-12	
Variables	Coef.	t	Coef.	t	Coef.	t	Coef.	t
rxs	0.534	[79.55]	0.492	[84.97]	0.461	[67.93]	0.434	[73.58]
hg	0.003	[0.24]	0.005	[0.35]	-0.001	[-0.06]	-0.002	[-0.15]
eg	-0.039	[-2.34]	-0.007	[-0.49]	-0.029	[-1.84]	-0.03	[-2.08]
age	0.378	[10.84]	0.487	[16.1]	0.562	[15.69]	0.569	[17.85]
female	0.095	[4.66]	0.09	[4.64]	0.127	[5.76]	0.077	[3.87]
uninsured	-0.024	[-0.72]	0.03	[0.99]	0.031	[0.88]	-0.077	[-2.23]
lw	-0.089	[-8.15]	-0.096	[-9.56]	-0.097	[-9.14]	-0.1	[-9.53]
white	-0.055	[-0.88]	-0.167	[-2.9]	-0.088	[-1.23]	-0.119	[-2.13]
asian	-0.146	[-1.9]	-0.376	[-5.32]	-0.277	[-3.35]	-0.255	[-3.66]
hisp	0.023	[0.77]	-0.056	[-2.08]	-0.059	[-1.83]	0.004	[0.14]
_cons	-1.563	[-10.54]	-1.453	[-10.23]	-1.606	[-10.64]	-1.415	[-9.66]
R^2	0.717		0.678		0.653		0.62	

Note: Std. Err. adjusted for clusters in DUID.

Table 6: OLS Prescription Drug Utilization (RXS) estimation results (t-statistics in parentheses)

	2007-08		2009-10		2010-11		2011-12	
Variables	Coef.	t	Coef.	t	Coef.	t	Coef.	t
hg	-0.033	[-0.88]	-0.102	[-2.89]	-0.038	[-0.97]	0.02	[0.57]
eg	-0.093	[-2.38]	-0.063	[-1.69]	-0.041	[-1.06]	0.022	[0.61]
age	2.229	[27.05]	1.902	[23.98]	1.979	[22.21]	1.939	[24.48]
female	0.27	[4.8]	0.314	[5.95]	0.275	[4.49]	0.345	[6.52]
uninsured	-0.667	[-8.21]	-0.816	[-11.11]	-0.758	[-9.1]	-0.906	[-11.74]
lw	-0.245	[-8.14]	-0.198	[-7.19]	-0.132	[-4.14]	-0.1	[0]
white	0.031	[0.17]	0.131	[0.84]	0.162	[0.93]	-0.017	[-0.09]
asian	-0.526	[-2.44]	-0.778	[-4.08]	-0.355	[-1.74]	-0.577	[-2.71]
hisp	-0.72	[-8.53]	-0.721	[-9.98]	-0.626	[-7.58]	-0.639	[-9.62]
_cons	-0.442	[-1.13]	0.086	[0.23]	-0.987	[-2.34]	-1.092	[-2.74]
R^2	0.184		0.157		0.14		0.147	

Note: Std. Err. adjusted for clusters in DUID.

Since OLS estimations are likely to be bias and inconsistent because of the endogeneity, 2SLS estimations are employed to handle this problem. Using two successive applications of the OLS estimation, the 2SLS approach is a particular type of IV estimation and the default method for regressing over-identified models. After running an IV (using *le* to instrument *hg*, and *linc* and *erg* to instrument *eg*) regression, we check the weak Instrument Diagnostics, under-identification (Kleibergen-Paap rk LM statistic), and Over-identifying (Hansen J statistic overidentification) test of all instruments. Using the Endogenous test, we check the hypothesis that the endogenous variables are exogenous based on orthogonality conditions. Finally, we use the Hausman command to perform the Hausman Specification test to compare the IV regression performances to the OLS regression.

Results from 2SLS estimations

To assess the pathways through which the changing of employment and health factors affect prescription drug utilization, we estimate the structural utilization and spending equations and the employment and health equation separately using the 2SLS approach. Table 7 presents the Structural Prescription Drug Utilization (*rxq*) estimation results (2SLS), and Table 8 shows the results from the Structural Prescription Spending (*rxs*) estimations (2SLS) for the total dataset, 2007-2016. And Table 9 presents the Structural Prescription Drug Utilization (*rxq*) estimation results (2SLS), and Table 10 shows the results from the Structural Prescription Spending (*rxs*) estimations (2SLS) for separate datasets, 2007-08, 2009-10, 2010-11, and 2011-12. The first stage estimations show the direct effect of wages (*lw*), income (*linc*), age (*age*), regional employment-population ratio (*erg*), and education on the endogenous variables, the gaining of employment (wages and non-wage income effects), and health status (education, age, and wages effects). And the 2SLS estimations suggest the indirect effects of the instrumental variables, *le*, *lw*, and *erg,* on the drug utilization and spending and the direct effects of other independent variables.

For the total data 2007-2016, according to the result in Table 7, the utilization of prescription drugs has a negative relationship with the changing of health status from bad to good, and a positive relationship with the changes of employment status from unemployment to employment. The total effects of the changing health status on drug utilization (*rxq*) are significant

with a value, -0.918. The overall effects of the gaining employment status on drug utilization are positive but insignificant with a value, 0.202. Results in Table 8 shows that there are negative effects (statistically significant) of both gaining healthy status and the changing employment status (*eg*) on prescription drug spending (*rxs*) for all the period, 2007-2016, with values -10.143 and -22.045, respectively.

Table 7: Structural Prescription Drug Utilization (rxq) estimation results (t-stat in parentheses)

	2007-2016	
	First-stage regressions	
hg	Coef.	p-value
rxs	-0.007	0.087
age	-0.043	0.082
female	-0.024	0.084
uninsured	0.089	0.003
lw	-0.012	0.070
white	0.018	0.278
asian	0.022	0.446
hisp	0.067	0.000
linc	0.001	0.843
le	0.069	0.000
erg	0.199	0.409
_cons	0.499	0.005
	F value	P-value
Excluded instruments F test	12.16	0.000
	First-stage regressions	
eg	Coef.	p-value
rxs	-0.010	0.020
age	-0.263	0.000
female	0.008	0.569
uninsured	0.051	0.143
lw	-0.155	0.000
white	0.031	0.077
asian	-0.005	0.841
hisp	-0.011	0.601
linc	0.014	0.000
le	-0.028	0.048
erg	0.086	0.053
_cons	2.737	0.000
	F value	P-value
Excluded instruments F test	12.820	0.000

	Instrumental variables (2SLS) regression	
rxq	Coef.	p-value
hg	-0.918	0.000
eg	0.202	0.445
rxs	0.462	0.000
age	0.529	0.000
female	0.045	0.016
uninsured	0.054	0.270
lw	-0.066	0.141
white	-0.014	0.568
asian	-0.181	0.000
hisp	-0.037	0.190
_cons	-1.51	0.073
Test of joint significance of endogenous Statistics	Chi2 (1)	P-value
Anderson-Rubin Wald test	F value: 14.81	0.000
Anderson-Rubin Wald test	44.47	0.000
Stock-Wright LM S stat.	43.16	0.000
Underidentification test (Kleibergen-Paap rk LM statistic)	23.107	0.000
Hansen J statistic (overidentification test of all instruments)	2.761	0.090
Hausman test	40.41	0.000

Note: OLS regression with robust standard errors and clusters (DUID)

Table 8: Structural Prescription Drug Spending (rxs) estimation results (t-stat in parentheses)

	2007-2016	
	First-stage regressions	
hg	Coef.	p-value
age	-0.022	0.319
female	-0.016	0.186
uninsured	0.067	0.009
lw	-0.014	0.021
white	0.008	0.583
asian	-0.008	0.749
hisp	0.061	0.001
linc	0.001	0.809
le	0.059	0.000
erg	0.351	0.056
_cons	0.347	0.012
	F value	P-value

Excluded instruments F test	9.62	0.000
	First-stage regressions	
eg	Coef.	p-value
age	-0.261	0.000
female	0.014	0.248
uninsured	0.036	0.219
lw	-0.160	0.000
white	0.018	0.257
asian	-0.006	0.794
hisp	0.004	0.808
linc	0.012	0.000
le	-0.027	0.025
erg	0.203	0.040
_cons	2.717	0.000
	F value	P-value
Excluded instruments F test	10.740	0.000
	Instrumental variables (2SLS) regression	
rxs	Coef.	p-value
hg	-10.143	0.021
eg	-22.045	0.000
age	-3.726	0.001
female	0.361	0.227
uninsured	0.654	0.407
lw	-3.977	0.000
white	0.610	0.129
asian	-0.691	0.236
hisp	-0.120	0.805
_cons	-69.116	0.000
Test of joint significance of endogenous Statistics	Chi2 (1)	P-value
Anderson-Rubin Wald test	F value: 16.02	0.000
Anderson-Rubin Wald test	48.1	0.000
Stock-Wright LM S stat.	4238.66	0.000
Underidentification test (Kleibergen-Paap rk LM statistic)	28.522	0.000
Hansen J statistic (overidentification test of all instruments)	151.168	0.000
Hausman test	27.30	0.001

Note: OLS regression with robust standard errors and clusters (DUID)

Table 9: Structural Prescription Drug Utilization (rxq) estimation results (t-stat in parentheses)

	2007-08		2009-10		2010-11		2011-12	
	First-stage regressions							
hg	Coef.	t	Coef.	t	Coef.	t	Coef.	t
rxs	-0.023	[-2.07]	-0.026	[-2.81]	-0.016	[-1.53]	-0.01	[-1.03]
age	0.004	[0.06]	-0.057	[-1.02]	-0.083	[-1.35]	-0.057	[-1.05]
female	0.004	[0.11]	0.007	[0.25]	0.01	[0.28]	-0.055	[-1.77]
uninsured	-0.039	[-0.6]	-0.025	[-0.47]	0.107	[1.84]	-0.023	[-0.34]
lw	-0.022	[-1.13]	-0.026	[-1.7]	0.008	[0.45]	-0.003	[-0.17]
white	0.08	[1.68]	0.016	[0.32]	0.035	[0.74]	-0.015	[-0.35]
asian	0.235	[2.6]	-0.09	[-1.32]	-0.015	[-0.19]	-0.05	[-0.68]
hisp	-0.105	[-1.78]	0.007	[0.14]	-0.04	[-0.71]	0.091	[1.78]
linc	0.004	[0.55]	0.007	[1.04]	0	[0.03]	0.01	[1.47]
le	-0.194	[-2.27]	-0.122	[-1.36]	-0.112	[-1.22]	0.038	[0.45]
erg	1.051	[5.23]	0.938	[0.11]	-1.72	[-1.29]	0.514	[2.03]
_cons	0.612	[1.89]	0.798	[2.86]	2.12	[3.73]	0.39	[1.2]
Test Statistics	F	P-value	F	P-value	F	P-value	F	P-value
Excluded instruments F test	7.57	0.000	2.85	0.036	5.88	0.000	2.43	0.06
Angrist-Pischke multivariate excluded F test	7.96	0.000	3.36	0.035	4.06	0.007	19.5	0.000
eg	Coef.	t	Coef.	t	Coef.	t	Coef.	t
rxs	-0.027	[-3.03]	-0.011	[-1.14]	-0.01	[-0.95]	-0.009	[-1.01]
age	-0.135	[-2.01]	-0.196	[-3.32]	-0.262	[-3.64]	-0.347	[-5.67]
female	0.058	[2.34]	-0.032	[-1.17]	-0.049	[-1.44]	-0.014	[-0.45]
uninsured	0.08	[1.2]	0.145	[2.31]	0.052	[0.73]	0.108	[1.51]
lw	-0.14	[-5.58]	-0.135	[-6.25]	-0.161	[-6.54]	-0.178	[-7.48]
white	0.006	[0.14]	0.052	[1.32]	-0.032	[-0.65]	-0.051	[-1.11]
asian	0.02	[0.3]	0.007	[0.12]	-0.021	[-0.27]	0.039	[0.53]
hisp	0.066	[1.08]	-0.059	[-1.47]	0.055	[0.94]	-0.045	[-0.94]
linc	0.012	[2.3]	0.018	[3.08]	0.019	[2.77]	0.016	[3.31]
le	0.089	[1.05]	0.069	[0.97]	0.041	[0.33]	0.004	[0.04]
erg	0.299	[1.78]	0.364	[2.59]	0.477	[2.6]	0.431	[1.97]
_cons	1.731	[4.92]	1.831	[5.68]	2.444	[5.33]	3.075	[7.73]
Test Statistics	F	P-value	F	P-value	F	P-value	F	P-value
Excluded instruments F test	5.99	0.000	11.86	0.000	2.71	0.029	4.85	0.002
Angrist-Pischke multivariate excluded F test	7.91	0.000	15.47	0.000	2.67	0.046	6	0.0005
	Instrumental variables (2SLS) regression							
rxq	Coef.	t	Coef.	t	Coef.	t	Coef.	t
hg	0.779	[2.64]	0.775	[1.87]	0.281	[0.82]	-2.191	[-0.63]
eg	0.228	[0.45]	-0.122	[-0.28]	0.224	[0.6]	1.309	[0.49]
rxs	0.575	[30.09]	0.525	[38.94]	0.476	[37.47]	0.435	[14.88]
age	0.364	[3.97]	0.458	[5.25]	0.66	[5.95]	0.86	[1.21]
female	0.102	[2.06]	0.089	[2.34]	0.155	[4.1]	-0.046	[-0.26]

uninsured	-0.052	[-0.58]	0.144	[1.6]	-0.064	[-0.79]	-0.22	[-0.52]
lw	-0.031	[-0.45]	-0.081	[-1.41]	-0.06	[-0.91]	0.122	[0.25]
white	-0.141	[-2.58]	-0.066	[-1.28]	-0.024	[-0.5]	-0.042	[-0.28]
asian	-2.269	[-2.32]	-0.172	[-1.85]	-0.167	[-2.14]	-0.339	[-1.04]
hisp	0.118	[1.66]	-0.032	[-0.5]	-0.006	[-0.11]	0.31	[0.72]
_cons	-2.732	[-2.47]	-2.163	[-2.42]	-2.762	[-2.22]	-3.992	[-0.61]
Test Statistics	Chi2 (1)	P-value	Chi2 (1)	P-value	Chi2 (1)	P-value	Chi2 (1)	P-value
Test of endogeneity	77.832	0.000	19.026	0.000	111.027	0.000	210.921	0.000
Underidentification test (Kleibergen-Paap rk LM statistic)	19.801	0.000	6.995	0.030	7.133	0.068	45.263	0.000
Hansen J Overidentifying restrictions	32.419	0.000	0.221	0.639	25.192	0.000	1147.72	0.000

Note: OLS regression with robust standard errors and clusters (DUID)

Table 10: Structural Prescription Drug Spending (rxs) estimation results (t-statistics in parentheses)

	2007-08		2009-10		2010-11		2011-12	
	First-stage regressions							
hg	Coef.	t	Coef.	t	Coef.	t	Coef.	t
age	-0.037	[-0.71]	-0.05	[-1.05]	-0.06	[-1.12]	-0.041	[-0.88]
female	0.019	[0.65]	-0.003	[-0.1]	0.002	[0.06]	-0.051	[-1.81]
uninsured	-0.016	[-0.3]	0.039	[0.82]	0.03	[0.63]	0.022	[0.38]
lw	-0.023	[-1.31]	-0.011	[-0.79]	0.003	[0.17]	-0.015	[-0.98]
white	0.061	[1.41]	-0.02	[-0.52]	0.018	[0.42]	0.002	[0.06]
asian	0.156	[2.09]	-0.038	[-0.6]	-0.036	[-0.5]	0.006	[0.08]
hisp	-0.064	[-1.25]	0.04	[0.91]	-0.042	[-0.83]	0.085	[1.89]
linc	0.005	[0.82]	0.003	[0.44]	-0.002	[-0.24]	0.008	[1.37]
le	-0.138	[-1.85]	-0.149	[-1.91]	-0.109	[-1.34]	-0.012	[-0.15]
erg	0.945	[5.6]	0.877	[0.96]	-1.811	[-1.41]	0.364	[1.27]
_cons	0.545	[1.94]	0.624	[2.45]	2.066	[3.81]	0.589	[1.86]
Test Statistics	F	P-value	F	P-value	F	P-value	F	P-value
Excluded instruments F test	7.57	0.000	2.86	0.036	5.86	0.000	1.24	0.294
Angrist-Pischke multivariate excluded F test	7.96	0.000	3.38	0.034	4.03	0.007	19.5	0.000
eg	Coef.	t	Coef.	t	Coef.	t	Coef.	t
age	-0.188	[-3.37]	-0.182	[-3.64]	-0.278	[-4.52]	-0.371	[-6.82]
female	0.05	[2.09]	-0.04	[-1.55]	-0.031	[-1.01]	-0.026	[-0.95]
uninsured	0.083	[1.43]	0.174	[3.11]	0.036	[0.58]	0.082	[1.29]
lw	-0.157	[-6.58]	-0.159	[-7.7]	-0.158	[-6.95]	-0.185	[-8.38]
white	-0.003	[-0.08]	0.047	[1.27]	-0.085	[-1.81]	-0.052	[-1.26]
asian	0.042	[0.64]	0.052	[0.91]	-0.057	[-0.86]	0.018	[0.29]
hisp	0.087	[1.65]	-0.062	[-1.63]	0.033	[0.63]	-0.028	[-0.65]
linc	0.013	[2.61]	0.018	[3.17]	0.02	[3.41]	0.015	[3.24]
le	0.134	[1.73]	0.132	[2.09]	0.021	[0.2]	-0.015	[-0.2]
erg	0.374	[2.27]	0.355	[2.54]	0.477	[2.55]	0.319	[2.36]

_cons	1.786	[5.7]	1.813	[6.23]	2.501	[5.95]	3.318	[9.13]
Test Statistics	F	P-value	F	P-value	F	P-value	F	P-value
Excluded instruments F test	5.99	0.000	11.85	0.000	2.71	0.029	4.12	0.006
Angrist-Pischke multivariate excluded F test	7.91	0.000	15.5	0.000	2.67	0.046	6	0.0005
	Instrumental variables (2SLS) regression							
rxs	Coef.	t	Coef.	t	Coef.	t	Coef.	t
hg	-0.133	[-0.16]	-0.424	[-0.63]	0.199	[0.24]	3.554	[0.49]
eg	1.207	[1.27]	1.377	[1.75]	0.709	[0.74]	0.608	[0.13]
age	2.431	[11.27]	2.145	[11.57]	2.299	[7.94]	2.373	[1.59]
female	0.123	[1.27]	0.329	[3.84]	0.258	[2.7]	0.512	[1.84]
uninsured	-0.705	[-4.19]	-1.07	[-5.1]	-0.877	[-5.8]	-1.037	[-2.9]
lw	-0.094	[-0.63]	0.035	[0.27]	-0.055	[-0.33]	0.029	[0.04]
white	0.225	[1.72]	0.204	[1.69]	0.308	[2.18]	0.449	[1.43]
asian	-0.405	[-1.7]	-0.715	[-3.5]	-0.337	[-1.66]	-0.013	[-0.04]
hisp	-0.641	[-4.08]	-0.35	[-2.32]	-0.55	[-4.21]	-0.705	[-0.92]
_cons	-3.034	[-1.29]	-3.394	[-1.69]	-3.306	[-1.04]	-6.415	[-0.54]
Test Statistics	Chi2 (1)	P-value	Chi2 (1)	P-value	Chi2 (1)	P-value	Chi2 (1)	P-value
Test of endogeneity	69.5	0.000	9.357	0.009	48.21	0.000	210.921	0.000
Underidentification test (Kleibergen-Paap rk LM statistic)	19.801	0.000	7.039	0.030	7.112	0.068	45.263	0.000
Hansen J Overidentifying restrictions	33.738	0.000	0.009	0.925	36.185	0.000	1147.72	0.000

Note: OLS regression with robust standard errors and clusters (DUID)

According to Table 9 results, prescription drugs' utilization has a positive relationship with the changing of health status and the changes in employment status from unemployment to employment most of the time. The total effects of the changing health status on drug utilization (*rxq*) are significant with values, 0.779, 0.775, 0.281, and -2.191 (insignificant), for 2007-08, 2009-10, 2010-11, and 2011-12, respectfully. The total effects of the changing health status on drug utilization (*rxq*) are significant with values, 0.779, 0.775, 0.281, and -2.191 (insignificant), for 2007-08, 2009-10, 2010-11, and 2011-12, respectfully. The total effects of the gaining employment status on drug utilization are positive but insignificant at 0.228, -0.122, 0.224, and 1.309, for 2007-08, 2009-10, 2010-11, and 2011-12, respectfully. Results in Table 10 show that the effects of gaining healthy status on prescription drug spending are insignificant for all four periods. The total effects of the changing employment status (*eg*) are also insignificant on *rxs*, only significant with 1.377 for 2009-10. The significant effect of *eg* in 2009-10 suggests that the longer time the individual to get employment, the more he/ she spends on drugs during the launch of the Affordable Care Act (ACA).

There are four instrument variables, education (for *hg*), *erg* (for *eg*), *linc* (for *eg*) and lw (for both *hg* and *eg*) for both drug utilization and expenditure estimations. For the total data, 2007-2016, education (*le*) has a significantly positive effect, 0.059, on gaining healthy status; wages (*lw*) has both direct and indirect effects on drug utilization, with negative effects on gaining employment status, (*eg*), -0.16, statistically significant. The non-wage income has significantly

positive effects on *eg*, 0.012, which make sense that people may spend a long time finding a high pay job, while with a higher non-wage income.

For the separate datasets, education (*le*) has a negative effect (*le,* -0.138, -0.149, -0.109, and -0.012, for 2007-08, 2009-10, 2010-11, and 2011-12, respectively) on gaining healthy status, statistically significant for three out of four estimations. While wages (*lw*) has negative effects on gaining employment status, (*eg*), -0.157, -0.159, -0.158, and -0.185, for 2007-08, 2009-10, 2010-11, and 2011-12, respectively. And the non-wage income has positive effects on *eg* (0.013, 0.018, 0.02, and 0.015, for 2007-08, 2009-10, 2010-11, and 2011-12, respectively), again, which makes sense people may spend a long time to find a high pay job. The higher the wage, the faster the individual wants to get a job; people tend to be more sensitive to wages.

The regional employment-population ratio (*erg*) has a positive relationship with the gaining of employment status (*eg*) for all datasets. This result suggests that, with a higher *erg*, individuals in the area tend to have a relatively longer time gaining employment. Why? On the one hand, the higher erg, the hard to find a job; on the other hand, the higher the *erg*, the more confidence the individuals will have on finding a high-pay job (with longer seeking time).

Age (*age*) has a negative relationship with the gaining of health status and gaining employment, except for 2007-2008 (positive but insignificant). As age increases, people tend to have less recovery time from less healthy to healthy status and less time on gaining employment.

Uninsured status (*UNINSURED*), a critical variable, has insignificant effects on both drug utilization (*rxq*) and expenditures (*rxs*) for the total period of 2007-2016. However, when checking the separate data periods, *UNINSURED* has significant effects on drug spending (*rxs*) with values of -0.705, -1.07, -0.877, and -1.037 2007-08, 2009-10, 2010-11, and 2011-12, respectively. The result shows that there are apparent differences in drug spending between the uninsured and insured during the launching of the ACA period comparing with those from the usual period. This difference might be the evidence of how the aiming (affordability of health insurance, lowering the uninsured rate) of the Affordable Care Act works in the sense of adverse selection problem. The significantly negative effect on the higher coefficient value of UNINSURED on both drug utilization and spending for 2010-2011 created a significant gap between uninsured and insured individuals, which is a sign of the ACA insurance benefit.

Female (*FEMALE*) has a positive effect on both drug utilization (significant) and spending (insignificant), statistically significant for four periods most of the time for the separate datasets, suggesting that women tend to use more drugs while may not spend more money. Among the race variables, *hisp* (Hispanic) always has significantly negative effects on drug utilization and spending, while white has significantly positive effects on drug spending.

Along with the 2SLS approach, we performed the excluded instruments test, endogeneity test, under-identification and overidentifying test, and Hausman test. Those tests suggest good and reasonable estimations for drug spending (rxs) and quantity (rxq) models. Tables 7&8 show the test results from the total data, 2007-2016, while Tables 9&10 are the test results from the separate datasets for four two-year periods. Test results indicate the validity of the three instrument variables for the endogenous variable, the gaining of employment and health status (*eg* and *hg*). Endogeneity tests' results reject the exogenous null hypotheses of *eg* and *hg*, while Kleibergen-Paap rk LM statistics indicate no under-identification at a 10% level. And the Hansen J statistics reject the presence of overidentification of all instruments (3 IV: *lw, erg*, and *le*).

As discussed theoretically in this research, *age* and *wage* have direct and indirect effects on drug utilization; however, 2SLS estimations involve the duplicates problem if we regress the *age* and *wage* as both instrument variables to *hg* and an explanatory variable for drug utilization as the same time. Thus, we need the 3SLS estimation to get more accurate and valid empirical results.

Results from 3SLS estimations

To handle the duplicates problem in 2SLS, we apply the Three-Stage Least Squares (3SLS) approach with equations (7, 8, 9, and 10) to estimate the structural coefficients on direct and indirect (employment and health) effects on the prescription drug utilization. Specifying a structural model for each endogenous repressor, the 3SLS estimator uses the cross-correlation of errors to produce a more precise and efficient estimation than 2SLS.

Table 11 presents the results of 3SLS estimations for the structural prescription drug investment equation (1) and (2), and health and employment production function equation (3) and (4) for the total data 2007-2016. The coefficients of the health changing status variable (*hg*) in both the structural prescription drug utilization (*rxq*) and spending (*rxs*) equations are positive and statistically significant, 2.811 and 2.759, respectively. The change of employment status has positive effects on both prescription drug utilization (significant) and spending as well, while an insignificant on spending equation. Thus, results from 3SLS suggest that the gaining of employment status plays an essential role in increasing drug utilization. People who used more time to gain employment tend to use more drugs while they may not spend more, which is a sign of delayed health services because of losing jobs.

Table 12 presents 3SLS estimations for the prescription drug with the separate datasets, 2007-08, 2009-10, 2010-11, and 2011-12. The effects of both health changing status variable (*hg*) and the changes in employment status (*eg*) on drug utilization and spending are quite different from those for the total period estimation. The coefficients of the health changing status variable (*hg*) in the structural prescription drug utilization (*rxq*) equation are positive but not significant, while negative, most of the time on spending but insignificant, neither. The insignificant effects on drug utilization, along with the insignificantly negative effects on spending suggest, the longer the recovery time of the health status, individuals may tend to use less expensive medicines (generic version) and reduce the drug spending. In contrast, the count of the prescription drug does not change. The employment status changes have statistically significant negative effects on prescription drug utilization and positive effects but insignificant on spending equation for three estimations, 2007-08, 2010-11, and 2011-12. However, during 2009-10, on the contrary, the gaining of employment, *eg*, has significant positive effects on drug spending. These results suggest that during the great recession and recovery period, people spent more time gaining employment while they might not use more prescription drugs or spend more on it. As the launch time of the Affordable Care Act, individuals behaved differently; the longer is gaining an employment status, individuals tend to spend more on prescription drugs.

For the total data 2007-2016 (See Table 11), the uninsured effects on both drug utilization (-1.401) and spending (-1.077) are negative and significant as expected. For the separate datasets (See Table 12), the uninsured effects on drug spending (-0.455, -0.855, -0.659, and -0.833) are also negative and significant as expected (insignificant) on the drug utilization. The evidence of change about insurance effects associated with the Affordable Care Act may influence the prescription drug spending somehow but no apparent effects on the count of the drugs individuals purchased (adverse selection). Under the immediate implementation of ACA, the gap of drug spending and quantity between the insured individuals and uninsured became smaller than those from the total period, 2007-2016. The healthy individuals may still choose to be uninsured under the ACA because the ACA placed no penalty on acquiring insurance until 2014, even though the uninsured status would discount the impact of the individual mandate tax penalty that would compel individuals to purchase care. Some individuals did enjoy the health benefit under the ACA policy as the launch period with significant and more considerable effects for 2010-2011 on both drug utilization (Coefficient: -4.089, t= -4.98) and spending. However, the effects will decay on the value of the coefficient afterward, and the significance of the ACA effects is not that strong (2011-12 for drug spending, insignificant for utilization).

Table 11: Structural Prescription Drug Utilization equation and health production function estimation results (t-statistics in parentheses)

Three-stage least-squares regression		
	2007-2016	
	Coef.	p-value
rxq		
rxs	0.881	0.000
hg	2.811	0.008
eg	1.785	0.006
age	3.348	0.000
female	0.103	0.011
uninsured	-1.401	0.000
lw	-0.014	0.866
white	0.116	0.043
asian	-0.503	0.000
hisp	-0.714	0.000
_cons	-6.761	0.001
chi 2 (P-value)	168.050	0.000
rxs		
hg	2.759	0.000
eg	1.169	0.127
age	2.087	0.000
female	0.042	0.449
uninsured	-1.077	0.000
lw	0.038	0.769
white	0.097	0.163
asian	-0.239	0.038
hisp	-0.556	0.000
_cons	-3.879	0.111
chi 2 (P-value)	486.020	0.000
Heteroscedasticity Tests	Chi2	P-value
Breusch-Pagan LM Test	8356	0.000
LR Test	16500	0.000
Wald Test	7600	0.000
Hansen-Sargan overidentification statistic	0.797	0.372
Hausman test	88.11	0.000

Table 12: Structural Prescription Drug Utilization equation and health production function estimation results (t-statistics in parentheses)

Three-stage least-squares regression								
	2007-08		2009-10		2010-11		2011-12	
	Coef.	Z	Coef.	Z	Coef.	Z	Coef.	Z
rxq								
rxs	0.343	[0.27]	1.673	[2.49]	-5.628	[-6.85]	4.882	[2.38]
hg	0.677	[1.04]	0.398	[0.56]	-0.745	[-0.27]	-19.691	[-1.02]
eg	0.387	[0.38]	-2.225	[-1.58]	-6.34	[-2.09]	-0.52	[-0.04]
age	0.813	[0.34]	-1.60	[-1.32]	11.843	[6.69]	-8.23	[-1.56]
female	0.089	[1.03]	-0.026	[-0.28]	0.463	[1.52]	-1.227	[-1.18]
uninsured	-0.157	[-0.27]	1.161	[1.89]	-4.089	[-4.98]	3.481	[1.39]
lw	-0.062	[-0.34]	-0.246	[-1.82]	0.086	[0.17]	0.217	[0.10]
white	-0.083	[-0.26]	-0.174	[-1.63]	2.089	[4.65]	-1.527	[-1.65]
asian	-0.275	[-2.12]	0.208	[0.77]	-1.15	[-1.86]	-0.78	[-0.51]
hisp	0.011	[0.02]	0.354	[1.41]	-3.356	[-5.51]	3.13	[1.27]
_cons	-2.712	[-2.21]	1.149	[0.46]	-12.437	[-1.28]	14.484	[0.48]
chi 2 (P-value)	525.3	(0.000)	190.62	(0.000)	59.23	(0.000)	11.67	(0.307)
rxs								
hg	-0.442	[-0.68]	0.325	[0.33]	-0.17	[-0.19]	3.93	[0.55]
eg	0.685	[0.69]	1.831	[1.85]	1	[1.06]	0.414	[0.08]
age	1.936	[9.29]	1.793	[8.21]	1.832	[6.06]	2.044	[1.57]
female	-0.055	[-0.57]	0.101	[1.09]	0.05	[0.52]	0.265	[0.74]
uninsured	-0.455	[-2.91]	-0.885	[-4.16]	-0.659	[-3.31]	-0.833	[-1.11]
lw	-0.135	[-0.97]	0.144	[1.07]	0.024	[0.15]	-0.021	[-0.02]
white	0.251	[2.28]	0.095	[0.77]	0.346	[3.08]	0.334	[1.32]
asian	-0.023	[-0.10]	-0.331	[-1.52]	-0.161	[-0.83]	0.099	[0.16]
hisp	-0.461	[-3.49]	-0.336	[-2.16]	-0.549	[-4.14]	-0.634	[-0.75]
_cons	0.092	[0.04]	-2.88	[-1.30]	-1.579	[-0.51]	-4.159	[-0.35]
chi 2 (P-value)	278.82	(0.000)	149.87	(0.000)	195.63	(0.000)	72.25	(0.000)
Heteroscedasticity Tests	Chi2(1)	P-value	Chi2(1)	P-value	Chi2(1)	P-value	Chi2(1)	P-value
Breusch-Pagan LM Test	62.08	0.000	1581.323	0.000	1831.748	0.000	2223.409	0.000
LR Test	63	0.000	2610.41	0.000	7615.65	0.000	7877.02	0.000
Wald Test	1411	0.000	1642.465	0.000	1813.42	0.000	2117.23	0.000
Hansen-Sargan overidentification statistic	0.177	0.674	0.363	0.547	0.239	0.625	0.047	0.829

In consistence with previous findings, we also find that *age* and *female* consume more prescription drugs, the effects of these two variables on both utilization and spending. *female* with significant positive effects on utilization while *age* with significantly positive effects on spending, indicating female individuals tend to consume more drugs with not considerably higher expenditure. In contrast, aged individuals tend to spend more on prescription medicines.

Overall system heteroscedasticity tests for 3SLS with Breusch-Pagan LM, Likelihood Ratio LR, and Wald statistics, suggest no overall system heteroscedasticity in all estimations. Hansen-Sargan overidentification statistics reject the overidentification problem of all instruments for all estimations. The Hausman specification test indicates that the coefficients of the OLS and 3SLS models are not equal statistically.

Comparing with the results form 2SLS estimation, the estimated effects for *hg* and *eg* are more consistent in the 3SLS estimation than those in 2SLS. The comparisons for the uninsured

effects for 2SLS and 3SLS on the prescription drug spending estimations also give out very close results except for 2007-08. These similar estimates of direct effects are indications of the robustness of these two alternative methods of estimation. What we expect from 3SLS are more precise estimates of the parameters given the validity of our specification and the use of the covariance among the disturbances.

Overall, the OLS biased estimates of the parameters due to the endogeneity of hg and *eg*, are inconsistent in the violation of OLS assumptions. Using 2SLS, the estimation of the parameters becomes consistent. Simultaneously, it still needs to be improved since *age* and *wage* (Duplicate problem in 2SLS) have both direct and indirect effects on drug utilizations. Finally, with 3SLS, estimations are both consistent and efficient than those obtained by OLS and 2SLS. The 3SLS approach can correct the endogenous problem (in OLS) by *hg* and *eg* and duplicates issue (in 2SLS) by *age* and *wage*.

CONCLUSION

This paper investigates the effects of gaining employment and health status on prescription drug utilization in the United States, emphasizing the switch of unemployment to employment status and less healthy to healthy status for working-age individuals, using the Two-year Longitudinal MEPS Data, 2007-2016. The findings suggest that the effects of changing status from unemployment to employment strongly affect prescription drug utilization, from 2007 to 2016, while no significant effects on drug spending. The longer time on gaining employment is associated with more prescription drug utilization, which is the sign of delaying health services behavior related to employment-based health insurance. However, our finding contradicts to those of Kozman et al. (2012). Why? By individually check for 2009-2010 and 2010-2011, we find the negative relationship between gaining employment and drug utilization; the longer it takes the individual to find a job, the lower the drug utilization, which suggests a change in individuals' preferences for drug utilization patterns during the great recession because of the extreme budget concerning (Fodman and Book 2010). Despite this, in general, there is a positive relationship between the gaining employment process and prescription drug utilization; hence, a good economy and shorter time for individuals' getting employment will lower the drug utilization in the sense that timely treatment could reduce the severity of illness.

We also find significant positive effects of gaining healthy status on both prescription drug utilization and spending during this period, consistent with the original suggestion of the Grossman model. Results from the simultaneous equation system demonstrate the pathways through which the gaining of employment and health factors influence prescription drug utilization and suggest an essential role for individual employment and health status and their determinants.

Besides, age has both a direct and indirect effect on prescription drug utilization and spending; we find that age had a positive relationship with both drug utilization and expenditure. Education has an indirect impact on drug utilization by affecting the individual's health status, finding suggests that the more the education years, the less likely, the faster the gaining of health status. Wages only has indirect effects on drug utilization by affecting both the changes of health status and employment status, no direct effects on drug utilization and spending. In our study, results suggest that the regional employment-population ratio indirectly affects drug utilization and spending by changing the individual's gaining of the employment process.

There are a few limitations in this research and potential improvement in future research. One limit is the MEPS data, which only includes the number of prescriptions the respondents filled, not the number of drugs the doctor written. However, as researchers acknowledged a complicated situation in the pharmaceutical area, the physicians are the chief players. The latter specifies the prescriptions used by individual patients (Alowi and Kani, 2019).

Another limitation of this analysis is the UNINSURED, which might be an arguable variable

here in this research since it only captures the uninsured individuals. This variable provides no information about the copay, deductible, and out of pocket, the most effective insurance factors on prescription drug utilization. Moreover, the insurance types also affect drug utilization; for example, Medicaid and Medicare recipients might utilize drugs differently than individuals that are privately insured or employment-based insurance recipients. However, the uninsured status can exactly give evidence of how the aiming (affordability of health insurance, lowering the uninsured rate) of the Affordable Care Act works in the sense of adverse selection problem. Checking the uninsured effects on drug utilization, we find no significant differences before and after 2009-2010.

Meanwhile, there was a significantly negative effect on the higher coefficient value of UNINSURED on both drug utilization and spending for 2010-2011, which created a significant gap between uninsured and insured individuals. This result confirmed the ACA insurance benefit, which suggests that individuals' prescription drug utilization decision was significantly affected by the insurance status during the launch of the Affordable Care Act, while only for a limited period. In this study, we have the regional employment to population ratio, a location variable, as an instrumental variable to capture the effect of gaining employment on prescription drug utilization. While in the future study, it might be possible to capture the changes in insurance status based on the differential expansion of Medicaid if we can effectively control for location in the future study.

REFERENCES

Alowi M., Kani Y. (2019) Promotion of Prescription Drugs and Its Impact on Physician's Choice Behavior. J Pharma Care Health Sys 6: 200. doi:10.4172/2376-0419.1000200

Bass, U.F., Woodward J.A., 1978. Services research report. Skills training and employment for ex-addicts in Washington, D. C.: A report on TREAT. DHHS Publication No. (ADM): 78-694.

Berger M, Leigh JP (1989), Schooling, Self-Selection, and Health, Journal of Human Resources 24 (3):433-455.

Davis, M., Poisal, J., Chulis, G., Zarabozo, C., Cooper, B., 1999. Prescription drug coverage, utilization, and spending among Medicare beneficiaries. Health Affairs 18(1): 231-243.

Fillenbaum, G.G., Hanlon, J.T., Corder, E.H., Ziqubu-Page, T., Wall, W.E., Brock, D., 1993. Prescription and Nonprescription Drug Use among Black and White Community-Residing Elderly. American Journal of Public Health 83 (11).

Fodeman J., Book R.A. (2010) "Bending the Curve": What Really Drives Health Care Spending. Backgrouder No. 2369.

Gibson, T.B., McLaughlin, C.G., Smith, D.G., 2005. A Copayment Increase for Prescription Drugs: The Long-Term and Short-Term Effects on Use and Expenditures. Inquiry 42: 293–310.

Grossman M., 1972. On the Concept of Health Capital and the Demand for Health. The Journal of Political Economy, 80(2): 223-255.

Grossman M (1999) "The Human Capital Model of the Demand for Health," NBER Working Paper 7078.

Grossman M (2000) The human capital model. In Handbook of Health Economics, Volume 1 (Eds. A.J. Culyer and J.P. Newhouse). Elsevier Science BV.

Haislmaier E, Badger D (2018) How Obamacare Raised Premiums. Backgrouder No. 3291.

Han, M., Rizzo J. A., 2012. The Health Returns to Medical Expenditures: Does Age Matter?

Henkel D (2011) Unemployment and substance use: a review of the literature (1990-2010). Curr Drug Abuse Rev 4(1):4-27.

Holahan, J., 2011. The 2007–09 recession and health insurance coverage. Health Affair (Millwood) 30(1): 145–152.

Kozman, D., Graziul, C., Gibbons, R., Alexander, G.C., 2012. Association between unemployment rates and prescription drug utilization in the United States, 2007–2010. BMC Health Services Research 12:435.

Leibowitz, A., Manning, W.G., Newhouse, J.P., 1985. The demand for prescription drugs as a function of cost-sharing. Social Science & Medicine. 21(10): 1063–1069.

Long SH. (1994) Prescription Drugs and the Elderly: Issues and Options. Health Affairs 13:2, 157-174.

Lusardi, A., Schneider, D., Tufano, P., 2010. The economic crisis and medical care usage. Natl Bur Econ Res Working Paper Ser, 15843:1–14.

Lusardi A, Schneider DJ, Tufano P (2015) The economic crisis and medical care use: comparative evidence from five high-income countries. Social Science Quarterly 96(1). DOI: 10.1111/ssqu.12076

Okunade, A. A., You, X., Koleyni, K., 2018. Cross-country Medical Expenditure Modeling Using OECD Panel Data and ARDL Approach: Investigating GDP, Technology and Ageing Effects. Health Econometrics in Contributions to Economic Analysis, Emerald Publishing.

Roe, C.M., McNamara, A.M., Motheral, B.R., 2002. Gender- and Age-Related Prescription Drug Use Patterns. Ann Pharmacother 36(1): 30-39.

Rosen S, Taubman P (1982) Changes in Life-Cycle Earnings: What Do Social Security Data Show? Journal of Human Resources 17(3): 321-338.

Silverman K., Robles E. (1999) Employment as a drug abuse treatment intervention: A behavioral economic analysis. In: Chaloupka F.J, Bickel W.K, Grossman M, Saffer H, editors. The economic analysis of substance use and abuse: An integration of

econometric and behavioral economic research. Chicago: University of Chicago Press pp. 279–299.

Simoni-Wastila, L., Ritter, G., Strickler, M.S., 2004. Gender and Other Factors Associated with the Nonmedical Use of Abusable Prescription Drugs. Substance Use & Misuse 39: 1089-1098.

Sjoqvist, F., Birkett, D., 2003. What is drug utilization research and why is it needed? WHO booklet “Introduction to Drug Utilization Research”.

Skoog, J., Midlov, P., Borgquist, L., Sundquist, J., Halling, A., 2014. Can gender difference in prescription drug use be explained by gender-related morbidity?: a study on a Swedish population during 2006. BMC Public Health 329: 1471-2458.

Stronks K., van de Mheen H, van den Bos J, Mackenbach JP (1997) The interrelationship between income, health and employment status. Int J Epidemiol 26(3):592-600.

Svarstad, B.L., Cleary, P.D., Mechanic, D., Robers, P.A., 1987. Gender Differences in the Acquisition of Prescribed Drugs: An Epidemiological Study. Medical Care 25(11): 1-23.

Sznitman, S.R., 2007. Drugs and gender: A contradictory project in interviews with socially integrated men and women who use drugs. Nordic Studies on Alcohol and Drugs 24: 107-126.

Theodorou M., Kaitelidou D., Galanis P., Middleton N., Theodorou P., et al. (2011) Quality of life measurement in patients with hypertension in Cyprus. Hellenic J Cardiol 52(5):407-15.

Truffer, C.J., Keehan, S., Smith, S., Cylus, J., Sisko, A., Poisal, J.A., Lizonitz, J., Clemens, M.K., 2010. Health spending projections through 2019: the recession’s impact continues. Health Affair (Millwood) 29(3): 522–529.

You, X., Okunade, A. A., 2017. Income and Technology as Drivers of Australian Healthcare Expenditures. Health Economics 26(7).

DECISION SCIENCES INSTITUTE
Comprehensive Cost Analysis of Prostate Cancer Treatment: Direct and Indirect Costs

Rachael Selepa
University of Wisconsin – La Crosse
Email: selepa.rachael@gmail.com

Uzay Damali
University of Wisconsin – La Crosse
Email: uzaydamali@uwlax.edu

Mehmet Kocakulah
University of Wisconsin La Crosse
Email: mkocakulah@uwlax.edu

ABSTRACT

Prostate cancer is the second most common cancer among men in the US (CDC, 2019). While the research on prostate cancer is extensive, this research does not estimate an overall cost, considering all the direct (e.g., appointments and medication) and indirect (e.g., dealing with complications) costs associated with treatments of prostate cancer considering different risk levels and stages of diagnosis. Addressing this research gap is critical for the optimal use of healthcare resources, so has potential to improve quality of care. We quantified this total cost through a systematic literature review.

KEYWORDS: Health Care, Cost Benefits Analysis, Healthcare Operations,

INTRODUCTION

One of every four deaths in the United States is due to cancer, with prostate cancer is the second most common cancer among men in the US (CDC, 2019). To address this healthcare concern, better screening tests and treatments have been developed over the years. The results have been optimistic, the rate of prostate cancer deaths are significantly reduced by better screening tests and treatments. However, the number of new cases and deaths per year are still increasing due to an increase in the US population (CDC, 2019). So the research and development on improving screening tests and treatments of prostate cancer continues to be relevant.

One stream of academic research on the topic of prostate cancer is the estimation of screening tests and treatment costs. An extensive amount of research has been published on this topic, and a general finding is that patients with lower risks (typically early stage diagnosis and cancer is progressing slowly) have lower cost and wider variety of screening and treatment options, while patients with higher risk (later stage diagnosis and cancer is progressing rapidly) have higher cost and less variety of screening and treatment options. Further, this research listed many different screening tests and treatment options (e.g., hormone deprivation therapy, radiation therapy, surgery) and estimated a cost for each.

However, this research does not estimate an overall cost, considering all the direct (e.g., appointments and medication) and indirect (e.g., dealing with complications) costs associated with screening tests and treatments of prostate cancer for early and late stage diagnosis (Bardi, 2013).

Estimation of an overall cost of treatment of prostate cancer is critical for the quality of care. Treatment of prostate cancer is a long-term process that requires complex decision making which requires an extensive amount of medical and cost related information. While medical information is available for both oncologists and patients, the costs, particularly overall cost information, is not available or well understood (Henrikson et al., 2014). Overall cost information is critical for the optimal use of healthcare resources, such as when the patients do not understand the overall direct and indirect costs of a long-term treatment after diagnosis, they may not afford other critical therapies in the later stages which may negatively affect the quality of the care. In addition, having an idea of total cost can help new technology adoptions in this area as people want to know the relative advantage of new technologies from the perspective of cost and experience. Addressing this research gap has the potential to improve quality of care in prostate cancer treatment.

The purpose of this research is to quantify the total cost (direct and indirect costs) of the first-line (initial choice of treatment) and second-line treatments (next choice of treatment, when the first-line of treatment is no longer effective) for prostate cancer, considering the level of risk (low, high and metastatic). Specifically, risk is discussed in terms of the stage of the cancer as well as the progression rate. We quantify this total cost through a systematic literature review. This review showed that first, depending on the level of risk, there is significant variance in the first-line and second-line therapies. In most cases, first-line therapies were cheaper than second-line therapies. Second, due to advancements in diagnostic technologies, more patients are diagnosed earlier (lower risk), and therefore they are able to choose a wider variety of less intensive treatment options, such as active surveillance or hormone deprivation therapy. And usually, patients in the second line therapy receive these treatments are more aggressive, including surgery, radiation therapy, hormone therapy, and chemotherapy. Third, while there is much insight about direct costs, the insights about indirect costs (e.g., treating toxicity) are significantly lacking.

LITERATURE REVIEW

We searched three groups of terminologies in PubMed. The first search phrase is “prostate cancer OR metastatic prostate cancer OR mCRPC.” Here mCRPC is metastatic castrate-resistant prostate cancer. The second terminology “cost OR treatment cost OR cost of treatments” terms to the search. The third phrase varied based on the treatment that we were focused on. For example, when focusing on low-risk first-line treatments we would use the phrases “active surveillance OR watchful waiting”. We further narrowed our search by excluding studies done outside the United States. In addition, we excluded articles promoting the use of one drug over another and those that did not contain any numerical cost information.

We created the following tables, below, to organize this research that outlines first- and second-line treatments for low risk, high risk, and metastatic prostate cancer. We further divided our information by direct and indirect costs. A first-line treatment refers to the first treatment specifically for prostate cancer that a patient undergoes. If the cancer no longer

responds or PSA levels continue to rise during first-line treatment then that treatment will be stopped and another treatment will be started, the second-line treatment. Low risk prostate cancer is defined as prostate cancer that is slowly progressing, or is at a low risk of progressing. Prostate cancer that is at risk of spreading quickly or increasing rising PSA levels would be associated with a high risk. If a patient's cancer has metastasized, that patient is no longer low-risk even if their cancer is still slow to spread throughout the body (Informed Health, 2018). However, that patient may have PSA scores low enough not to be considered high risk. For these reasons, we have metastatic treatments standing alone.

Table 1: First-Line Treatments Mean Cost Summary

	First-Line Direct Costs	Yearly Cost	Life-time Cost	First-Line Indirect Costs
Low Risk	Active Surveillance and Watchful Waiting	$8,902	$70,780	Toxicity management and some pain management costs were imprecisely estimated while other indirect costs were not estimated at all. Costs not estimated include travel time and expenses, hospital/ clinic staff wages, etc.
	Radical Prostatectomy	12,431	N/A	
	Brachytherapy	N/A	71,978	
	Brachytherapy and Hormone Therapy	81,543	N/A	
High Risk	Radical Prostatectomy	12,431	N/A	
	Radiation Therapy	79,200	132,939	
	Radiation Therapy and ADT	17,385	23,488	
Metastatic	ADT	Varies by drug	Varies by drug	
	Chemotherapy	N/A	N/A	

Table 2: Second-Line Treatment Mean Costs Summary

	Second-Line Direct Costs	Yearly Cost	Life-time Cost	Second-Line Indirect Costs
Low Risk	Salvage Prostatectomy	$30,881	N/A	Toxicity management and some pain management costs were imprecisely estimated while other indirect costs were not estimated at all. Costs not estimated include travel time and expenses, hospital/ clinic staff wages, etc.
	Radiation Therapy	79,200	132,939	
High Risk	Salvage Prostatectomy	30,881	N/A	
	Radiation Therapy and ADT	29,968 – 93,894	177,704	
	ADT	Varies by drug	Varies by drug	
Metastatic	Chemotherapy	N/A	N/A	
	Sipuleucel-T	93,000	93,000	

Low Risk First-Line Direct Costs

Based on the literature, we identified treatment options for first-line treatments of low risk prostate cancer as active surveillance or watchful waiting, a radical prostatectomy, and radiation therapy.

When prostate cancer is first diagnosed and low risk, patients have multiple options whose price range varies considerably. From the patient's perspective, the cheapest treatment option one could choose is watchful waiting and schedule check-ups at the sign of any new symptoms. These check-ups are typically covered by private insurance or Medicare/Medicaid (Nguyen et al., 2019). Active surveillance would be the next financially best choice for the patients initially. The cost to the patient in the first year is about $8,902. However, if the cancer does not progress for many years, the lifetime costs can rise to $70,780. As for the healthcare system, active surveillance and watchful waiting cost about $8 billion a year. Patients could also undergo a radical prostatectomy. This treatment costs about $12,431, more than one year of active surveillance but with the added possibility of little to no treatment if the cancer is successfully removed. Lastly, brachytherapy supplemented by hormone therapy can incur the most cost for patients with about $71,978 in a lifetime but is more effective in slowing disease progression. No matter what costs are estimated in these first-line treatments, a patient's treatment choice is restricted by the speed of the cancer's growth and their current overall health status.

Active Surveillance and Watchful Waiting

When prostate cancer is diagnosed early, one of the most common treatment options are active surveillance or watchful waiting. Both include close monitoring of disease progression. However, active surveillance uses tests to track cancer growth while watchful waiting looks for a change in symptoms over time. Tests conducted in active surveillance include prostate specific antigen (PSA) blood tests, digital rectal exam (DRE), transrectal ultrasound (TRUS), and biopsies. DRE tests are the most common and least expensive, however PSA tests are best for a more specific diagnosis (Saigal & Litwin, 2002). On a national level, Saigal and Litwin estimate the cost of

screening at about 8 billion USD with about 2 billion estimated for PSA screening alone (2002). The cost to the healthcare system for a single patient is estimated at about $3,415 for DRE only, $5,516 for DRE and PSA tests, and $11,135 for a patient to have DRE, PSA tests, and TRUS (2002). From a payer's perspective, the mean annual costs for screening is estimated at $8,902 and lifetime costs at $70,780 (Nguyen et al., 2019). This lifetime cost estimate could be even higher if a patient is diagnosed early enough and their cancer does not progress for many years.

Radical Prostatectomy

There are two surgery options, radical and salvage, for prostate cancer that removes the prostate and surrounding tissue including the lymph nodes. A radical prostatectomy is offered when the cancer is still localized to the prostate. As a first-line therapy, the intention is to remove as much cancerous tissue as possible before it can spread. In a 2012 study of SEER Medicare Data by Prasad and colleagues, a radical prostatectomy is estimated to cost $12,431. This includes inpatient, outpatient, physician wage, emergency room visits, readmission, and additional surgical or radiologic procedures. These costs were only included in a total estimate, therefore, we were not able to identify itemized costs for a radical prostatectomy.

Brachytherapy

Radiation therapy can be administered internally or externally and has become more precise as technology evolves. Improving the precision in which the radiation is given also improves treatment effectiveness but also limits toxicity in the body. When used as a first-line treatment, patients are treated with brachytherapy and oftentimes hormone deprivation therapy simultaneously (Brennon & Straus, 2019). Brachytherapy, a form of internal radiation, is when radioactive seeds are placed near or in a tumor. Due to the proximity of the seeds to the tumor and surrounding organs, this treatment releases a higher dosage of radiation (Prostate Cancer Foundation, n.d.). This has the potential to lead to higher cost incurred due to toxicity. The high probability of toxicity and other side effects may explain results found by the study conducted by Mahmood and colleagues that showed that the use of brachytherapy has been declining since 2005 (2014). Though this decline could also be a result of better targeting available with technological improvements of external radiation therapies. Despite its decline, brachytherapy is still a viable treatment option. Unfortunately, due to its decrease in popularity few studies isolate the cost of this type of radiation. According to Vu et al. brachytherapy estimates a lifetime cost of $71,978. Though brachytherapy has a high initial cost it reduces overall lifetime costs of treatment because it decreases the risk of metastatic castrate-resistant prostate cancer (2017).

Brachytherapy and Hormone Therapy

Androgen deprivation therapy (ADT) is a type of hormone therapy that prevents androgens, male hormones, from reaching and fueling prostate cancer cells. Low-risk prostate cancer or non-progressing prostate cancer are not treated with ADT alone as the risks and long term side effects outweigh the benefits in low-risk patients or localized prostate cancer (Prostate Cancer Foundation, n.d.). When brachytherapy and ADT are used in tandem, low risk patients may see the best treatment results but at an average annual cost of $81,543 (Vu et al., 2017).

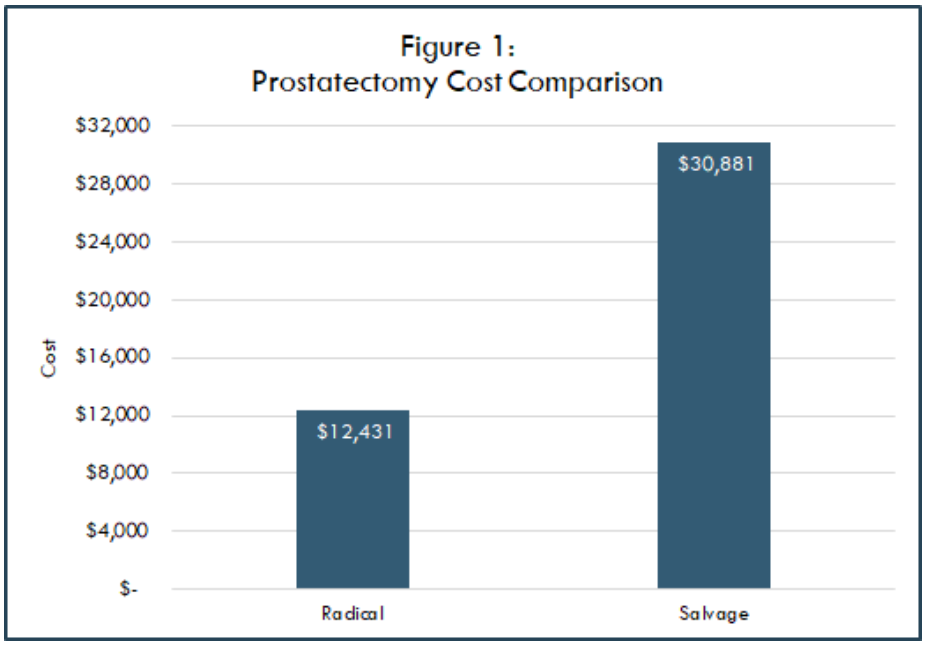
Low Risk Second-Line Direct Costs

If a patient experiences no change in PSA levels despite treatment or a change in symptoms but still their prostate cancer remains relatively low risk they will move to a second line of treatment. Second-line treatments for lower risk patients include salvage surgery or switching to a more aggressive treatment of radiation therapy.

Should a patient's prostate cancer progress but remain low risk, second-line treatments options are similar to their first-line options, given their overall health has remained constant. Younger and healthier patients may choose to have a salvage prostatectomy. Despite shorter and less adverse side effects, compared to radiation therapy, few choose surgery due to the cost, at $30,881, and less effective cancer control (Prasad et al., 2012). The more common second-line treatments for low-risk prostate cancer are radiation therapy. There is little distinction in the literature between external and internal costs. However, IMRT, an external treatment, is estimated at $81,543. If a patient chooses internal radiation therapy, they should expect a lifetime cost of $71,978 (Vu et al., 2017). A more general lifetime estimate for radiation therapy is about $79,200 (Nguyen et al., 2019).

Salvage Prostatectomy

When prostate cancer progresses or relapses and is still localized in the prostate and patients have already undergone a first-line treatment they may choose to undergo a salvage prostatectomy. In this surgical operation the prostate and any surrounding tissue that may contain cancer cells or tumors is removed. According to Prasad et al., factors such as other available treatment options, greater risk of complications, and ineffective cancer control have led to less than 2% of patients to choose a prostatectomy (2012). This study also found that salvage prostatectomy cost $30, 881, nearly $19,000 more than a radical prostatectomy illustrated in Figure 1 below (Prasad et al., 2012).



Radiation Therapy

External radiation therapy (e.g. external beam radiation therapy (EBRT), image-guided radiation therapy (IGRT), and intensity-modulated radiation therapy (IMRT)) are more popular for second-line treatments than for first-line treatments for patients with low-risk prostate cancer. Brachytherapy (internal radiation) is a recommended second-line radiation treatment for low-risk patients but brachytherapy in addition to EBRT is also standard treatment (Dominic et al., 2016). Brachytherapy has an estimated lifetime cost of $71,978 (Vu et al., 2017). Nguyen et al. estimates a lifetime average cost of $132,939 and a median annual cost of $79,200 for radiation therapy, but this estimate does not distinguish between what kind of radiation therapy (2019). Vu et al. did estimate IMRT with a lifetime cost of $81,543 (2017) which is rather close to the combined radiation estimate from Nguyen et al. above. Unfortunately, the literature lacks any other estimates that differentiate between radiation therapy treatments. Therefore, our estimates will use the median estimate of $79,200 for a broad definition of radiation therapy and $81,543 for IMRT where it is appropriate. Though this will make our estimates less precise when comparing treatments for high and low risk prostate cancer. Table 3 summarizes and compares the cost of possible radiation treatments for low risk prostate cancer.

Table 3: Cost Comparison of Radiation Treatments

Treatment	Annual Cost	Lifetime Cost
External*	$79,200	$132,939
Internal**	N/A	$71,978
Brachytherapy + ADT**	$81,543	N/A
*Nguyen et al., 2019 **Vu et al., 2017		

Low Risk Indirect Costs

A prostatectomy of any kind may be the best chance of eliminating the cancer from the body but it comes with more than just physical side effects. Surgery can cause incontinence, impotence and include the removal of testicles. These can all lead to a decline in mental health, therefore it is not uncommon for a patient's doctor to recommend some form of therapy during and after treatment.

Radiation therapy causes side effects that can be treated with over the counter drugs unless symptoms are severe enough for a prescription. Below in Table 4, is an incomplete list of side effects that one might experience with radiation therapy and the cost to treat them that a patient might expect.

Table 4: Cost of Side Effect Treatments: Radiation Therapy

Side Effect	Treatment	Cost
Abdominal Pain	Diet change, prescribed pain killers	$22.85
Anemia	Erythropoiesis-stimulating agents (EAD)	$491.00
Diarrhea	Over-the-counter products	$11.67
Erectile dysfunction	Sildenafil (Viagra), tadalafil (Cialis), vardenafil (Levitra)	$74.00
Incontinence	Medical devices	$1.50 - $10.00
	Medication	$19.60
	Surgery	$17,500
Neutropenia	ESAs, antibiotics	$358.22
Thrombocytopenia	Lower radiation or prescribed (oprelvekin)	($6,865.62)
*See Appendix for cost calculations		

High Risk First-Line Direct Costs

The majority of prostate cancer diagnoses happen in early or regional stages of the disease with a low chance of the cancer spreading. However, there are some cases where the cancer may appear localized but have a high risk of spreading quickly. In such cases the goal is to control the cancer before it can do any more harm. Treatments to do so include radical surgery, radiation therapy, and radiation therapy with ADT. If a patient has already had a first-line treatment of their low risk cancer and the cancer has progressed or relapsed to become high risk, then they may receive any of the treatments in this section. However, the treatment will be considered their second-line treatment, despite the fact that this would be their first treatment for their high-risk cancer.

To be first diagnosed with high risk prostate cancer the main treatment goal is to lower PSA levels quickly to slow the progression of the cancer. Of the first-line treatments available for high risk prostate cancer a radical prostatectomy not only has the lowest direct costs but also has the highest probability of the patient going into remission. The cost to surgically remove as much cancerous tissue as possibly is estimated at about $12,431. However, many patients are not comfortable with such a surgery. The most popular treatment options include some form of radiation therapy. A patient may expect to pay an average of $70,780 annually for external radiation therapy with a lifetime cost of about $132,393. As high risk prostate cancer requires more aggressive treatment, some patients may have a pre-treatment of ADT. This additional treatment can increase the cost of treatment to about $88,164 annual and about $177,704 over a lifetime.

Radical Prostatectomy

As stated above, the goal in high risk prostate cancer treatment is to control it while it is most manageable. The quickest way to do so is to remove any and all cancerous tissue in and around the prostate in a radical prostatectomy. Prasad and colleagues estimate this to cost about $12,431 (2012).

Radiation Therapy

Radiation options are very similar between low and high risks cancer groups. The most notable differences are in the frequency, dosage, and combination of treatments. In low risk prostate cancer treatments, brachytherapy might be combined with small dosages of ADT. For high-risk prostate cancers, internal and external radiation will be used together to increase the effectiveness of the treatments. These treatments usually involve higher dosages, more frequent treatments (weekly or daily), or both in an attempt to stop the cancer from spreading any further in the body. Broadly speaking, the mean annual estimate for radiation therapy is $70,780 and $132, 939 over a lifetime (Nguyen et al., 2019). For IMRT, a common treatment for the cancer that has spread outside the prostate cancer, a lifetime estimate is about $81,543 (Vu et al., 2017).

Radiation Therapy and ADT

In some cases, ADT may be a “pre-treatment” to external beam radiation therapy for a more effective treatment (Nguyen et al., 2019). It is more common to combine radiation and ADT for high risk prostate cancer as it is a more aggressive way to lower PSA levels and prevent the cancer from metastasizing. Studies have shown that this combination has led to increased long-term survival (Prostate Cancer Foundation, n.d.). ADT added to radiation reduces disease progression, has a lower frequency of severe toxicity, and the risk of death from prostate cancer (Warde et al., 2011). Nguyen and colleagues mention that the initial ADT cost $10,410. Adding ADT to radiation therapy increased this cost to $17,384. Within five years that is a cost increase from $23,199 to $23,488 over five years (2019). In this instance, ADT is a pre-treatment and not a stand-alone, full regiment hormone therapy. Therefore, the estimated treatment cost of ADT at $17,384 in the initial year is substantially lower than our estimates later in this report. This cost would then have to be added to the cost of radiation therapy as they are being used in tandem. For simplicity, we added these ADT estimates to the mean radiation costs to calculate the first annual cost of $88,164.

High Risk Second-Line Direct Costs

If the previously discussed treatments fail to reduce PSA levels and the progression of the cancer, a patient may switch to an even more aggressive treatment. If one has not already had a surgery, they may undergo a salvage prostatectomy. If they were only receiving radiation therapy, they will likely have ADT added to their treatment. Patients can also be added to an ADT-only therapy.

At this point we start to see that as prostate cancer progresses treatments become more aggressive which increases the direct costs. Salvage prostatectomy costs nearly $19,000 more than a radical prostatectomy. Nguyen and colleagues calculated first year costs of radiation therapy and ADT to be between $29,968 and $93,894, demanding on health, with a lifetime

average of $177,704 (2019). First-line treatment costs ranged from $23,199 to $23,488. ADT-only has costs ranging from $10,290 to $95,485 with an average cost of $62,619 without the added cost of treating toxicity.

Salvage Prostatectomy

Should a patient's prostate cancer progress or relapse, a salvage prostatectomy would be one way to remove as much of the cancerous tissue as possible at an estimated $30,881 (Prasad et al., 2012). However, not many will be eligible for this second-line treatment if they were already high risk since there will be higher risks associated with the operation.

Radiation Therapy and ADT

When a patient's cancer has become high risk one option for them is to have a full regiment treatment of ADT and radiation therapy. Previously, we discussed the number of radiation therapies available. Later, we will discuss the different ADT drugs and dosages. These variations in radiation therapy and ADT as well as the overall health of a patient can make estimating treatment costs difficult. A mean annual cost estimate of combined radiation therapy and ADT for earlier treatment of healthy patients is about $29,968 and as much as $93,894 for a patient nearing a terminal stage. Over a lifetime, the mean cost is $177,704 (Nguyen et al., 2019).

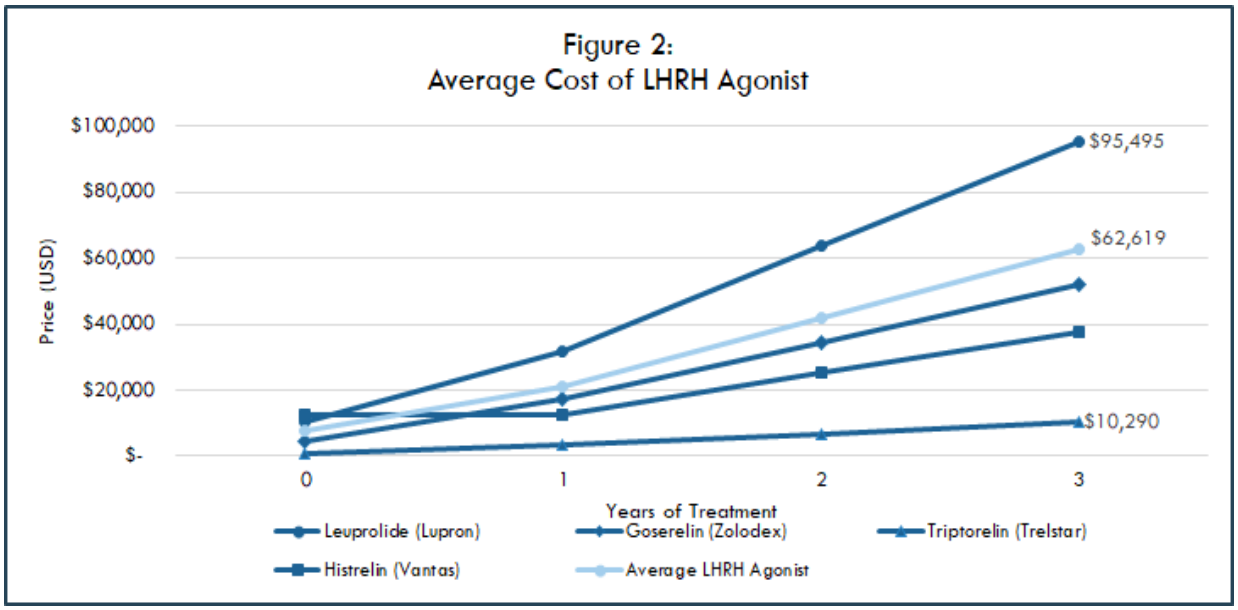
ADT

In treating prostate cancer, androgen deprivation therapy is used to two forms of luteinizing-hormone releasing hormones. Luteinizing-hormone releasing hormones (LHRH) initiate the release of testosterone. LHRH agonist and antagonist block the release of this hormone which then causes cancer cells to "starve" as testosterone acts as a fuel for prostate cancer cells. LHRH agonist can cause flare ups of testosterone which can cause bone pain if the cancer has already metastasized, or spread to the bones, so it is more popular in earlier treatment (Prostate Cancer Foundation, n.d.).

Using drug information from the University of Wisconsin Genitourinary (UW GU) Center's charge master, the average cost of four common LHRH agonist drugs over three years were calculated. Three years is the average length of ADT treatment (Prostate Cancer Foundation, n.d.) and each drug has a different dosage amounts and frequencies that were noted within the charge master spreadsheet provided by the UW GU Center. These drugs are Leuprolide (Lupron), Goserelin (Zolodex), Triptorelin (Trelstar), and Histrelin (Vantas). Leuprolide had twelve different dosages with varying frequencies of the injection. Goserlin has three different dosages but all are given every three months. Table 5 breaks down the costs for each drug from the first treatment to the third year average. Figure 2 compares the cost of the four drugs and their combined averages from the first treatment to the third year average.

Table 5: LHRH Agonist Cost Comparison

LHRH Agonist (Brand Name)	Procedure	Cost per Dosage	Year 1	Year 2	Year 3
Leuprolide (Lupron)	Leuprolide Acetate (6 Month) For 45 MG	$14,989.44	$29,979	$59,958	$89,937
	Leuprolide Acetate (3 Month) For 11.25 MG	$6,350.89	$25,404	$50,807	$76,211
	Leuprolide Acetate (4 Month) For 30 MG	$10,028.19	$40,113	$80,226	$120,338
	Average Leuprolide	$10,456	$31,832	$63,663	$95,495
Goserelin (Zolodex)	Goserelin Acetate Implant 10.8 MG (3 Month)	$6,277.11	$25,108	$50,217	$75,325
	Goserelin Acetate Implant 3.6 MG (3 Month)	$2,374.86	$9,499	$18,999	$28,498
	Average Goserelin	$4,326	$17,305	$34,608	$51,912
Triptorelin (Trelstar)	Triptorelin (3 Month) For Inj 11.25 MG	$857.52	$3,430	$6,860	$10,290
Histrelin (Vantas)	Histrelin Acetate Implant Kit 50 MG	$12,578.26	$12,578	$25,157	$37,735



Triptorelin has the lowest initial cost at $858, Histrelin has the highest at $12,578. However, Histerlin is a once a year treatment while Triptorelin is given four times a year. After three years of treatment, Triptorelin still has the lowest average cost at $10,290 and Leuprolide has the largest average cost at $95,485.

The most common LHRH antagonist that patients are treated with is Degarelix. From the dosage information from the UW GU the first dosage of Degarelix is the largest at 120 mg (Brennon & Straus, 2019). Then every 28 days, patients are given an 80 mg dose for about three years. A 120 mg injection cost about $2,963.95 and an 80 mg injection cost about $1,937.80. By the end of year one the cost accumulates to $24,280. By year two and three the cost increases to $42,632 and $63,947, respectively.

High Risk Indirect Costs

Before toxicity and other side effects are considered in an LHRH agonist treatment, patients will receive an anti-androgen during their treatment to prevent PSA flare ups. ADT also comes with long-term side effects that would typically be prevented due to testosterone such as weight loss or gain, loss of libido, mood changes, and depression (Prostate Cancer Foundation, n.d.). These types of treatment side effects are not easily quantifiable. However, the cost of anti-androgens and treatment for other side effects are seen in Table 6 and Table 7, respectively.

Table 6: Average Cost of Anti-Androgens

Drug	Cost per Dosage	Year 1	Year 2	Year 3
Bicalutamide	$59	$234	$439	$703
Flutamide	$71	$285	$570	$855
Enzalutamide	$740	$4,438	$8,876	$13,314
Apalutamide	$5,370	$62,222	$124,444	$186,665
Average	$1,560	$16,795	$33,590	$50,384

Table 7: Cost of Side Effects: LHRH Agonist

Side Effect	Treatment	Cost
Abdominal Pain	Diet change, prescribed pain killers	$22.85
Anemia	Erythropoiesis-stimulating agents (ESAs)	$491.00
Breast Tenderness	Removal of breast tissue	$3,978.00
	Tamoxifen	$19.85
Erectile dysfunction	Sildenafil (Viagra), tadalafil (Cialis), or vardenafil (Levitra)	$75.00
Hot flashes	Medroxyprogesterone	$8.24
Osteoporosis	Zoledronic acid	$703.86
*See Appendix for cost calculations		

Metastatic First-Line Direct Costs

Once prostate cancer has spread to the bones, or metastasized, radiation therapy and surgery are no longer the best treatment options. A prostatectomy might occur on a case-to-case basis. ADT is also another possible treatment route, however, Docetaxel chemotherapy is standard care for metastatic prostate cancer. Immunotherapy has also been used against metastatic prostate cancer, however, what few drugs have been approved by the U.S. Food and Drug Administration are not financially available for most patients, thus immunotherapy is not a likely choice as a first-line therapy (Brennon & Straus, 2019). At this disease stage, patients can also enroll in clinical trials though there was not enough literature to make a definitive conclusion.

Degarelix, enzalutamide and Abiraterone have been proven to be the most effective ADT for metastatic prostate cancer. Perhaps due to this limited number of drugs has led to a higher cost compared to other hormonal deprivation therapies. An LHRH agonist can cost as little as $13,000 or as much as $95,485. However, they are not the recommended treatment due to causing flare up in PSA levels which would require additional cost to treat. The LHRH antagonist, Degarelix, has an average cost of $24,280. After three years of treatment it cost about $63,947. Enzalutamide is an anti-androgen that can be used as a primary as well as a supplemental treatment. This duality offers a lower average treatment cost of $6,657. The newer drug Airaterone is harder to price due to brand name and insurance coverage and can range from $1,921 to about $10,000 for a three-month supply. The course of treatment costs an average of about $86,467.

Should a patient undergo chemotherapy instead of hormone deprivation, they will most likely be treated with Docetaxel. From one data source treatment is estimated at about $18,457. With insurance, other studies have identified a total treatment cost ranging from about $2,588 to $13,169.

ADT

Being diagnosed with metastatic prostate cancer requires an aggressive approach to quickly decrease PSA levels and slow down disease progression. A LHRH agonist is not recommended as a treatment as it would cause PSA levels to flare up that could further "feed" cancerous cells throughout the body, though there is a possibility of an agonist being prescribed anyway. Patients are more likely to be treated with the LHRH antagonist Degarelix as it does not cause flare ups. Two other common, aggressive drugs for hormonal therapy are Enzalutamide and Abiraterone and are typically given for castrate resistant prostate cancer.

As previously stated, Degarelix cost about $24,280 within the first year. By the end of the three-year treatment the total cost raises to $63,947. Should a patient be treated with an agonist, direct costs between $10,290 and $95,485 are expected.

Enzalutamide is an anti-androgen that blocks testosterone in areas throughout the body besides the testicles. One use is as a supplement to an LHRH agonist treatment to prevent flare-ups. Whether or not this drug is being used as a secondary or primary therapy, the literature did not provide any previous cost data on Enzalutamide treatments. Therefore, the following information is taken from the data available through the UW GU but is not an accurate representation due to the small sample size. According to the UW GU data, a one-month supply of Enzalutamide costs about $739. The median treatment duration is 10.7 months (Schultz N.M., Flanders S.C., Wilson S., et al., 2018). To account for patients that may have a shorter or longer treatment duration, the drug cost was calculated at the sixth, eighth, tenth month, and twelfth month of treatment for a cost of $4,438, $5,917, $7,396, and $8,876 respectively. Or an average drug cost of $6,657.

Abiraterone or abiraterone acetate is a new and effective drug for castrate resistant prostate cancer. Unlike drugs and their side effects, Abiraterone causes inflammation and prednisone (an anti-inflammatory) is prescribed in tandem when treating prostate cancer. For this reason, the cost of prednisone is considered a direct cost. As Enzalutamide and Abiraterone are the primary drugs to treat prostate cancer at this stage, existing research compares their effectiveness, not their costs. Due to this lack of data and previous cost research, the calculated costs are not a complete representation of the direct price of these drugs. The median length of treatment for Abiraterone is 13.7 months (Schultz N.M., Flanders S.C., Wilson S., et al., 2018). As above, an average was found using costs calculated at durations around this median. From the eighth month to the eighteenth month the average treatment cost was $86,467. Note that the UW GU data contained the brand drug, Zytiga, which was an outlier with costs between $8,00 and $10,000 for a three-month supply of tablets. The generic drug costs about $1,921 for a three-month supply.

Chemotherapy

There are two main chemotherapy drugs that have been proven to treat prostate cancer and be approved by the Food and Drug Administration, Docetaxel and Cabazitaxel. Cabazitaxel is a newer drug that has primarily been used as a second-line treatment, after Docetaxel. Some studies have mentioned using it as a first-line treatment but there was little substantial information about cost. Instead, Cabazitaxel and its costs are discussed in the second-line therapy section.

Docetaxel is an injection given over the course of six weeks. From the data available through the UW GU, one dosage costs $3,076. The total cost of the treatment is about $18,457. In 2017, Dr. Armstrong and co-researchers used Medicare and commercial claims data to

calculate an average cost per treatment. For a Medicare claim the cost was about $2,588 while it was $13,169 for a commercial claim. The significantly higher costs for commercial claims can be attributed to the higher contract rates paid by commercial insurers (Armstrong et al., 2017).

Metastatic Second-Line Direct Costs

If hormone deprivation therapy with Degarelix fails, a patient might then be treated with Abiraterone, enzalutamide, or chemotherapy. The anti-androgen drug, Enzalutamide, has an average treatment cost of $6,657. Abiraterone is a newer drug and a patient would expect to pay an estimated average of about $86,467.

In the case that a patient's first-line treatment was Abiraterone or Enzalutamide it is likely that their second-line treatment will be Docetaxel, a chemotherapy. Docetaxel is a six-week treatment that can cost as high as about $18,457. With insurance claims, cost can range from about $2,588 to $13,169 (Armstrong et al., 2017). Cabazitaxel is also a chemotherapy, though it is almost exclusively used as a second-line treatment after docetaxel. It is given intravenously every three to four weeks for an average of four cycles (Terada, Kamoto, Tsukino,et al, 2019). Using the UW GU cost of about $33,660 per injection, the total cost is estimated at about $134,640 for four treatment cycles.

There is a third option of immunotherapy, specifically Sipuleucel-T. Immunotherapy is not a popular treatment option as many patients may not be able to afford it. However, numerous studies have supported the effectiveness of Sipuleucel-T and have led to a more wide-spread use of this specific therapy. Sipuleucel-T is given in three doses with one to two weeks between each dose. First the patient donates their own cells so that they can be treated before being returned to the patient through an IV (Cleveland Clinic Cancer, 2019). Cost data from the UW GU estimates this treatment at $93,000.

Metastatic Indirect Costs

Enzalutamide and Abiraterone are hormonal deprivation therapies, therefore they have the same long-term side effects as ADT where a patient might consider therapy. Most other side effects can be treated with over-the-counter drugs or prescriptions if needed. Enzalutamide can cause bone pain in some patients which is commonly treated with Denosumab or Zoledronic acid. Both can vary in cost depending on insurance coverage and commercial coupons available.
Docetaxel, Cabazitaxel, and Sipuleucel-T have side effects that can be bought off the shelf or by prescription. Docetaxel and Cabazitaxel have two more costly side effects, peripheral neuropathy, when nerves that carry messages between the brain and the rest of the body, and chemo brain, cognitive damage due to chemotherapy. Peripheral neuropathy is treated with occupational and physical therapy. Chemo brain might also require therapy to repair damaged circuits but can often be performed at home with memory and thinking exercises. Due to the variety of different needs for peripheral neuropathy and chemo brain vary between patients and price charged between facilities a cost was not able to be calculated.

Caveats

The main limitation to consider in this literature review is the various times and time frames that these studies were conducted. The inflation costs were not calculated but should be considered with all estimates throughout this report.

Newer, more relevant data was also limited throughout this project. In some cases, newer studies were not accessible. In others, there has not been any recent and relevant information on treatments. For example, radium-223 is a viable treatment for cancer and bone pain but there is little data outside of clinical trials.

Conclusion

There are many uncertainties during cancer treatments due to the uniqueness of each patient. However, these uncertainties prevent doctors and patients from being fully informed when making a treatment plan. This study has shown that patients in earlier stages or receiving first-line treatment have less invasive options that have limited toxicity and side effects with lower healthcare utilization costs. More aggressive treatments have higher initial costs as well as the need for other healthcare services, such as hospital and staff utilization, outpatient/inpatient costs, and emergency care, as well as greater toxicity treatment costs. This opens the possibility of new or better clinical support technologies. If oncologists could better track disease progression and more precisely target cancerous tissues, then they could give a more effective first-line treatment. This would have the potential to lower cost to patients as they will not have to pay for a more aggressive and thus more expensive treatment. It can also increase the efficiency of the healthcare system to apply a more localized treatment which would limit excess use of resources.

APPENDIX

The following table provides cost estimates obtained from the UW GU charge-master, goodrx.com, and other online sites that offer discounts on prescription medicine or sell over-the-counter drugs for symptom relief. A patient's insurance coverage may lead to a different amount then what is presented.

Side Effect	Treatment	Average Cost	Total Average
Abdominal pain	Ibuprofen	$15.63	
	Naproxen	$30.04	
	Advil	$8.92	
	Motrin	$15.65	
	Aleve	$8.90	
Prescription only	Naprelan	$739.97	
	Anaprox	$51.03	
	Voltaren	$32.66	
			$22.85*
Anemia	Aranesp	$1,847.00	
	Epogen	$336.00	
	Mircera	$445.00	
	Procrit	$537.00	
			$491.00*
Breast Tenderness	Removal surgery	$3,978.00	N/A
	Tamoxifen	$19.85	N/A
Diarrhea	Loperamide	$13.00	
	Imodium A-D	$12.00	
	Bismuth subsalicylate	$10.00	
			$11.67
Erectile dysfunction	Sildenafil (Viagra)	$14.00	
	Tadalafil (Cialis)	$28.00	
	Vardenafil (Levitra)	$180.00	

			$74.00
Hot Flashes	Medroxyprogesterone	$8.24	N/A
Incontinence	Catheters (each)	$1.50	
	Compression device (package)	$10.00	
Medications	Tamsulosin (Flomax)	$12.00	
	Alfuzosin (Uroxatral)	$13.00	
	Silodosin (Rapaflo)	$30.00	
	Doxazosin (Cardura)	$22.00	
	Terazosin	$21.00	
			$19.60
Neutropenia			
Antibiotics	Cipro (ciprofloxacin)	$11.00	
	Certraxel (ciprofloxacin)	$11.00	
	Cilozan (ciprofloxacin)	$11.00	
	Ocuflox (ofloxacin)	$13.00	
	Floxin (ofloxacin)	$13.00	
			$11.80
ESAs	Aranesp	$1,847.00	
	Epogen	$336.00	
	Mircera	$445.00	
	Procrit	$537	
			$791.25
Osteoporosis	Zoledronic acid	$703.86	N/A
Thrombocytopenia	Oprelvekin (Neumega)	$4,543.73	
	Platelet transfusion	$9,187.50	
			$6,865.62
*Median calculated due to outliers			

REFERENCES

Armstrong, A., Bui, C., Fitch, K., Sawhney, T., Brown, B., Flanders, S., Chambers, J. (2017).

Docetaxel chemotherapy in metastatic castration-resistant prostate cancer: Cost of care in Medicare and commercial populations. *Current Medical Research and Opinion*, 33(6), 1133-1139.

Bardi, J. How prostate cancer therapies compare by cost and effectiveness. (2013). States News Service, p. States News Service, Jan 4, 2013. Retrieved from https://www.ucsf.edu/news/2013/01/13370/how-prostate-cancer-therapies-compare-cost-and-effectiveness

Brennon, M. J., RN, & Straus, J., RN. (2019).*UW Health Chargemaster Medications*[Excel]. Madison: UW GU Center. This file contains raw data about medication prices. For calculated prices over time see the Medication Cost file.

Cleveland Clinic Cancer. (n.d.). Provenge. Retrieved August 18, 2019, from http://chemocare.com/Provenge

Flannery, K., Drea, E., Hudspeth, L., Corman, S., Gao, X., Xue, M., & Miao, R. (2017).

Haque, W., Butler, E., & Teh, B. (2017). Stereotactic body radiation therapy for prostate cancer review. *Chinese Clinical Oncology*, 6(Suppl 2), 1-13.

Henrikson, N., Tuzzio, B., Loggers, L., Miyoshi, E., & Buist, T. (2014). Patient and oncologist discussions about cancer care costs. Supportive Care in Cancer, 22(4), 961-967.

InformedHealth.org. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2006-. What is low-risk prostate cancer and how is it treated? 2018 Feb 22. Available from: https://www.ncbi.nlm.nih.gov/books/NBK487255/

Mahmood, U., Pugh, T., Frank, S., Levy, L., Walker, G., Haque, W., Lee, A. (2014). Declining use of brachytherapy for the treatment of prostate cancer. Brachytherapy, 13(2), 157-162.

Nguyen, C., Lairson, D., Swartz, M., & Du, X. (2019). Cost-effectiveness of adding androgen deprivation therapy to radiation therapy for men with advanced prostate cancer from a u.s. payer's perspective. *Journal of Managed Care & Specialty Pharmacy*, 25(2), 228

Prasad, Gu, Kowalczyk, Lipsitz, Nguyen, & Hu. (2013). Morbidity and costs of salvage vs. primary radical prostatectomy in older men. *Urologic Oncology: Seminars and Original Investigations*, 31(8), 1477-1482.

Prostate Cancer Foundation. n.d. Retrieved August 12, 2019, from https://www.pcf.org

Saigal, C., & Litwin, S. (2002). The Economic Costs of Early Stage Prostate Cancer. *PharmacoEconomics,* 20(13), 870

Schultz N.M., Flanders S.C., Wilson S., et al. (2018) Treatment Duration, Healthcare Resource Utilization, and Costs Among Chemotherapy-Naïve Patients with Metastatic Castration-Resistant Prostate Cancer Treated with Enzalutamide or Abiraterone Acetate: A Retrospective Claims Analysis. Adv Ther. 2018;35(10):1639-1655. doi:10.1007/s12325-018-0774-1

Terada, N., Kamoto, T., Tsukino, H. et al. The efficacy and toxicity of cabazitaxel for treatment of docetaxel-resistant prostate cancer correlating with the initial doses in Japanese patients. BMC Cancer 19, 156 (2019). https://doi.org/10.1186/s12885-019-5342-9

Vu, Blas, Lanni, Gustafson, & Krauss. (2017). Cost effectiveness of prostate boost with high dose rate brachytherapy versus intensity modulated radiation therapy in the treatment of intermediate-high risk prostate cancer. *International Journal of Radiation Oncology*, *Biology, Physics*, 99(2), S37-S38.

Warde, P., Mason, Ding, Kirkbride, Brundage, Cowan, . . . Parulekar. (2011). Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: A randomised, phase 3 trial. The Lancet, 378(9809), 2104-211

DECISION SCIENCES INSTITUTE

Antecedences of Impulsive Donation Behavior on Social Media: Perspectives of Campaign Characteristics and Social Capital

Thi Hong Nhung Nguyen
Department of Business Administration, College of Management
National Cheng Kung University
R48077012@mail.ncku.edu.tw

Ching-Ying Huang
Department of Business Administration, College of Management
National Cheng Kung University
isastat@mail.ncku.edu.tw

ABSTRACT

The rise of social platforms has facilitated the popularity of online impulsive donation. From perspective of non-profits, understanding the impulsive donor's behaviors is imperative to maximize fund-raising goals. Drawing on perspectives of charitable campaign characteristics and social capital theory, this study is primarily to investigate mechanisms and antecedences that drive user's impulsive donation on social media context.Three key antecedences of user's impulsive donation behaviors including: perceived trust on charities, attitude toward charitable campaign and personal impulsiveness were empirically validated**.** Next, the Elaboration Likelihood Model is first-ever extended in the non-profit literature to further explain the changes in an impulsive donor's cognitive persuation**.** This research has also confirmed the essential roles of social experiences via SNSs in driving user's online impulsive donation. The current paper might helps the non-profits on how to cultivate campaign features and develop marketing strategies to persuade and encourage users an urge to donate on social media.

KEY WORDS: Non-profits, Impulsive donation, Charitiable campaign, Social media

INTRODUCTION

Public donation to charities plays a crucial part in modern communities. In particular, the growth of charitable campaigns on social platforms has considerably contributed to the popularity of online giving. As the fact that a significant proportion of charitable donations comes from impulse donors over times (Bennett, 2009). The impulsive donation is understood as a heuristic unplanned giving which is usually stimulated by environmental triggers and fleetly made at a spur-of the-moment. Several charitable campaigns promoted on social media, rely much on donors making fast and impulsive decisions. In other word, the sort of urge to donate impulsively is crucial to maximize such potential sources of charitable fund. Hence, understanding the impulsive giving behavior among online users is really imperative to the success of every charitable campaign or fundraising projects. To our best knowledge, the number of research studying on impulsive philanthrophy is very limited in literature. A previous study by (Bennett, 2009) as the foremost among little research concludes that socio-demographic characteristics; impulsiveness trait; negative attitude toward impulsive behavior; donation history; prior knowledge of hospices; subjective norms and personal involvement are key determinants predicting impulsive giving decision taken by browsers of charity websites. Meanwhile, social network platforms are distinguished to the websites by social experiences and social resources generated via interactive environment (Li, Guan, Hou, & Chong, 2019; L. Liu, Suh, & Wagner, 2017). Notably, these salient discrepancies may address different environmental stimuli that influence on donor's decision making. Therefore, it is necessary to dig deeply into exploring impulsive donation

on social media. This empirical study attempts to fill this knowledge gap about underlying antecedences of impulsive donation behavior. We examined impulsive donation behavior implicitly in the context of monetary donations on social media.

A systematic approach should be acquired to investigate these antecedences under both cognitive and affective facets of impulsive donation-making decisions. Three major determinants of online impulsive donation are emphasized on perceived trust; attitude toward charitable campaigns and empathy. First, perceived trust - public confidence in charities, is indispensable in faciliating online donation transactions. Trust is transfered on charities as intermediaries between donors who are urged to give to beneficaries who are strangers living from a far distance. Second, an attitudinal measure of charitable campaign is firstly developed to demonstrate user's mental and emotional expressions toward a charitable campaign. A favor attitude toward charitable campaigns will convey people supportive responses to a chartitable cause (i.e., impulsively giving). Third, empathy is widely highlighted as a prominent element of intention to donate. Donors who experience the affective state of the other's distressed circumtances tend to provoke a stronger desire to helps donatees immediately. In addition, impulsiveness trait is incoporated in the model since it is highly evidenced a linear association with impulsive decisions.

Furthermore, this study examines the impacts of charitable campaigns on online user'perspective and impulsive donation. Campaign's characteristics and social capitals generated on social media can be considered as critical environmental stimuli. Social capital elicits collective actions as philanthropy, its underlying mechanism on impulsive donation yet has not been highlighted. Under a perspective of campaign's characteristics, the success of charitable campaigns on social media is framed to the context of message persuasion. Campaign's messages driven by non-profits' objectives related to raising awareness; to convince potential donors to involve different persuasive efforts to influence online users. Nevertheless, the current literature have not empirically investigated the impacts of these environmental cues on user's cognitive process in changing attitudes, consequently, predicting behavioral responses. Motivated by these research gaps, we attempted to integrate both attributes of social capital and characteristics of campaigns as stimuli of online impulsive donation. Drawing on the Elaboration likelihood model (ELM), two research questions are clarified:(1) what are antecedences of the persuasion mechanism that influences user's confidence in charities and attitude changes in response to campaign's stimuli, and (2) how different persuasion routes induced by charitable campaigns on social media is characterized. Besides, grounded in the social capital theory, this research aims to investigate how social experiences and social values on social media affect online impulsive donors's decision.

There are several contributions addressed in this empirical study. Theoretically, it is considered as the first-ever in-depth study drawing a comprehensive picture of online impulsive donation in the context of social media. Framed by the S (stimuli)- O (organism)- R(response) paradigm, we conduct a systematic investigation on the underlying mechanism of environmental stimuli from social experiences and characteristics charitable campaigns on impulsive donation intention through perceived trust, attitude toward campaign and empathy. Practically, the findings guide the non-profits how to market campaign incentives and marketing strategies on social media to reach larger potential donors and accomplish charity goals. The study followed a quantitative approach, begins reviewing relevant literature and develop hypotheses to test, next the research methodology is described and is accompanied by empirical results. Finally, the research finding discussion and implications are concluded.

LITERATURE REVIEW

Social media matter charities

Non-profits such as non-governmental organizations and charitable organizations are on a mission to make positive changes in the world. Charities involve to relief victims of natural disasters and social causes, aids to enhance education; health and life quality of the poor; disadvantaged people and homeless children…etc. The world thus becomes a better place thank to human compassion and philanthropy. Every non-profit has a unique story that exhibits social concerns and inspires pro-social behaviors. Over past decades, in order to call for donation, charity organizations spread out messages mainly through traditional advertising vehicles (e-mails, pod-casts,radios, newspapers, TVs...). Thus far, these methods seem getting less effective in response to the non-profit marketing's requirements, whilst charities are even facing a much more competitive environment. Advantageously, the radical evolution of information technology advances has extremely offerred a wide range of opportunities that allow non-profits to harness civic engagement; to reach online users community as potential donors; and to faciliate charitable campaigns's objectives (Bhati & McDonnell, 2020; Carboni & Maxwell, 2015; Guo & Saxton, 2018). Social media bridge people to people; bridge charities to people as well (Farrow & Yuan, 2011). The vast adoption of social network sites among users has been incredibly affecting how individuals socialize; exchange information; and also make donations. Finding has shown that people are more and more willing to donate online across the world (Ahn, Sura, & An, 2018).

In the rapid evolution of ITC eras, the far-reaching connections of social offer a window of opportunities that allow worldwide charities to connect their dedicated volunteer communities and accomplish fundraising goals (Ahn, Sura, & An, 2018; Bhati & McDonnell, 2020). By 2019, the global social media users comprises one of third world population at all ages (estimated 3.48 billion people) (Digital, 2019). The vast adoption of these online networks has been extremely changing human lifestyles by how people socialize; acquire information; and also make donations. Prior studies imply that an online donation platform has a great impact on donors enagement process (Carboni & Maxwell, 2015; Waters, Burnett, Lamm, & Lucas, 2009). Recent survey has indicated a high preference for online donation among people across the world, among those, millennials are most inspired to giving by social media. Statistics present that mong 55% of people who engage with non-profits on social media end up supportive responses, 59% donated money, 53% volunteer. In addition, non-profits also more acceleratedly engage in online campaign strategies, for illustration, 95% of worldwide non-profits have a Facebook page with average 1.2 posts shared daily; 83% are active on Twitter with average 5.3 tweets updated per day; 40% have a Instagram account (NPsource, 2018).

By taking advantages of social media, non-profits can do better at fostering its popularity and fundraising regardless of whether they are large, medium or small-sized. For example, with a 10% increase in the number of "likes" on charity campaign's Facebook posts is allied with a 1% increase gained in the number of donations (Bhati & McDonnell, 2020). From a recent study examining on how non-profits faciliate social media to draw attention of donors, the findings indicate that higher frequency of posting news and number of conversations with online users would strongly lead to higher civic advocacy (Guo & Saxton, 2018). Thus far, embracing the immediacy of social media as powerful communication channels is an essential cost-effective strategy for non-profits to reach more potential donors and expand its influence across geographical boundaries. Social networking sites can be an ideal vehicle to reach a large pool of donors if charities understand how external factors affect to user's cognitive and affective responses.

Online impulsive donation behavior

Social behavior is a joint function of two systems processing behavioral decisions, in which the reflective system elicits behavior through knowledge about facts and values, whereas the impulsive system emphasizes on stimuli associations and rules of thumb motivations (Strack & Deutsch, 2004). Similarly, in the field of prosocial behavior such as philanthrophy, givers also donate through two distinct ways: a deliberation process and an impulsive

process. The deliberative philanthrophy is a thoughful planned donation. Deliberative donors often scruntinize relative information and data to evaluate non-profit performance and their use of money in helping charity recipients. It means that these donors want to make the reflection for ensuring the efficacy of donation before giving it away. In constrast, the impulsive philanthrophy is a heuristic unplanned donation. Impulsive donors are more easily affected by an attractiveness and response spontaneously to feelings of generosity. Impulsive donation decisions are usually emanated from environmental stimuli and fleetly made at a spur-of the-moment. These acts of giving are more likely induced emotional satisfaction and cognitive reaction (Hausman, 2000). Many scholars have attempted to link various types of impulsive behaviors, such as impulsive buying and impulsive giving, then concluded the similarities of both. Over past decades, several studies determined that impulse buying encompasses hedonic or affective components (Hausman, 2000; Rook & Fisher, 1995). Unlike to impulsive buying, impulsive donation is beyond internal compassion that can bring about benefits to well-being of not only givers but also the whole community.

Notably, the rise of social platforms has facilitated the popularity of online impulsive philanthrophy. A significant portion of charitable donations comes from impulse donors over times (Bennett, 2009). From perspective of non-profits, understanding the impulsive giving behavior among online users is really imperative to their success. Because the sort of quick donation is important to maximize such potential sources of fund. The previous study by (Bennett, 2009) suggested a number of key determinants predicting impulsive giving decision to hospice organizations taken by browsers of charity websites. They are socio-demographic characteristics; impulsiveness trait; negative attitude toward impulsive behavior; donation history; prior knowledge of hospices; subjective norms and personal involvement. Howerver, to extend the online donation context to social media with its distinctive characteristics, it is required an in-depth investigation on determinants and underlying mechanism that influence user's impulsive donation-making decisions.

Elaboration likelihood model (ELM) of persuasion

The Elaboration likelihood model (ELM) is a persuasion theory rooted in social psychology. The tenet of the ELM is to describe attitude formation routes. The ELM specifically explains how the influence process shapes an individual's perception and, subsequently, behavior when he or she exposes to messages (Petty & Cacioppo, 1986). The ELM identifies two alternative routes to persuation, the central route and the peripheral route, based on different sorts of information processed by an individual (e.g., issue-relevant arguments or explicit cues). Also, the dual-process of persuasion further specifies condition under which an individual would be more influenced by one route rather than the other (Bhattacherjee & Sanford, 2006). These two routes are characterized by degree of elaboration- the extent to which a person cognitively processes an argument. The central route to persuation is manifested when the degree of elaboration is high, in this case, the recipient primarily bases on critical thinking to scrutinize the issue-related arguments and comprehensive considerations. In contrast, when degree of elaboration is low, the peripheral route to persuation is manifested as the recipient processes the message based on relatively simple cues without requiring a deep analysis of the issue-related arguments. The degree of elaboration varies with a person's motivation and cognitive ability (Petty & Cacioppo, 1986). It means that the person following the central route spends more cognitive efforts (deeper processing and time-consuming) than those following the peripheral route. However, these two mechanisms are both crucial to a person's attitude formation.

This study examines the impacts of charitable campaign's characteristics on online user' perception and impulsive donation. Charitable campaign information on social media can be considered as a stimuli-based decision-making environment. Charity project, charity organization are found out as the measure of the general attitude toward online donation (Treiblmaier and Pollach, 2006); predict intention to donate (Reddick and Ponomariov, 2012). In fact, every charity has a unique story about the organization's work and

beneficiary circumtances to share and connect with the online community via feeds, posts, tweets, images and videos on its fanpage. These generated contents may involve different persuasive efforts to influence online audiences. Particularly, charitable campaign's contents are driven by non-profits' objectives related to raising awareness; to convince potential donors; to encourage the public trust, thereby, can be regarded as persuasive messages. The success of charitable campaign's contents on social media is addressed to the context of persuasion. In literature, charitable campaign's characteristics have been examined as stimuli of the person's intention to donate. For instance, (L. Liu et al., 2017) suggested three elements: reputation of initiator; project popularity and content quality as features of crowfunding project that have positive impacts on online user's donation to micro charities. Howerver, these findings have not clarified yet how the depth of a person's cognitive process different persuasive stimuli is, and also not explained whether a person is more influenced by issue-relevant arguments or explicit cues. Therefore, in this paper, the ELM model is adopted at the aim to investigate (1) how the persuasion mechanism induced by features of charitable campaigns on social media is characterized and (2) what are the determinants of the persuasion mechanism that influences potential donors's impulsive donation. We propose the stimuli of audience's attitudes and perceived trust toward charity campaigns on social media through central route is content quality; and through peripheral route are sources of credibility and campaign popularity. We hope that by applying the ELM framework in explaining the underlying mechanism of a person's perceieved trust and attitude formation toward charitable campaigns on social media, it would further enrich the current literature with a deeper understanding of impulsive donor's decision-making process.

Social capital theory

The heart of the Social Capital Theory is embodied in productive resources which are leveraged from the network of social relationships. Social capital incorporates various aspects of a social context that facilitate member's subsequent behaviors. Generally, these facets are classified into three main dimensions: structural, relational, and cognitive (Nahapiet & Ghoshal, 1998). Structural dimension is embedded in the connectedness of the social system with a net of ties. The relational dimension is embedded in personal relationships developed through a history of interactions drived by social motives (approval, prestige, sociability). The cognitive dimension exhibits shared interpretations and paradigms of meaning among members. According to (Ellison, Steinfield, & Lampe, 2007), social capitals can be generated and mobilized via social relations within a computer-mediated communication environment.

In non-profit sector, social capital is examined as a crucial driver of an individual' charitable giving (Y. Kim & Lee, 2014; Wang & Graddy, 2008). Characteristics of social network plaforms have been largely recognized as external stimulus that positively affect user's awareness and behaviors toward philanthrophy (Ahn, Sura, & An, 2018; Y. Kim & Lee, 2014; Li et al., 2019; Waters, Burnett, Lamm, & Lucas, 2009). The majority of studies have extensively examined how social media matter the rise of online giving and how non-profits take advantages of social network sites to boost up charity performance. (Li et al., 2019) stated that users involving in charitable campaigns online do not only simply conduct online giving transactions, but also experience a social intimacy by posting and sharing stories, information (posts), photos, videos regarding charitable campaigns with others. These interactions among people help them be more active in pro-social behaviors. Furthermore, (Adler & Seok-Woo, 2002) implied that the most salient benefit that social capital brings is solidarity including civic engagement. In general, a social platform plays a crucial role in generating social experiences that connect the whole community, build trusts among actors, faciliate the spill-over effects of member's involvement in social causes. These characteristics of charitable campaigns on social media makes them distinguished to other charitable campaigns launched in traditional contexts.

A prior study investigating donor's impulsive decision in the context of websites suggested subjective norms is critical to donor's decisions (Bennett, 2009). Meanwhile, due to differences between social media and website contexts, it is imperative to understand user's decision-making process in the field of online implusive donation. To our best of knowledge, there has no research paper that studies the impacts of social capital on online user's impulsive donation toward charitable campaigns on social media. Hence, we hope that the finding of the current study will shed a light in philanthrophy marketing literature by exploring how social experiences on social network platforms affect online donors's perception and the urge to donate impulsively. Moreover, the underlying mechanism of social capital's impacts on perception and behavior of online donors (i.e., impulsive donors) in response to charitable campaigns on social media is still underscored. A recent study has examined the influence of social capitals on donor's empathy and intention to donate charitable crowdfunding by three variables: social influence, social interaction, social proximity (Li et al., 2019). Howerver, they have ignored the significance of cognitive capital in generating mutual understandings and empathic concerns among online users, consequently, it might positively affects donation decision. Therefore, so as to comprehensively investigate influences environmental stimuli on donors'repsonse, in this study, we integrated three structural, relational, cognitive dimemensions of social capital which are respectively manifested by social bonds (interaction ties), subjective norm and shared visions.

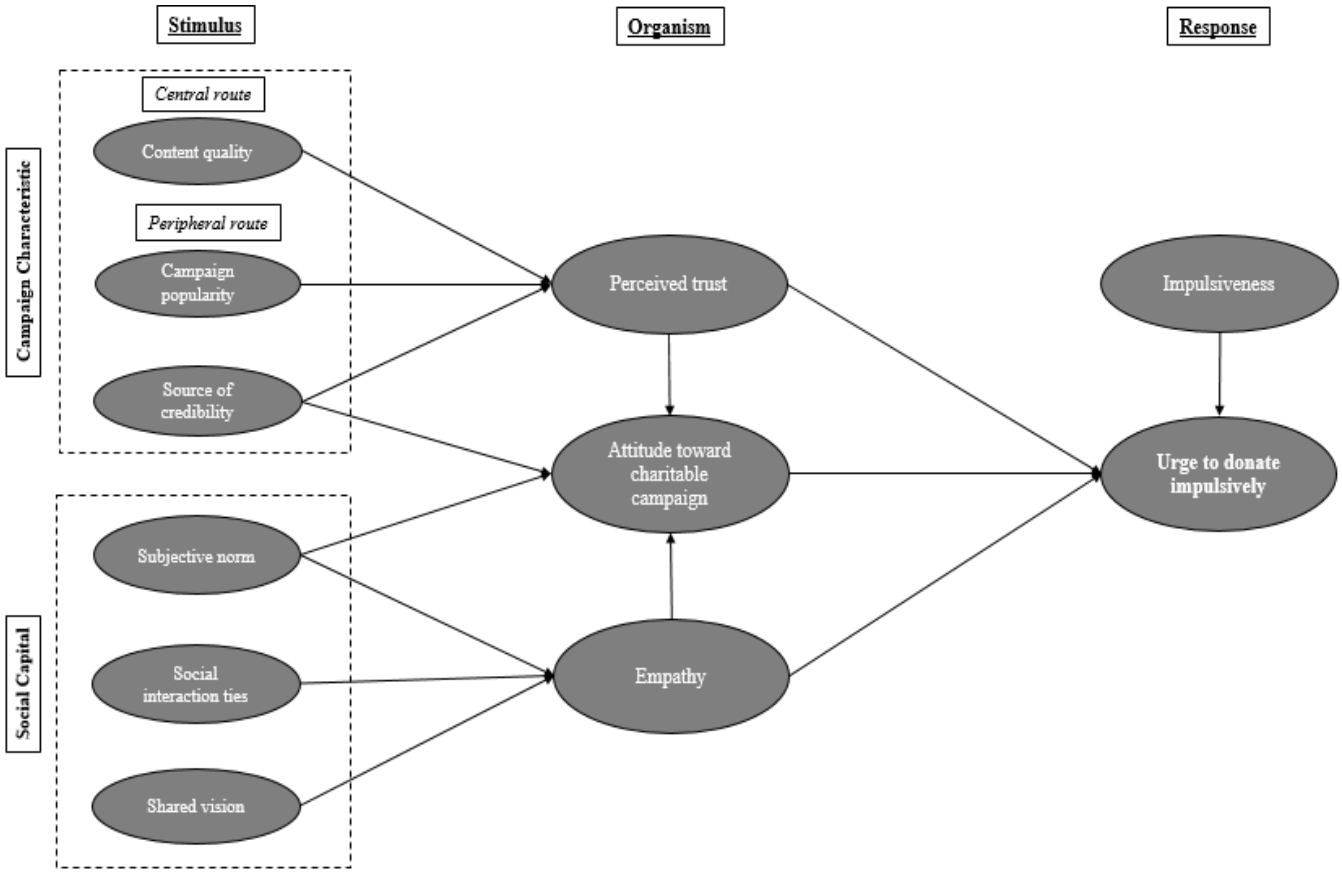
HYPOTHESES DEVELOPMENT

S-O-R theoretical framework

In literature, the stimuli-organism-response (S-O-R) model has been extensively adopted to understand online user behavior. S-O-R model posits that stimuli perceived in the environment will stimulate an individual's internal evaluations as organism factors, which subsequently leads to possible responses. By applying the S-O-R paradigm, this empirical study examines how environmental cues- campaign's characteristics and social capital – stimulate donor's perceived trust; attitude toward charitable campaign and empathy, which in turn impact the impulsive donation intention toward charitable campaigns on social media context.

Figure 1: Research framework

Antecedences of Impulsive Donation Behavior on Social Media



Response factor: *Urge to donate impulsively and impulsiveness*

In literature, the concept of urge to buy impulsively is defined as a state of desire triggered by an object in the environment that elicits a sudden, powerful, and persistent feeling to buy something immediately (Y. Liu, Li, & Hu, 2013). The urge to buy impulsively is claimed to be a reliable antecedent of actual impulse buying behavior (Liu, Li, & Hu, 2013; Parboteeah, Valacich, & Wells, 2009). Similarly, in the sphere of charities, impulsive donation behaviors are often triggered explicitly by environmental stimuli and fleetly made at a spur-of the-moment without pre-planned considerations. The urge to donate impulsively is also a vital predictor of actual impulsive donation. In this paper, we modified the urge to donate impulsively as the state of desire stimulated by environmetal triggers, that makes online users prone to donate spontaneously. We proposed two main sources of stimulus: appeals from charity campaigns's characteristics and social captials generated among users on social media. By adopting a systematic framework of S-O-R, we also suggested three key antecedences of urge to donate impulsively: perceieved trust in charitable campaign, donor's attitudes and empathy. The urge to donate impulsively is further examined under effects of personal impulsiveness.

Impulsiveness plays a crucial role in both offline and online consumer decision making. Impulsiveness is a kind of socio- psychology trait that is characterized by four main dimensions: absence of planning, elicited by a stimulus, spur-of-moment decision, and cause emotional or cognitive reaction (Piron, 1991). When an individual is highly impulsive, he or she tends to take action immediately without pre-planning. The linear association between personal impulsiveness and impulsive decision has been evidenced in previous studies (Bennett, 2009; Rook & Fisher, 1995). The findings stated that personal impulsiveness significantly encourages an individual to impulsively giving away. Therefore, in this study, we would like to further confirm that greater personal impulsivenes might predict higher urge to donate impulsively to charity campaigns on social media. When he or

she is in high impulsiveness, he or she will spontaneously execute an urge to donate for the charity recipients to fulfill desires to help. The following hypothesis was developed:

Hypothesis 1: Higher personal impulsiveness leads to higher urge to donate impulsively

Organism factors

* ***Attitude toward charitable campaign***

Attitude encompasses both mental and emotional expressions towards an object, which in turn shapes the individual's perception and guides behaviors (Ajzen & Fishbein, 2005). Within a long history of socio-psychological researches on attitude, most scholars primarily implicate that a person's attitude reflects his or her evaluations regarding to a certain target or behavior that he or she experienced with. Commonly, attitude is identified as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken 1992). In their comprehensive review, Eagly & Chaiken have also explained how attitude can significantly explicit cognitive, affective and behavioral responses. The relationship between attitude and behavior shows a high consistency in both theoretical and practical fingdings. In the literature of philanthrophy marketing, (Webb, Green, & Brashear, 2000) clarified two related but distinct attitudinal determinants of donation behavior are attitude toward helping others (*attitude toward behavior*) and attitude toward charitable organizations (*attitude toward target*). Findings indicate that attitude toward helping others and toward charitable organizations are both closely associated with the breadth of donation behavior, however, only attitude toward charitable organizations significantly explains to donor's magnitude of donation in term of how much money actually donated (Webb et al., 2000). In the current study, we develop one more specific attitudinal measure named "*attitude toward charitable campaign*". Charitable campaign (project) is a planned set of activities performed by a charitable organization over a period of time in order to achieve primary objectives related to philanthrophy and social well-being (i.e., helping strangers, relief victims, enhance the welfare of the needy people…). Based on the nature of charitable campaign, we found that attitude toward charitable campaign might be integrated by both individual's attitude toward helping behviors and attitude toward a target charity. We adopt this concept to develop an attitudinal measure of charitable campaign as online user's psychological tendency expressed by personal judgements, evaluations toward a charitable campaign with some degree of favor or disfavor.

In literature, a number of studies show that attitude toward donation; charity in general positively impact on user's intention to donate (Ahn et al., 2018; Ranganathan & Henley, 2008). Online users further prone to make a donation inhesitately and impulsively because they think that their acts of giving are essentially devoted to positive impacts on the other's difficulties. In general, positive attitude toward giving will convey positive attitude toward charitable campaigns, which in turn leads people to behave in a positive maner (Ranganathan & Henley, 2008). When a person feels good about donation, he/she will be more urged to donate for self-rewards and social needs. Previous evidence implies that unfavourable attitudes towards impulsive donation exerted a significantly negative impact on impulsive donate behavior, and vice versa (Bennett, 2009). In addition, we suggest that it is essential to investigate the influence of attitude toward charitable campaign on impulsive donation behavior. Therefore, in order to examine donor's attitude toward charitable campaign on social media, we attempt to test the positive link between attitude and urge to donate impulsively as followed:

Hypothesis 2: Positive attitude toward charitable campaign leads to higher urge to donate impulsively

* ***Percieved trust***

In this study, trust refers to the extent to which people are confidently believed in charities (D. J. Kim, Ferrin, & Rao, 2008). Notably, public confidence in charities is indispensable for their success in reaching out potential donors and fund-rasing goals. When reviewing the literature, we have found that there are three main reasons determining the significance of public trust to charitable giving, especially to social media donation. First, embedded in group membership, trust is regarded as one of the most imperative attitudinal factors promoting civic engagement (Taniguchi & Marshall, 2014). Once trust is increased, the social bonds are also strengthen, which in turn affect positively to individual's perceptions toward charitable giving. Second, IS researchers have found that trust plays a vital part in faciliating online transactions where exist information asymmetry (Chiu, Hsu, & Wang, 2006). Within the sphere of non-profit envirnonment, giving to charities involves helping the strangers who are living from a far distance. Charities as intermediaries are in charge of transfering charitable gifts from donors to beneficaries. In these cases, trust and positive evaluations toward charities are essential when donors are urged to do such good things for grant recipients (Wiepking & Maas, 2009). According to a recent survey examining public trust in charities in UK, the majority of respondents (37%) advocated that transparency about the fund allocation made it to the end cause is the most critical to their confidence in charities (Populus, 2018). Third, donation is an act of volunteery liberality whereby a donor sacrifices of a thing in favor of another without expecting anything in return. Hence, donors tend to lay down their trustworthiness in charities (understood as "*trust transfer*") and empower them to accomplish social causes on behalf. Researchers may all agree that perceived trust on charities and their campaigns does matter not only donor's perception but also donation behaviors. In fact, evidence has revealed that those who trust in charities are more likely to give and make repeat donations than those who do not (Populus, 2018). (Li et al., 2019) examine donor behavior to charitable crowdingfunds has further explained that potential funders often base on their own judgement of the project's credibility to decide whether donate or not. There is a large number of research exploring the positive relationship between trust and donor's attitude and willingness to donate. However, none of them has specifically focused on impulsive donation behavior. Thus, we attempt to fill the research gap by developing a hypothesis which demonstrates the positive association between trust and attitude; and impulsive donation as follow:

Hypothesis 3: Percieved trust has a positive impact on attitude toward charitable campaign

Hypothesis 4: Higher perceieved trust leads to higher urge to donate impulsively

* ***Empathy***

In a large body of literature, empathy is highlighted as a salient determinant of prosocial behavior. The conception of empathy is identifined by three core dimensions: emotional, cognitive and behavorial. According to (Eisenberg & Miller, 1987), empathy respesents the affective state that "stems from the apprehension of another's emotional state or condition". Also, in non-profit sector, empathy is often studied as the extent to which an individual feels compassion for the those seeking helps. A plenty of evidences have implied that empathic concern significantly inhibits the donation behaviors (Verhaert & Van den Poel, 2011). People are in higher degree of empathy are known to be more geogenerous givers, merely because they can better figure out potential donatee's needs (Wiepking & Maas, 2009). Non-profits's efforts have ultilized visual effects of such appealing marketing approach through emotive pictures, video, live-streaming in order to draw concerns of online users who shared compassions to social causes. A finding suggested that web pages which are designed with emotive-orientations will lead to higher likelihood of impulsive donations (Bennett, 2009). In the current study, we proposed that those who have a positive attitude toward donation are those who always have high feeling of compassion and empathic concern to the less fortunate. Since, when they expose to the other's circumtances and experience the emotional state of the other. They tend to elicit a strong desire to help and positive perceptions toward charitable campaign, which in turn makes them more inclined to

give immediately. These sort of quickly supportive responses will significantly generate emotional benefits to donors themselves (i.e., feeling of good, out of distress, self-reward). Thereby, there are two hypotheses developed to investigate the positive relatonship between empathy and attitude toward charitable campaign; empathy and urger to donate impulsively in the context of charitable giving on social media.

Hypothesis 5: Empathy has a positive impact on attitude toward charitable campaign

Hypothesis 6: Higher empathy leads to higher urge to donate impulsively

Impacts of charitable campaign's stimuli

* ***Central route factor:*** *Content quality*

The ELM model describes the central route with a high degree of elaboration in processing messages (Petty & Cacioppo, 1986). In the central route, the person always evaluates the received information by scrutinizing the pros and cons of its conveyed values. The recipient perceives messages as helpful, persuasive and trustworthy if these issue-relevant arguments are in high quality. Argument quality echoes information reliability and the persuasive strength (Bhattacherjee & Sanford, 2006; Stuart, Teng, Khong, Goh, & Chong, 2014). Therefore, argument quality is the most critical dimension of the central route (Teng, Khong, Chong, & Lin, 2017). In this study, the persuasive strength of charitable campaigns depends primarily on contents posted on social network platforms by charities, it is highly related to content quality.

In the context of online donation, content quality is defined as the degree to what a user considers that the charitable campaign-related information is of a high quality (L. Liu et al., 2017). The IS literature posits that information quality encompasses information accurarcy; relevance; completeness; ease to understand, format and updatedness (Xu, Benbasat, & Cenfetelli, 2013). Charitable campaign-related information posted on social media are generated under forms of text, images, video…that comprise different persuasive efforts conveyed to recipients. These users who pay attention to charitable campaigns will be more motivated, spend more time and cognitive efforts to judge the quality of contents and accquire detailed information before forming attitudes. The quality of contents delivered by charities further enables donors to recognize which charity is of high trustworthiness and compatible to the others with low credibility (L. Liu et al., 2017). Hence, we propose that charitable campaign's content quality positively affects the user's perceived trust toward a charity.

Hypothesis 7: Content quality has a positive impact on perceived trust

* ***Peripheral routes factors:***

The ELM model describes the peripheral factors as stimuli in the persuasion context that can shape a person's attitude without scrutinizing the information quality or message arguments (Petty & Cacioppo, 1986). In the context of online donation, we proposed that the peripheral stimuli can be heuristic cues of persuasion including sources of credibility and source likeability (attractiveness of campaign) (Wiener & Mowen, 1986). Together with the use of social media as a vehicle to spread charitable campaign's messages to larger donors, the significant influence of source of credibility and campaign popularity can be played out to convince online users to involve in philanthrophy giving.

Sources of credibility

Bhattacherjee & Sanford (2006) suggested that the peripheral route of persuation is mainly resprented by sources of credibility. Sources of credibility refers to the extent to which a meassage source is perceived to be credible by message recipients. It does mean that in order to persuade the message recipients, should let them know trustworthy sources of information. Existing literature reveals that information is doubted less since sources of

credibility deprives from (1) expertise which represents a person's ability to provide the correct information in terms of knowledgeable, experienced, competent (2) trustworthiness which indicates the extent to which a person believed a communicator would tell the truth (McGinnies & Ward, 1980). Previous studies demonstrate that when a person are highly elaborated with received messages, source of credibility has little impact on attitude formation because she or he will analyze the argument rather than evaluate peripheral cues. Dissimilarly, when a person is not involved in a message, source of credibility has been addressed to be an essential predictor of attitude changes (Chaiken & Maheswaran, 1994; Petty & Cacioppo, 1986; Teng et al., 2017). A study by (Brinol, Petty, & Tormala, 2004) further implies that source of credibility positively influences consumers's attitudes by making them be more confident in thoughts. In the context of purchasing donation, a finding reveals that consumers' willingness to donate in retail stores is inclinated when they perceive the retailers are trustworthy and credible (Savas, 2016). In the sphere of online donation, charitable campaign's messages and information are spread out on social media by multiple actors: expertises in charity, celebrities, influencers, supporters, public agents…who are widely recognized as knowledgable and credible communicators. Online user's perception towards charitable campaigns is also be shaped by source credibility. When users peceived the sources of message are creditable, their trusts on charities are thereby increased (L. Liu et al., 2017). Therefore, we propose the positive link between source credibility and user's attitude and perceived trust as follow:

Hypothesis 8: Sources of credibility has a positive impact on attitude toward charitable campaign

Hypothesis 9: Sources of credibility has a positive impact on perceived trust

Campaign popularity

The attractiveness of a charitable campaign is considered as a source likeability for a user to process messages through the peripheral path. It is characterized by the poularity of campaign which refers to the quantity of shares (retweets), comments and likes to a charitable campaign launched on social media. Apparently, when users observe charitable campaign's posts with a large number of shares, likes and comments, they will perceive it as popular. They also believe that the campaign is on and off drawing a great concern among users community, then feel obliged to join the crowd. Source attractiveness such as post popularity is critical to the online user's attitude formation (Teng et al., 2017). This phenomenon can be explained by the *bandwagon effect* which demonstrates a simple logic as "if others think that this is a good story, then I should think so too" (Sundar, 2008). Bandwagon heuristic cues are considered as the most prevailing shortcuts for cognitively processing online messages, with the results upon observing other users' rating actions (Lin, Spence, & Lachlan, 2016; Sundar, 2008). In e-marketing literatue, researches have shown that number of likes and friend's likes on Facebook brand pages are heuristic cues that positively impact on online buyer's attitude toward brand; brand trust; brand involvement; and purchasing decisions (Phua & Ahn, 2016; Sabate, Berbegal-Mirabent, Cañabate, & Lebherz, 2014). In the field of online donation, popularity of charitable project works as a signal that stimulates potential donor' perceived credibility on that project (L. Liu et al., 2017). Based on above arguments, we proposed that the popularity of charitable campaign on social media is considered as a vital environmental trigger that can enhance online donors' trust on charities. Hence, the below hypothesis was developed:

Hypothesis 10: Campaign popularity has a positive impact on perceived trust

Impacts of social capitals stimuli

* ***Subjective norms***

Interactive relationships are embedded in a social context that can affect an individual's affective and cognitive reactions. In the social capital theory, subjective norm is a well-

defined relational dimension which embodies how much judgements of families, friends, colleagues and relatives are important to an individual's beliefs, attitudes and subsequent decisions (Fishbein & Ajzen, 1975; Nahapiet & Ghoshal, 1998). Therefore, a behavorial response is performed under a social normative pressure when an online user percieved that other peoples who matter to him/her would approve his/her decision. In the context of online donation, subjective norm is examined as a significant factor predicting user's giving decision (Bennett, 2009; Li et al., 2019). Since charity donation is a pro-social behavior which should be encouraged by civic engagement in order to bring about beneficial outcomes for the whole society. People always highly appreciate and expect others to show compassion for people who seeking helps. These social expectations may shape individual's perception toward making a donation for charity project donatees. Li et al, (2019) further explain that social capital generated from influencial people's expectations also has a great impact on an online user's empathic concerns toward donatee, which in turn encourages intention to make a donation. Hence, in the current paper, we also proposed that online user's attitude and empathy would be more affected by subjective norms toward making an urge to donate charity campaigns. The following hypotheses are developed:

Hypothesis 11: subjective norms have positively impacted on donor's attitude toward charity campaign.

Hypothesis 12: subjective norms have positively impacted on donor's empathy.

* ***Social interaction ties***

In the context of a community network, social bonds among people play an important role in establishing strong emotional bonds that allow them to comprehend what others think and feel. A sizable body of research has supported that online community users are likely to evoke emotional attachment or emotional closeness (e.g., empathy, share feelings, emotional supports, caring…) with others through experiences of appropriate interpersonal relationships (Brodie, Ilic, Juric, & Hollebeek, 2013; Ren et al., 2012; Williams, 2006). Since perceptions of social ties increases credibility assessments, bring in consensus from others (Ren et al., 2012). Based on above arguments, we also can noted that when people are in strong-tie relationships (as frequent and intensive interactions) with family; friends; and society, the more understandings they shared and the higher empathic concerns they convey to gain insight of others circumtances. Several findings also proved that close-knit human ties are essential to motivate prosocial behavior such as charitable giving (Adloff, 2009; Farrow & Yuan, 2011; Wang & Graddy, 2008). Social media can generate social capitals by making users feel connected to the online community, thereby, faciliate opportunities for collectivism. Users are easily hearing to charitable giving opportunities through public posts and tweets shared by both charity organizations and community members. Thereby, through compelling interpersonal channels, individuals are exposured more closely with the circumtances of donatee on such virtual environment. In that case, it can be possible for them to raise empathy up towards the less fortunate. Recent studies have shown that online social interactions effectively enable users to feel the emotions of the others, named as empathic emotions (Li et al., 2019; Park & Rhee, 2019). Therefore, we proposed a hypothesis that there exists a positive relationship between social interaction ties and donor's empathy toward a social cause in the context of online donation.

Hypothesis 13: Social interaction ties have positively impacted on donor's empathy.

* ***Shared vision***

(Fischer & Reuber, 2011) state that the interpersonal interactions on social media (i.e., Twitter) can effectively generate cognitive capital for user commmunities. Cognitive dimensions are identified by "attributes like a shared code or a shared paradigm that facilitates a common understanding of collective goals and proper ways of acting" within a

community (Nahapiet & Ghoshal, 1998). Among those, shared vision is a critical to bring members of human networks together under collective goals and aspirations (Cohen, 2001 #2122;). In virtual communities, the common visions and core values shared by members help them see the meaning of collective behavior, which in turn endure benefits of the whole community (Chiu et al., 2006). In a nutshell, shared visions can lead to mutual understandings. These common understandings are more enriched once people have capabilities to walk a mile in the other's shoes. In online communities, donors are also sharing mutual desires to care for those seeking helps. They all visualize to a better world and positive changes in the lives of people. They always provoke empathic concerns to the less fortunate and make efforts to giving. These charity outcomes are further played out when non-profits take advantages of social media in connecting commonly held values among online users. Because when online community's members are exposured to a social cause, they might exhibit emotionally intense feeling of empathy and social supports to fullfill social needs. However, the relationship between cognitive capital (shared visions) and empathy has not been yet explored in previous studies, particulary in the context of online social networks. Therefore, in the current paper, we attempt to fill this gap by examining the role of shared charitable visions in driving donor's empathy toward a social cause on social media. We developed a hypothesis signifying their positive correlation as follow:

Hypothesis 14: Shared visions have positively impacted on donor's empathy

METHODOLOGY

Data collection

This research intentionally approached social media users via online-based surveys. The surveys were releashed and shared widely from a bulk of internet sources including: Facebook, Instagram, Twitter, Amazon M-turk to recruit the data sample. These target respondents are those who are familiar with charitable campaigns on social media. Respondents were asked to recall the most recent charitable campaign they have read or involved, it is not required that they have donated money, just simply have acknowledge about that campaign. Then they are asked to fill out a web-linked questionnaire. To make sure only the right participants can take part in answering the survey, we have two short filtering questions to check whether they have experienced with charitable campaigns on SNSs by asking them list out name of the most recent charitable campaign on social media they knew and its launched period of time. A total of 213 valid samples were consumed for further analyses. The sampling process conducted without missing data.

The socio-demographics of the population were summarized in Table 1. Majority of participants are female (70 %) who are younger generations aged 25 to 30 years old (34.7%), followed by 31 to 35 years old (20.2%). Regard to education and income backgrounds, it found that most have a Bachelor's degree (71.4%) with average monthly earning from USD 1000 to 2000 (39%). Among commonplace SNSs, Facebook users make up the largerst proportion with over 50% of participants who are active almost 3 hours per day. Their donation experiences were also surveyed and it turns out that participants are enthusiatically engaged in social philanthrophy over time with twice donations per year on average (nearly 40%).

Table 1. Socio-demographic characteristics.

Characteristics	Frequency (N=213)	Percentage (100%)
Gender		
Female	149	70.0
Male	64	30.0
Age		

Between 18 and 24	10	4.7
Between 25 and 30	74	34.7
Between 31 and 35	43	20.2
Between 36 and 40	22	10.3
Between 41 and 45	23	10.8
Above 45	41	19.2
Education		
High school or below	15	7.0
Bachelor's degree	152	71.4
Master's degree	42	19.7
Doctoral degree	4	1.9
Monthly income		
Under US $ 1,000	16	7.5
Between US $ 1001 to $ 2,000	83	39.0
Between US $ 2,001 to $ 4,000	78	36.6
Above US $ 4,000	36	16.9
Past donation experiences (frequency in year)		
Almost never	22	10.3
Once	53	24.9
Twice	85	39.9
Three times	23	10.8
More than three times	30	14.1
Social network site (SNS) preferences		
Facebook	107	50.2
Twitter	40	18.8
Instagram	58	27.2
Others	8	3.8
Average day- time spent on SNS		
Less than 1 hour	27	13.8
Between 1 to 3 hours	72	36.9
Between 3 to 5 hours	38	19.5
More than 5 hours	14	7.2

Measurement

A pool of eleven measuring contructs was developed in the proposed research framework (see Table 2). Most of items were adopted and slightly modified from previous studies. A seven-point Likert scale was adopted for measuring all latent items (ranging from 1=strongly disagree to 7=strongly agree). There are totally four reversed-scored items included in a survey, then all items were identically coded for running data analysis.

Table 2. Measure development

Construct	References	Definition	Item	Description
Content quality	(Bhattacherjee & Sanford,	The extent to what an individual believed that	CQ1	Overall, I would give the content quality of the charity campaign high marks

(CQ)	2006; Xu et al., 2013)	the information provided about a charity campaign is of a high quality (e.g., informative, complete, transparent, objective, valuable, up-to-date…)	CQ2	Overall, I would give a high rating in terms of the content quality for the charity campaign.
			CQ3	In general, high-quality content for the charity campaign is provided
Sources of credibility (SOC)	(Bhattacherjee & Sanford, 2006; Sussman & Siegal, 2003)	The extent to what a receipient perceived that the source of charity campaign-relevant information are trustworthy and creditable.	SOC1	The charity campaign information are posted and shared by influencers/ reputed person or creditable organization
			SOC2	The charity organization/individual posted/shared charity campaign information was trustworthy.
			SOC3	The charity organization/individual posted/shared charity campaign information was reliable.
			SOC4	The charity organization/individual posted/shared charity campaign information appeared to be an expert/specialist on this issue.
Campaign popularity (CP)	(Chang, Yu, & Lu, 2015)	The degree to what an individual perceived a charity campaign is popular on social media when they see its post's number of likes/shares/comments.	CP1	I think charity campaign with more people pressing like, sharing, and responding positively to it posts is trustworthy
			CP2	I think charity campaign with more people pressing like, sharing, and respondingpositively to it posts is reliable
			CP3	I think charity campaign posts with more people pressing like, sharing, and responding positively to its posts are believable.
Shared vision (SV)	(Chiu et al., 2006)	An individual's perception of the extent to which members in a social community shared the mutual vision; goal and value about philanthrophy practices	SV1	Members in my online community share the same vision of helping others and solving social causes.
			SV2	Members in my online community share the same goal of making positive changes in society.
			SV3	Members in my online community share the same value that helping others is pleasant.
Subjective norm (SN)	(Bennett, 2009)	The perceived expectations from others that influence an user to make a donation to the charity campaign on social media	SN1	People who matter to me would strongly approve of my making a donation to the charity campaign.
			SN2	Members of my family would strongly approve of my making a donation to the charity campaign.
			SN3	My friends would strongly approve of my making a donation to the charity campaign
Social interaction ties	(Chiu et al., 2006)	The intensity of the social relationships, and the amount of time	SI1	I maintain close social relationships with my online community

(SI)		spent, and frequency of information exchanged among individuals of online communities or between individual and charity organizations	SI2	I spend a lot of time interacting with my online community
			SI3	I have a frequent communication with my online community
Perceived trust (PT)	(D. J. Kim et al., 2008)	The extent to which people are confidently believed in a charity.	PT1	I perceived that the charity to be credible.
			PT2	I perceived that the charity to be believable.
			PT3	I perceived that the charity to be trustworthy.
			PT4	The charity gave me an impression that it keeps promises and commitments by doing that campaign
Empathy (EMP)	(Verhaert & Van den Poel, 2011)	Feelings of sympathy and compassion for distressed others (Davis, 1994)	*How well these following statements describe yourself? (1=not at all; 7= very much)*	
			EM1	I often have tender, concerned feelings for people less fortunate than me
			EM2	Sometimes, I don't feel sorry for other people when they are having problems (reversed)
			EM3	When I see someone being taken advantage of, I feel kind of protective toward them
			EM4	Other people's misfortunate usually do not disturb me a great deal (reversed)
			EM5	When I see someone being treated unfairly, I sometimes don't feel pity for them (reversed)
			EM6	I would describe myself as a pretty soft-hearted person
			EM7	I quite often being touched by things that I see happen
Attitude toward charitable campaign (ATT)	(Eagly & Chaiken 1992)	An individual's psychological tendency expressed by personal judgements, evaluations toward a charitable campaign with some degree of favor or disfavor.	*Let's think about the charitable campaign, how do you feel if you donate to? (1=strongly disagree; 7= strongly agree)*	
			ATT1	Good
			ATT2	Positive
			ATT3	Favorable
			ATT4	Pleasant
Impulsiveness trait (IP)	(Bennett, 2009; Hausman, 2000)	An individual's tendency includes sudden urges to take an unplanned action or take action immediately without evaluation and deliberation of the consequence	IP1	I often do things spontaneously
			IP2	Sometimes I feel like doing things on the spur of the moment
			IP3	I carefully plan the things I do (reversed)
			IP4	"Just do it" describes my attitude to getting things done
			IP5	Sometimes I'm a bit reckless about what I do

Urge to donate impulsively (UDI)	(Y. Liu et al., 2013)	The state when confronting the stimuli in the environment and urge desire to immediately donate to fulfill prosocial need	UDI1	As I exposure to charity campaign posts, I had the urge to donate to fulfill my desire to help.
			UDI2	While reading charity campaign posts, I have a desire to donate that did not pertain to my desire make positive changes.
			UDI3	While reading charity campaign posts, I have the inclination to donate to fulfill my desire to help.

RESULTS

Measurement tests

To test whether the measuring constructs in the developed hypotheses are emprirically valid or not, a confirmatory factor analysis (CFA) was employed by using PLS-Smart 3.0 software. All latent items were framed under reflective types. The remained items exhibited standardized factor loadings that exceeded the minimum value .70 at a significance level of .001. Measurement tests were performed to assess the reliability and validity of the mentioned eleven constructs. The CFA results are reported in table 3. *Reliability* representing the internal consistency of each contruct was evaluated by composite reliability (CR) and Cronbach's alpha. All constructs reached acceptable levels of reliability based on Cronbach's alpha exceeded the threhold value of .70 (Hair Jr, Hult, Ringle, & Sarstedt, 2016). Additionally, the composite reliability are all accepted with values ranged from .814 to .889. These results showed that all constructs were considerably reliable. *Convergent validity* was satisfactionarily assessed by the average extracted variances (AVE) were above .50, suggested that measuring items explained a large portion of the variation and implied a good convergent validity (Anderson & Gerbing, 1988). The results of the Fornell-Larcker criterion (see Table 4) showed all constructs's square root of AVE were larger than correlation among latent variables, indicated an adequate *discriminant validity* (Chin, 1998).

Table 3. Descriptive statistics and PLS-CFA results.

Construct	Item	Mean	S.D	Factor loading	AVE	C.R	Cronbach's alpha
Attitude toward charitable campaign (ATT)	ATT1	5.52	1.17	0.787	0.667	0.889	0.833
	ATT2	5.53	1.32	0.813			
	ATT3	5.36	1.21	0.835			
	ATT4	5.49	1.22	0.830			
Campaign popularity (CP)	CP1	5.45	1.32	0.851	0.675	0.861	0.758
	CP2	5.33	1.31	0.760			
	CP3	5.36	1.24	0.851			
Content quality (CQ)	CQ1	5.52	1.10	0.860	0.664	0.855	0.745
	CQ2	5.32	1.31	0.755			
	CQ3	5.38	1.26	0.826			
Empathy (EM)	EM1	5.42	1.25	0.722	0.527	0.886	0.851
	EM2	5.41	1.33	0.717			
	EM3	5.28	1.35	0.767			
	EM4	5.23	1.35	0.713			

	EM5	5.32	1.34	0.744			
	EM6	5.31	1.32	0.724			
	EM7	5.17	1.40	*deleted*			
Impulsiveness trait (IP)	IP1	5.12	1.54	0.863	0.549	0.853	0.774
	IP2	4.93	1.58	0.794			
	IP3	5.41	1.34	*deleted*			
	IP4	4.94	1.57	0.802			
	IP5	4.63	1.72	0.745			
Perceived trust (PT)	PT1	5.31	1.34	0.801	0.613	0.864	0.790
	PT2	5.48	1.27	0.794			
	PT3	5.29	1.35	0.769			
	PT4	5.35	1.32	0.768			
Social Interaction (SI)	SI1	5.27	1.37	0.851	0.661	0.853	0.741
	SI2	5.22	1.32	0.723			
	SI3	5.16	1.30	0.857			
Subjective norms (SN)	SN1	5.44	1.27	0.795	0.636	0.839	0.714
	SN2	5.49	1.31	0.744			
	SN3	5.43	1.10	0.850			
Source of credibility (SOC)	SOC1	5.40	1.28	0.791	0.588	0.851	0.767
	SOC2	5.44	1.21	0.725			
	SOC3	5.36	1.31	0.807			
	SOC4	5.32	1.28	0.742			
Shared visions (SV)	SV1	5.39	1.26	0.849	0.673	0.860	0.756
	SV2	5.40	1.28	0.760			
	SV3	5.43	1.22	0.849			
Urge to donate impulsively (UDI)	UDI1	5.43	1.31	0.872	0.605	0.814	0.752
	UDI2	4.89	1.60	*deleted*			
	UDI3	5.37	1.28	0.880			

Table 4. Construct correlation matrix.

	ATT	CP	CQ	EM	IP	PT	SI	SN	SOC	SV	UDI
ATT	**0.817**										
CP	0.650	**0.821**									
CQ	0.656	0.782	**0.815**								
EM	0.580	0.488	0.521	**0.726**							
IP	0.392	0.557	0.582	0.299	**0.741**						
PT	0.692	0.828	0.830	0.546	0.504	**0.783**					
SI	0.560	0.632	0.636	0.588	0.560	0.651	**0.813**				
SN	0.686	0.724	0.730	0.629	0.431	0.773	0.685	**0.798**			
SOC	0.622	0.827	0.841	0.530	0.571	0.844	0.622	0.730	**0.767**		

SV	0.626	0.766	0.751	0.616	0.519	0.788	0.697	0.814	0.764	**0.821**	
UDI	0.617	0.770	0.760	0.486	0.635	0.762	0.600	0.706	0.738	0.698	**0.778**

** Diagonal figures are the square root of the extracted average variance*

Common method variance

In addition, since this research was conducted based on self-reported surveys, it is necessary to carry out the common method variance (CMV) tests at the aim to discover possible external factors that affect to the internal validity of measurements. First, Harman's one-factor test claim that the CMV was not exhibited when the highest unrotated factor explained only 42.2 %, less than 50% of the variance (Podsakoff, Mackenzie, Lee, & Podsakoff, 2003). Second, the collinearity assessment approach in SEM-PLS was further applied, results showed that all variance inflation factors (VIF) values of all constructs were not exceeded threshold value of 3.3, indicated that CMB was not a pervasive issue in the current research (Kock, 2015).

Hypotheses testing

Table 5. Path coefficient results

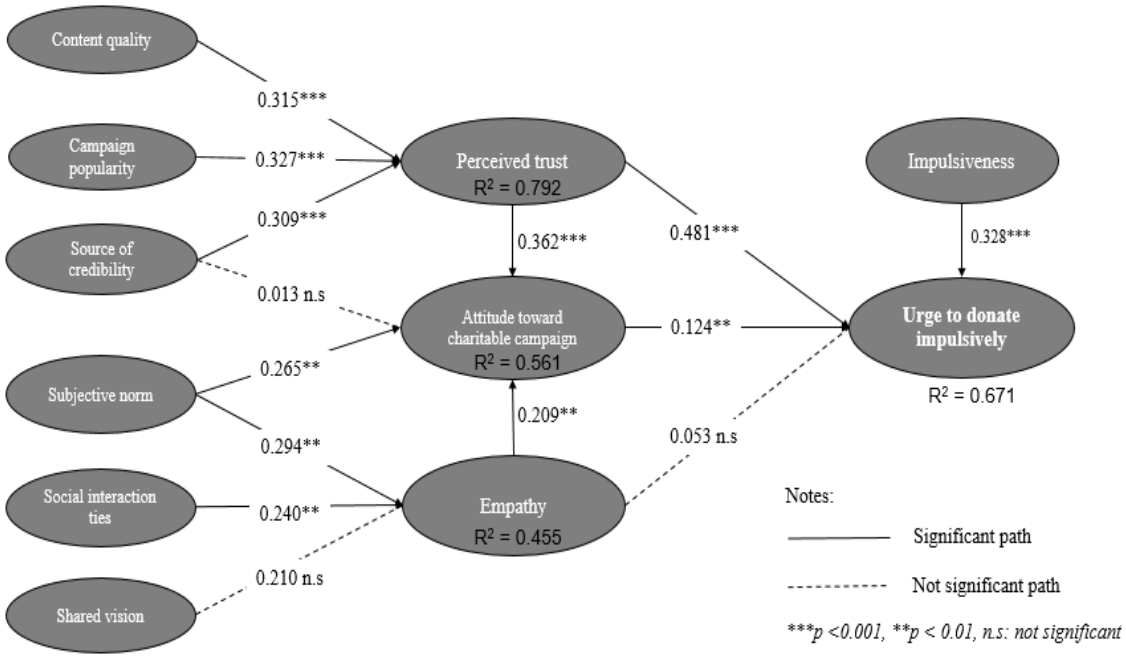
Hypothesis	Path	Beta	S.E	t-value	Conclusion
H1	Impulsiveness trait --> Urge to donate impulsively	0.328***	0.056	5.862	Supported
H2	Attitude toward charitable campaign --> Urge to donate impulsively	0.124**	0.059	2.125	Supported
H3	Perceived trust --> Attitude toward charitable campaign	0.362***	0.101	3.571	Supported
H4	Perceived trust --> Urge to donate impulsively	0.481***	0.070	6.852	Supported
H5	Empathy --> Attitude toward charitable campaign	0.209**	0.067	3.098	Supported
H6	Empathy --> Urge to donate impulsively	0.053 n.s	0.057	0.932	*Not supported*
H7	Content quality --> Perceived trust	0.315***	0.068	4.598	Supported
H8	Source of credibility --> Attitude toward charitable campaign	0.013 n.s	0.113	0.111	*Not supported*
H9	Source of credibility --> Perceived trust	0.309***	0.076	4.087	Supported
H10	Campaign popularity --> Perceived trust	0.327***	0.083	3.917	Supported
H11	Subjective norms --> Attitude toward charitable campaign	0.265**	0.095	2.798	Supported
H12	Subjective norms --> Empathy	0.294**	0.110	2.677	Supported
H13	Social interaction --> Empathy	0.240**	0.109	2.197	Supported
H14	Shared visions --> Empathy	0.210 n.s	0.142	1.473	*Not supported*

** Note: ***p <0.001, **p < 0.01, *p < 0.05; using a significance level of 0.05,t-value > 1.96.*

The SEM-PLS approach was used to test direct paths among variables which were proposed in the research model. The statistical results are reported in Table 5 and graphically demonstrated in Figure 2. As presented in the testing framework, the urge to donate impulsively as a depedent variable which was possitively explained by personal impulsiveness (β=0.328, p<0.001); attitude toward charitable campaign (β=0.124, p<0.01); and perceived trust (β=0.481, p<0.001). Thus, H1, H2 and H4 were significantly supported. However, the theoretically assummed positive relationship between empathy and urge to donate impulsively was not exhibited since the path coefficient (β=0.053) was not significant (p>0.05; t-value < 1.96), meanwhile, empathy has impacted on attitude toward charitable campaign with positive β=0.209 at the significance level of p<0.01. Thus, indicated H5 was supported but H6 was not supported. Besides, attitude toward charitable campaign was also positively linked with perceived trust (β=0.362, p<0.001) and subjective norms (β=0.265,

p<0.01), but not with source of credibility (β=0.013, p>0.05, t=0.111). Hence, H3 and H11 were supported, H8 was not supported. Regard to charitable campaign characteristics: content quality, campaign popularity and source of credibility were all presented a positive relationship with perceived trust with path coefficients correspondingly to β=0.315; 0.327; and 0.309 significant at p<0.001. These results implied H7, H9, H10 were supported. Among three social capitals, subjective norms (β=0.294) and social interaction ties (β=0.290) have shown significant associations to empathy at a significance level of p<0.01, however, empathy was not significantly related to shared visions (β=0.210, t=1.473). Therefore, it concluded that H12, H13 were supported, H14 was not supported.

Figure 2: Structural equation modeling (PLS-SEM) results



DISCUSSION & CONCLUSION

Drawing on perspectives of chairitable campaign characteristics and social capital theory, this study is primary to investigate mechanisms and antecedences that drive online user's impulsive donation on social media context. The empirical findings have some key insights to both theoretical and practical implications. First, this study attempts to fill research gap by providing understandings into the literature of online donation behaviors, in particular, impulsive donation making-decisions on social network environment. Empirical results have validated three key antecedences of user's impulsive donation behaviors including: perceived trust on charities, attitude toward charitable campaign and personal impulsiveness. Framed by the S-O-R paradigm, a systematic testing model has comprehensively examined the effects of environmental stimuli on impulsive donation intention under two main facets: features of charitable campaign and social capitals on social media settings.

Second, the Elaboration Likelihood Model is first-ever extended in the non-profit literature at the aim to explain the changes in a user's perceieved trust and attitude toward charities and online donation. It would further enrich the current knowledge with a deeper understanding of impulsive donor's behaviors. The present study categorized impulsive donor' decision-making processes as through either the central route or peripheral route under different conditions. Findings imply that the persuasion mechanism is induced by characteristics of charitable campaigns. Three determinants: content quality, campaign popularity and sources of credibility have indirectly

influenced a user's impulsive donation through a mechanism effect of perceived trust on charities. When impulsive donors process through the central route, they will be more motivated in deliberating the quality of campaign-related contents and scrutinizing messages from charities (Bhattacherjee & Sanford, 2006). The higher quality of contents enables donors to perceive which charity is of high credibility (L. Liu et al., 2017). Meanwhile, through the peripheral route, impulsive givers will perceive higher trust on a charity if they give judgements based on heuristic cues of persuasion such as sources of credibility and campaign popularity. Among them, campaign popularity exhibits a larger impact on perceived trust, indicated that impulsive donors are more likely process the charity's messages through the peripheral routes, mostly based on their perceptions of how much a campaign is popularized among user communities.

Third, this research has further confirmed the essential roles of social experiences via SNSs in driving user's online donation (Y. Kim & Lee, 2014; Li et al., 2019). However, by specifically examining impulsive donation behaviors in the context of social media, evidences show that subjective norms and social interaction ties have indirectly impacted on impulsive donor's decision. Instead, their associations were stimulated by user's positive attitude toward campaigns and empathic concerns. It also means that social media matter the online impulsive giving when it enables users to experience interpersonal interactions with other users through sharing stories, information regarding charitable campaigns. Under effects of these social influences and social experiences, impulsive donors tend to more actively involve in online giving when they have higher empathy and attitudes changed in a favor manner.

Last but not least, practical contributions of this study should not be ignored. Since such crucial findings would guide the non-profits how to cultivate campaign incentives and develop marketing strategies on social media to persuade larger potential donors and call for the sort of quick donations from user networks. Our study suggests that it is extremely imperative for chairities to build public confidences; to inspire positive attitudes and empathic concerns among online communities so as to draw user' impulse to donate for campaigns. Charities should strategically enhance high-quality contents, popularity of campaign and sources of credibility to elicit higher perceived trustworthiness for impulsive donors. Social network platforms can be a cost-effective marketing vehicle to spread out charitable campaigns to reach potential a large number of participants when charities know how to take its advantages and leverage social capitals to facilate user's emotional connectedness and interactions.

REFERENCES

Adler, P. S., & Seok-Woo, K. (2002). Social Capital: Prospects for a New Concept. *The Academy of Management Review, 27*(1), 17-40. doi:10.2307/4134367

Adloff, F. (2009). What encourages charitable giving and philanthropy? *Ageing & Society, 29*(8), 1185-1205.

Ahn, J.-c., Sura, S., & An, J.-C. (2018). Intention to donate via social network sites (SNSs). *Information Technology & People*.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin, 103*(3), 411.

Bennett, R. (2009). Impulsive donation decisions during online browsing of charity websites. *Journal of Consumer Behaviour: An International Research Review, 8*(2‐3), 116-134.

Bhattacherjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *MIS quarterly*, 805-825.

Brinol, P., Petty, R. E., & Tormala, Z. L. (2004). Self-validation of cognitive responses to advertisements. *Journal of consumer research, 30*(4), 559-573.

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research, 66*(1), 105-114.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of personality and social psychology, 66*(3), 460.

Chang, Y.-T., Yu, H., & Lu, H.-P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research, 68*(4), 777-782.

Chin, W. W. (1998). The partial least squares approach to structural equation modelling. In G. A. Marcoulides (Ed.), *Modern methods for business research*. Mahwah, New Jersey, United States of America: Lawrence Erlbaum.

Chiu, C.-M., Hsu, M.-H., & Wang, E. T. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision support systems, 42*(3), 1872-1888.

Digital. (2019). Digital report 2019. Retrieved from https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143-1168.

Farrow, H., & Yuan, Y. C. (2011). Building stronger ties with alumni through Facebook to increase volunteerism and charitable giving. *Journal of Computer-Mediated Communication, 16*(3), 445-464.

Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media:(How) can interactions on Twitter affect effectual thinking and behavior? *Journal of business venturing, 26*(1), 1-18.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*: Sage Publications.

Hausman, A. (2000). A multi-method investigation of consumer motivations in impulse buying behavior. *Journal of consumer marketing, 17*(5), 403-426.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems, 44*(2), 544-564.

Kim, Y., & Lee, W.-N. (2014). Networking for philanthropy: increasing volunteer behavior via social networking sites. *Cyberpsychology, Behavior, and Social Networking, 17*(3), 160-165.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration (IJeC), 11*(4), 1-10.

Li, B., Guan, Z., Hou, F., & Chong, A. Y.-L. (2019). How Social Experience Encourages Donation Intention to Charitable Crowdfunding Projects on Social Media: Empathy and Personal Impulsiveness. *Proceeding of the 23rd Pacific Asia Conference on Information Systems*.

Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in human behavior, 63*, 264-271.

Liu, L., Suh, A., & Wagner, C. (2017). *Donation behavior in online micro charities: An investigation of charitable crowdfunding projects.* Paper presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.

Liu, Y., Li, H., & Hu, F. (2013). Website attributes in urging online impulse purchase: An empirical investigation on consumer perceptions. *Decision support systems, 55*(3), 829-837.

McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin, 6*(3), 467-472.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of management review, 23*(2), 242-266.

NPsource. (2018). The Ultimate List Of Charitable Giving Statistics For 2018. Retrieved 05/04/2020 https://nonprofitssource.com/online-giving-statistics/

Park, E. Y., & Rhee, J. H. (2019). Who Clicks on Online Donation? Understanding the Characteristics of SNS Users during Participation in Online Campaigns. *Sustainability, 11*(13), 3674.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24): Springer.

Phua, J., & Ahn, S. J. (2016). Explicating the 'like'on Facebook brand pages: The effect of intensity of Facebook use, number of overall 'likes', and number of friends''likes' on consumers' brand outcomes. *Journal of Marketing Communications, 22*(5), 544-559.

Piron, F. (1991). Defining impulse purchasing. *ACR North American Advances*.

Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 839-903.

Populus. (2018). Trust in Charities, 2018. *Charity Commission for England and Wales*.

Ranganathan, S. K., & Henley, W. H. (2008). Determinants of charitable donation intentions: a structural equation model. *International journal of nonprofit and voluntary sector marketing, 13*(1), 1-11.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS quarterly*, 841-864.

Rook, D. W., & Fisher, R. J. (1995). Normative influences on impulsive buying behavior. *Journal of consumer research, 22*(3), 305-313.

Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Lebherz, P. R. (2014). Factors influencing popularity of branded content in Facebook fan pages. *European Management Journal, 32*(6), 1001-1011.

Savas, S. (2016). Factors affecting donations in US retail stores: a conceptual framework. *Journal of Retailing and Consumer Services, 33*, 178-185.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review, 8*(3), 220-247.

Stuart, D., Teng, S., Khong, K. W., Goh, W. W., & Chong, A. Y. L. (2014). Examining the antecedents of persuasive eWOM messages in social media. *Online Information Review*.

Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility*: MacArthur Foundation Digital Media and Learning Initiative.

Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information systems research, 14*(1), 47-65.

Taniguchi, H., & Marshall, G. A. (2014). The effects of social trust and institutional trust on formal volunteering and charitable giving in Japan. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 25*(1), 150-175.

Teng, S., Khong, K. W., Chong, A. Y. L., & Lin, B. (2017). Persuasive Electronic Word-of-Mouth Messages in Social Media. *Journal of Computer Information Systems, 57*(1), 76-88. doi:10.1080/08874417.2016.1181501

Verhaert, G. A., & Van den Poel, D. (2011). Empathy as added value in predicting donation behavior. *Journal of Business Research, 64*(12), 1288-1295.

Wang, L., & Graddy, E. (2008). Social capital, volunteering, and charitable giving. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 19*(1), 23.

Waters, R. D., Burnett, E., Lamm, A., & Lucas, J. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public relations review, 35*(2), 102-106.

Webb, D. J., Green, C. L., & Brashear, T. G. (2000). Development and validation of scales to measure attitudes influencing monetary donations to charitable organizations. *Journal of the academy of marketing science, 28*(2), 299-309.

Wiener, J. L., & Mowen, J. C. (1986). Source credibility: On the independent effects of trust and expertise. *ACR North American Advances*.

Wiepking, P., & Maas, I. (2009). Resources that make you generous: Effects of social and human resources on charitable giving. *Social Forces, 87*(4), 1973-1995.

Williams, D. (2006). On and off the'Net: Scales for social capital in an online era. *Journal of Computer-Mediated Communication, 11*(2), 593-628.

Xu, J., Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: an empirical test in the e-service context. *MIS quarterly*, 777-794.

DECISION SCIENCES INSTITUTE

Statistics and Big Data – Will they Tango?

Rakesh Kulkarni.
Indian Institute of Technology Bombay.
Address: #601, MANAS, Hill Side,
IIT Campus, Powai,
Mumbai – 400076
India
E-mail address: kulkarni.rakesh1@gmail.com
Phone: +91-908 219 9091

ABSTRACT

With the advent of big data technologies and the excitement they have managed to generate, companies and research institutions alike have taken a fascination to big data. This paper provides insights into the acceptability of big data tools among Statisticians and whether big data technologies would evolve to become amenable to use by Statisticians. The evolution and acceptance of technology within the Information Technology and Information Systems framework is studied. The Giddens' Structuration theory, the Actor- Network theory and the theory of Performativity are used to draw conclusions using an iterated triangulation method.

KEYWORDS: Statistics, Big data, Agency, Technology and Technology acceptance model.

INTRODUCTION

Big data is generated from a plethora of sources, including but not limited to Internet clicks, mobile transactions, user-generated content, social media, sensor networks and business transactions (Gerard et al. 2014). All other areas that generate and use large amounts of data such as genomics, health care, engineering, research and development, operations management, the industrial Internet, and financial processes contribute to the fascination for Big data technologies. For example, a participant in a Formula 1 car race generates 20 gigabytes of data from the 150 sensors on the car that can help analyze the technical performance of its components, but also the driver reactions, pit stop delays, and communication between crew and driver that contribute to overall performance. The use of large-scale data to predict human behavior is gaining currency in business and government policy practice, as well as in scientific domains (CERN) and where the physical and social sciences converge (recently referred to as "social physics")

Some of the tools used in the big data environment are given in Table 1:

Big Data Tools	Uses
NoSQL	Querying on large volumes of data
MongoDB	An open source software framework for distributed storage and processing of large amounts of data on clusters of computers
Apache Hadoop	An open source, non-schema, NoSQL Database
Apache Ambari	Managing and monitoring Apache Hadoop clusters
Hadoop MapReduce	Parallel processing of large datasets
Hive	Data summarization
Mahout	Machine learning and data mining library
Zookeeper	Coordination Service for distributed applications

Table 1: Tools used in the big data environment

Structured and Unstructured data:

Structured data is data that resides in a record or a file within a fixed field. It can be data types such as numeric data, names, addresses and dates. Unstructured data is the opposite of structured data and includes text files, videos and images.

For the purposes of this study I would be looking at structured data.

Big data vs. “Normal” data:

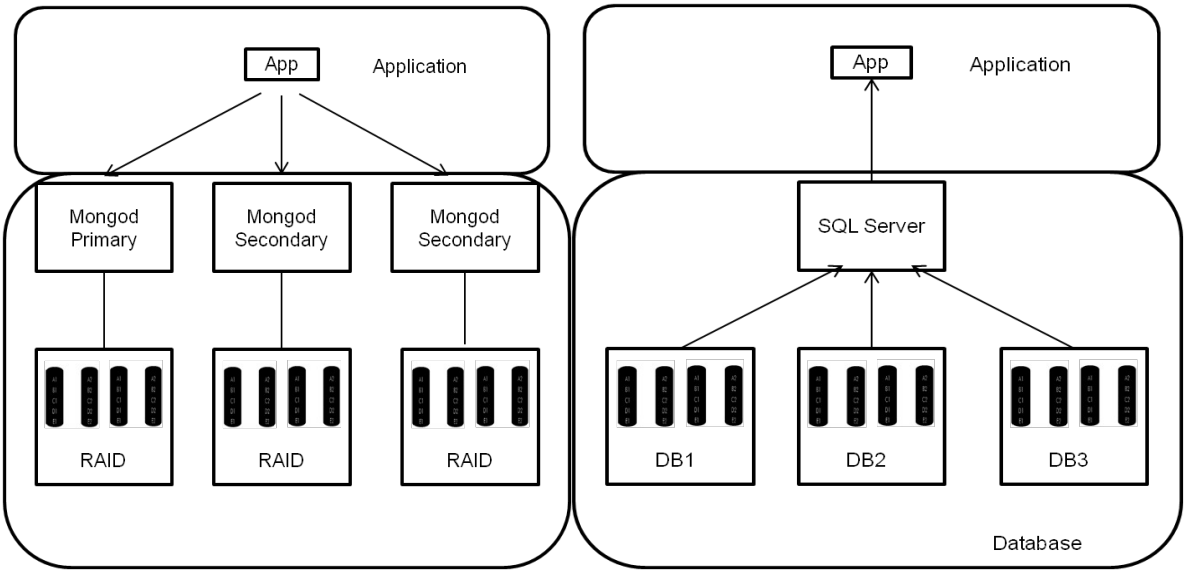


Fig.1. Big data vs. “Normal data”

The primary differences between Big data and “Normal” data are as follows:

1. Population vs. Sample – With the disruptive way in which big data technologies allow us to store humungous amounts of data, businesses might be tempted to use the entire population to try and avoid a sampling bias.

2. Correlation vs. Causality – Causality cannot be determined using large amounts of data and we can at best be able to find the strength of a linear relationship.

3. Algorithm-to-data vs. Data-to-algorithm – Due to the manner in which the data is stored in a big data context, the algorithm has to “go to” to the data stored in different clusters and the results aggregated and not vice-versa. In the traditional data storage methods, data, even if stored in different clusters would collectively be “taken to” the algorithm.

LITERATURE REVIEW

Given the rapid evolution of, and interest in big data by industry which has leapfrogged the discourse to popular outlets, the academic press has been forced to catch up. Academic journals in numerous disciplines have not yet covered this topic in significant detail. This being mentioned in an editorial in a leading academic journal shows the lack of debate on big data (Gerard et al., 2014). The most disruptive innovation attributed to big data technologies is that storage of high volume data has become possible. There are multiple sources of high volume data and there are five key sources of high volume data (Gerard et al., 2014):

(1) Public data - data typically held by governments, governmental organizations, and local communities.

(2) Private data – data held by private firms, non-profit organizations, and individuals.

(3) Data exhaust - data collected passively, data with very little value to the original data collection partner. This includes data on usage of mobile phones and Internet searches.

(4) Community data - a filtered form of unstructured data, primarily text into social and e-commerce networks. An example would be consumer reviews of a product written on these sites.

(5) Self quantification data – data that are revealed by the individual through quantifying personal actions and behaviors. Examples include smart-watches and shoes that monitor exercise and movement.

Given the potentially earth shattering innovation, then, how has the Statistics community accepted big data or is it interested in it at all? The question is pertinent because a new method of storing “humongous” data would be useful in business decision making only if it can be analyzed and is amenable to statistical modeling and as the editorial states, the typical statistical approach of relying on 'p' values to establish the significance of a finding is unlikely to be effective because the immense volume of data means that almost everything is significant (Gerard et al., 2014). Likewise, arriving at false correlations with alarming frequency is a possibility using our typical statistical tools. While more complex techniques would mean over-fitting the data, basic Bayesian statistics and stepwise regression methods may work.

Some other techniques proposed to be useful are soft computing techniques that are not well accepted by statisticians such as cluster analysis, data fusion and integration, data mining, genetic algorithms, machine learning, natural language processing, neural networks, network analysis, signal processing, spatial analysis, simulation, time series analysis, and visualization.

The way in which big data is stored makes it imperative to use aggregation techniques and it is this pitfall of big data which lies in focusing too much on aggregates or averages and too little on outliers. But, in the vastness of a big data universe, the outliers may be large enough in number to have significant effects. Given the way in which data collection happens, big data techniques

applied to group dynamics and behavior can potentially transform management theory and practice.

Apart from the multiple sources from which big data is collected, there are some distinctive features inherent in big data. These are (Fan et al., 2014):

1. Heterogeneity - Big data is obtained from multiple sources and is representative of different sub-populations.
2. Noise accumulation - the estimation of several parameters in parallel could be dominated by cumulative estimation error or noise.
3. Spurious correlations - pairs of variables are found to be interdependent when they are not.
4. Incidental endogeneity - violates an assumption of regression that the predictors are not correlated with the residual term.

Incidentally, large corporations were the first to explore the big data phenomenon. (Schlegel, 2015) studies different companies (Dell, Schneider, IBM among others) and the manner in which they are using big data (for predictive modeling, optimization) while noting the need to develop more advanced visualization techniques and insisting on embracing the data. Big data Analytics capabilities and tools studied by Schlegel are depicted in Figure 2.

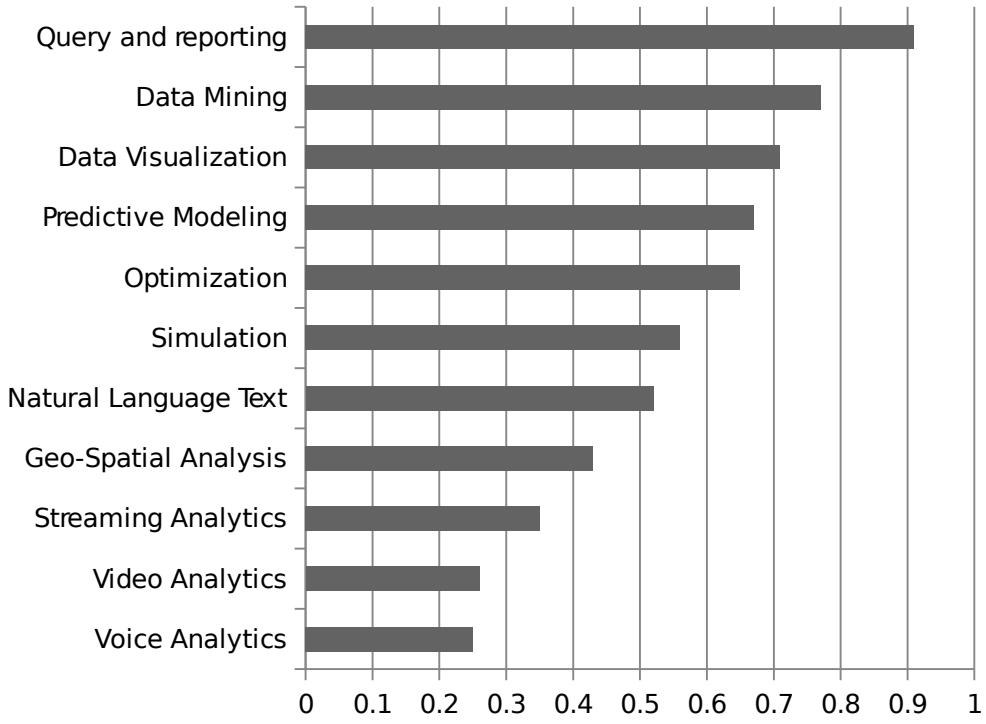


Fig.2. Big data Analytics capabilities and tools studied by Schlegel.

IBM has a now more widely accepted definition of big data in terms of “4Vs”:

1. Volume: “Data to Scale – Terabytes to petabytes and more”

2. Variety: “Data in many forms – Structured, unstructured, text, multimedia”
3. Velocity: “Data in Motion – Analysis of streaming data to enable decisions within seconds”
4. Veracity: “Data Uncertainty – Managing the reliability and predictability of imprecise data types” (Schlegel, 2015).

Big data is being accumulated at a furious pace by the scientific community. The Large Hadron collider is a case in point. The Large Hadron collider produces close to 30 petabytes of data every year. Whether scientists are able to utilize this data to generate new theories is a moot question. Frické (Frické, 2015) considers whether “big data, in the form of data-driven science, will enable the discovery, or appraisal, of universal scientific theories, instrumentalist tools, or inductive inferences”. As he points out such aspirations are similar to the “now discredited” inductivist approach to science. On the one hand, big data would allow us to use large and comprehensive samples, extensive testing of theories that is also less expensive, and continuous examination of theories. On the flip side, data driven science could encourage data collection alone, as against testing and experimentation and “hornswoggling” or “unsound statistical fiddling” (Frické, 2015). Frické (Frické, 2015) argues that theory is needed at every juncture of development of inductive algorithms, statistical modeling, and scientific discoveries and that these would not progress based on data alone.

The Research Questions

I believe the problem mentioned in my paper's title can be addressed by looking at three pertinent research questions:

1. Causality: With the advent of big data and its disruptive techniques of storing data, does statistics have to adapt to it, or is big data a passing fad since it would not survive without enough statistical modeling algorithms that can run on it?
2. Finality: Does technology, in our context big data, follow statistics or do statisticians (in terms of implementation) have to adapt to the latest in technology or do they just follow each other like night and day?
3. If neither happens, what would be the best way to put big data to use?

THEORETICAL DEVELOPMENT

One way of understanding whether big data and statistics would complement each other is by studying the evolution of technology and its adaptation by the intended users. The areas of Information Science and Information Technology are closely allied with both data and business decision making. By studying the evolution of Enterprise Resource Planning (ERP) software, we would be in a position to gain insights into the relationship between big data and statistics. In the Information Science context, Rose et al., study literature concerning the relationship between technology and organization – whether technology causes effects in an organization or do humans determine how a new technology is put to use (Rose, J et al., 2005). Theories of Information Technology innovation focus on the concept of agency. Let us look at three main theories below.

Theories of Information Technology Innovation

The three theories are Giddens' Structuration Theory, Actor-Network Theory and the Theory of Performativity and the briefly described in this section.

Giddens' Structuration Theory

Structuration based approaches are biased towards human agency over technological agency. Here, the human subject is the center of the action. Agency in Structuration theory is shaped by structure, while structure is produced and evolved, by the actions of humans in society. The relationship between structure and agency is seen as "a 'virtual order' of transformative relations...that exists, as time-space presence, only in its instantiations in (reproduced social) practices and as memory traces orienting the conduct of knowledgeable human agents" that is, apparently material resources such as IT which "might seem to have a 'real existence,' become resources only when incorporated within processes of structuration". Further, "technology does nothing, except as implicated in the actions of human beings" (Giddens et al., 1998). The Structuration theory does not allow for social rules to be embedded into technology during its design since this would imply a material existence of structures which Giddens denies (Giddens et al., 1998).

Actor-Network Theory

These investigations treat human and technological agency on par, but have assumptions of symmetry and do not account adequately for differences between humans and machines. Here, the human subject is not the center of the action. Agency and agents are replaced by actors and actants, where actants can be human or non-human (Latour, 1987). Action here is defined in terms of one particular type of consequence: a role in a socio-technical network and particularly the historical development of those networks (Latour, 1991). Since agency is not restricted to humans, any technological outcome is constrained by the physical or material needs of developing such a technology along with the resistance offered by the humans.

Theory of Performativity

Under this theory, "material performances and human agencies are both implicated in the other (human agency is always materially performed, just as material performances are always enacted by human agency), and neither are given apriori but are temporally emergent in practice" (Orlikowski, 2005). This view allows the recognition of unanticipated conditions and unintended consequences of the interrelationships between the two, allowing us to understand "institutional outcomes, social purposes, and human reflexivity."

RESEARCH METHODOLOGY

Given the nature of the questions to be answered, descriptive methods need to be used as against the quantitative techniques. One can use the approach of "iterative triangulation" (Lewis, 1998) which uses existing case study evidence, literature review and intuition, thus letting ourselves exploit exhaustive empirical information and at the same time gain insights into the current situation. Iterated triangulation, as the name suggests, involves selecting relevant cases

and iterating between case evidence, literature review and intuition until one is able to form a cohesive opinion, backed by facts. Dean et al., called case studies the ‘‘predominant mode of inquiry into the effects of new technology’’ (Dean et al., 1992).

FINDINGS

The differences between Giddens’ Structuration Theory and the Actor-Network Theory can be characterized as a problem of agency (Rose et al., 2005). This problem of agency can then studied empirically through ERP systems. Agency here refers to actions which have consequences or in Giddens’ terms – “the capability to make a difference”. A causal relationship between actions of particular agents and their effects is thus studied. Agency can lie either with the humans having effects on the technology or it may lie with the technology thus effecting the humans. ERP systems are referred to as a technological imperative that organizations must adapt to. In their 3 case studies of ERP implementation, the authors find that ERP was entirely a function of human agency at “Omega” and its success or failure was dependent on the actions of the managers of the company. The system did bring in some order when implemented at a company named “Martin” with no existing systems in place. SAP implementation at “BCTel” faced resistance from the workers' union when it tried to replace an existing legacy system with the union attributing job cuts and thus agency to both the management and the “Armageddon Machine” (Rose et al., 2005). Infrastructures can be viewed as actors and ERP software as an agency. SAP was a powerful actor and an ally in starting the process of change in a company. With the increase in the complexity of ERP infrastructure, there is a corresponding increase in SAP's resistance to control and it becomes a more independent actor (Hanseth et al., 2000).

The Actor-Network Theory regards humans and non-human as equally endowed with the power to act (Holmstrom et al., 2001) and technology is an actor because it has been endowed with the ability to act through its position in the network. In analyzing the failure of the Swedish cash card, the authors analyze interests of the human actors exclusively including banks, merchants, customers, among others, thereby ignoring the agency of technology (Holmstrom et al., 2001).

The over fitting inherent in big data is captured in Google’s flu tracking system, the Google Flu Trends’ (GFT) results of 2013 (Lazeret al., 2014) where for influenza-like illness it predicted more than twice the proportion of doctor visits. The GFT consistently overestimated cases of flu over a period of 3 years. The assumption of randomly distributed errors is violated and the homoscedasticity exhibits seasonality with the direction and the magnitude of errors varying with the time of year. A forecast using these forecasts only took the GFT into a vicious cycle. This could have been avoided by following traditional statistical modeling tools. “Big data hubris” led to the use of Big Data as a substitute for, rather than being a supplement to, time tested data collection and analysis, which in turn led to these vastly inflated predictions.

Companies have adopted big data technologies at a furious pace. At Cloudera, a leading vendor of Hadoop services to e-commerce, telecommunications, media, and retail customers that includes the likes of Facebook, over two million jobs were run on MapReduce within a year, using over 5000 machines and approximately 1.6 exabytes of data (Chen et al., 2012). From 2009 to 2010 data cleansing and processing of intermediate data sets has grown while processing for the final computation has shrunk. Thus, while Cloudera's subset of clients mentioned earlier have seen an increased customer base and more data collection, the final

stage of data processing to provide analytical insights for business decision making have remained the same (Chen et al., 2012).

Big data is considered in some quarters to have the potential to transform new age and traditional businesses alike. “It is expected to create a revolution far more powerful than the analytics that were used in the past by making better predictions and smarter decisions” (McAfee & Brynjolfsson, 2012). In this case study, researchers at the Johns Hopkins School of Medicine, found that they could use data from Google Flu to predict surges in flu-related emergency room visits a week before warnings came from the Centers for Disease control. The accuracy and validation of these predictions does not find mention. PASSUR Aerospace is a provider of decision-support technologies for the aviation industry that collects a wide range of information about every plane that it “sees”, every 4.6 seconds. This yields a huge and constant flood of data. The company stores all this data collected over a decade, creating an “immense body of multidimensional information”. This it believes allows high end analysis and pattern matching. Scheduling flights for an airline based on forecasting the landing of its planes is worth several million dollars a year at individual airports. The study concludes noting that not all the winners will be using big data to transform decision making although the surest way to transform decision making lies with big data (McAfee & Brynjolfsson, 2012).

Introduction of IT can immensely help hospitals and health care delivery and outcome. Use of modern Hospital Information Systems offers tremendous opportunities to support health care professionals and to increase the efficiency, effectiveness and appropriateness of care. However, it is estimated that close to 60-70% of all Hospital Information Systems fail (Ammenwerth et al., 2006). Two pertinent questions that need to be answered in this context are: What are the "socio-organizational" factors that influence adoption of an IT system in a given socio-organizational context? And based on the answers to the first question, is there any way to predict the effects of an IT system in a certain context? A framework of - "Fit between Individuals, Task and Technology" (FITT framework) is used to answer the first question which is based on the idea that IT adoption in a heath care setting depends on the fit between the attributes of the users (such as skill, motivation, interest), of the attributes of the health care tasks (such as task complexity, manpower planning, scheduling) and of the attributes of technology (such as usability, robustness, functionality, performance). All the attributes of the users and of the technology and of the health care environment and of the supported health care tasks are studied in a manner similar to Actor-Network theory to understand IT adoption processes. The German University hospital which is studied in the case looks at the quality of output before and after the introduction of a new IT system. While some of the nurses from different wards took to the new processes (fit between individuals and task) enthusiastically, some were only moderately interested (Ammenwerth et al., 2006). The Healthcare Information Systems professionals’ reaction to the new technology was similar to that of the nurses’ reaction to the new systems (fit between individuals and technology). The fit between task and technology however was problematic and parts of the new investment were either deemed unnecessary in some wards or were realized to be inadequate in others. Interventions were made in terms of training on the new processes, software and investments in hardware to address the issues. The output after the intervention was positive and all wards were seen to be working better than before the new Healthcare Information System was introduced.

As the Giddens’ Structuration theory suggests, the human subject is the center of the action and human agency is preferred over technological agency. Not surprisingly, lack of user acceptance has long been an impediment to the success of new information systems. A technology acceptance model (TAM) can be built and is defined as one that “specifies the causal relationships between system design features, perceived usefulness, perceived ease of use,

attitude toward usage, usage behavior and thus underscoring the importance of incorporating the appropriate functional capabilities in new systems" (Davis, 1993). In rating the existing e-mail system and a text editing software in a large American corporation, the responses of 112 professionals and managerial employees' were analyzed. In the Ordinary Least Squares and hierarchical regressions performed on the data, attitude had a significant effect on usage, perceived usefulness had a significant effect on attitude, ease of use had a significant effect on attitude and on usefulness, the system had a significant effect on ease of use and on attitude, but had no significant effect on use, with e-mail perceived to be easier to use than the text editor. The system had no significant effect on usefulness, thus e-mail and text editors give similar perceived impacts on job performance, except for differences in ease of use. In short, characteristics of the system influence human behavior through motivational variables and have no direct effect on use. There is a strong effect of usefulness on actual use, directly and indirectly, through attitude. Perceived ease of use has a small effect on attitude. However, the significant effect of system characteristics on attitude towards using indicates that perceived usefulness and perceived ease of use are not the only attributes mediating between system and attitude. A limitation in the study is that some users may be compelled to use a new technology by the management and do not find it to be useful or easy to use, which the TAM model does not capture. While ease of use and user-friendly interfaces are important, usefulness of the technology is of primary importance and cannot be ignored. Other TAM studies have found computer self-efficacy acts as a determinant of perceived ease of use both before and after hands-on use (Venkatesh & Davis, 1996). User acceptance is tested based on user perceptions.

Organizational variables like social norms play a vital role in influencing user acceptance of new technology, along with other factors. Social norms and job requirements are found to be more important in predicting use than workers' perceptions about ease of use and usefulness (Lucas et al., 1999). Brokers' and sales assistants' responses in a major investment bank studied in a TAM framework along with variables like social norms and prior performance revealed that social norms and prior performance are more important than user perceptions in determining user acceptance (Lucas et al., 1999). Poor performers perceive that using a system can improve their performance in this study (Lucas et al., 1999).

Intrinsic motivation plays an important role along with perceived ease of use and perceived usefulness (Venkatesh et al., 2002). In a study, participants completed a questionnaire, before and after training, measuring perceived usefulness, perceived ease of use, intrinsic motivation, and intention to use variables immediately after training. Usage the technology was tracked for 12 weeks as a measure of short term use. Perceptual, motivational, and intention measures were taken again at this point, and subsequent actual technology usage was tracked for an additional 12 weeks (Venkatesh et al., 2002). The theory based on TAM and intrinsic motivation suggests that intrinsic motivation influenced perceived ease of use and perceived usefulness; perceived ease of use influenced perceived usefulness; and intrinsic motivation, perceived ease of use and perceived usefulness all have significant positive effects on behavioral intention. Analyzing the data using a Structural Equation Modeling technique (EQS), it was found that intrinsic motivation influenced perceived ease of use and perceived usefulness, perceived ease of use influenced perceived usefulness significantly, perceived ease of use and perceived usefulness have significant positive effects on behavioral intention. However, intrinsic motivation was found to have no direct significant effect on behavioral intention, but influenced it indirectly through perceived ease of use and perceived usefulness. Behavioral intention influenced usage behavior and immediate use was the only significant predictor of continued usage. Thus, intrinsic motivation did not directly influence intention to use technology. Managerial interventions made the technology more acceptable to the users over time are considered to be

a powerful strategy for inducing widespread adoption and use of new technology (Venkatesh et al., 2002).

As seen in the previous case studies, managerial intervention played an important role in user acceptance of new technologies. Managerial incentives and control are thus to be examined as important components of managerial influences, which are linked to IT usage via a principal-agent model (PAM) (Bhattacherjee, 1998). Managers have the ability to motivate users' behavior within organizations by designing appropriate incentives and control structures intended to make IT usage acceptable.

Analytics technologies are decision support systems at the end of the day. Major factors influencing the use of decision support systems are attributes of the decision makers such as age, educational level, experience, and cognition; attributes of the decision support system such as response time, accuracy, timeliness, relevance of output; and attributes of the implementation process such as user involvement, user training, and top management support (Fuerst & Cheney, 1982). When studying 8 systems and 64 subjects from the oil industry, it was concluded that the most important variables affecting decision support system usage were accuracy of output, user training during the implementation process, relevance of output, and the decision maker's experience, based on multivariate regression (Fuerst & Cheney, 1982).

The tendency of the users towards initial use of an innovation and intentions to continue such use in the future voluntarily can shed light on user acceptance of technology. Current usage is not a significant predictor of future use intentions which suggests that momentum generated by initial use cannot be relied upon for continued, sustained use of a technological innovation (Agarwal & Prasad, 2007). Perceptions of voluntariness significantly explain current usage. Continued usage intentions, however, are not affected. People will continue to use the system only if they are able to view its benefits unequivocally. Perceptions of voluntariness may be important for acceptance behavior initially, but for continued use, users decide based on their own evaluations of the innovation.

DISCUSSION AND CONCLUSIONS

Big data is a phenomenon that is already here, and was not built in consultation with the statisticians, since due consideration was not given to the way in which statisticians have traditionally analyzed data. This is contrary to the systems approach followed by all the successful analytics software organizations in the world today. R software is a case in point. The colossal success of Silicon Valley is attributed to the systems thinking prevalent in the area where professionals with varied backgrounds helped make the IT industry a magnificent success story. But now that it is here, will it evolve? Or would Statisticians adapt to it by creating new algorithms/metrics to make sense of the data? If neither happens, what would be the best way to put it to use?

Widely used Statistical modeling techniques such as time series modeling and other techniques that require the entire data set to be used will not be amenable to big data analysis. The existing techniques that can be used on big data are limited to iterative and soft computing algorithms like K-means clustering, neural networks, genetic algorithms among others. As of today, there are all of 5 statistical algorithms that can run on a Mahout and MapReduce framework.

The Swedish Cash card system was a failure since the users did not adapt to it. However the machines became adopted globally after banks hard sold them to the merchants, with Citibank even providing the card swiping machines to the merchants for free or for a nominal rent, without making changes to the technology itself. The same cannot be said about big data since

it is already available for free and would involve massive investments, in terms of time and development of Intellectual Property, in developing new algorithms to run on that framework that mimic existing statistical techniques. Analytics companies are clearly undecided and open source contributors seem uninterested given their innumerable contributions, including Bayesian methods to other open source software such as R.

Google's teams working on big data and traditional statistics obviously failed to collaborate on the GFT project while being aware of each others' existence. It is of paramount importance to rely on sound statistical modeling techniques rather than on the size of the data alone. More data does not necessarily mean better predictive accuracy.

Given that a large number of organizations with mammoth training budgets have adopted the big data technologies, there is no dearth of a workforce skilled in technologies using big data. Skills such as SQL querying are easily transferable across multiple platforms. As was seen in the German University Hospital case study, interventions in terms of investments in training, software improvements and investments in hardware made the new technology acceptable. Thus, it is not the lack of skill or training that has prevented enough statistical models being built on the big data framework which would in turn help in business decision making, but the lack of amenability of the technology to most statistical models.

As underscored in the TAM studies, the characteristics of the system influence human behavior through motivational variables and have no direct effect on use. Big data is not a panacea for poor performance of an existing analytics set up. The hype generated by big data and "Big data hubris" have encouraged some organizations to jump the big data bandwagon without being able to improve business decision making accruing from the increase in the size of the data alone. Managerial interventions may be responsible for the acceptance of big data technologies in some of the corporations, since the corporations have already invested in these technologies. Since huge investments of resources are involved in the development of new technologies, user acceptance testing and better still, involving the end users during the development phase of a new technology in a systems approach that characterizes the success of Silicon Valley is of paramount importance.

Big data will not replace time tested database tools or statistical modeling techniques. Its impact on analytics will remain limited and would be used in conjunction with the existing systems since it does not purport to change the way in which structured data is analyzed and is merely a new means of storing data.

REFERENCES

Agarwal, R. & Prasad, J. (1997). The Role of Innovation Characteristics and Perceived Voluntariness in the Acceptance of Information Technologies. *Decision Sciences*, 28: 557–582.

Ammenwerth E, Iller C, Mahler C. (2006). IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study, *BMC Medical Informatics & Decision Making*, 6.

Andrew McAfee & Erik Brynjolfsson (2012). Big data: the management revolution, *Harvard business review* 10/; 90(10):60-6, 68, 128

Bhattacherjee, A. (1998), Managerial Influences on Intraorganizational Information Technology Use: A Principal-Agent Model. *Decision Sciences*, 29: 139–162

Davis, Fred D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, 38.3: 475-487.

Dean, J.W. Jr., Yoon, S.J., Susman, G.I. (1992). Advanced manufacturing technology and organization structure: empowerment or subordination? *Organization Science* 3 (2), 203–229.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2), 293–314.

Fuerst, W. L. & Cheney, P. H. (1982), Concepts, Theory, and Techniques: factors affecting the perceived utilization of computer-based decision support systems in the oil industry. *Decision Sciences*, 13: 554–569.

George Gerard, Martine R. Haas & Alex Pentland. (2014) Big Data and Management. Editorial. *Academy of Management Journal*, Vol. 57, No. 2, 321-326.

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration. *University of California Press.*

Giddens, A. & Pierson, C. (1998). Conversations with Anthony Giddens: Making sense of modernity. *Stanford University Press.*

Gregory L. Schlegel. (2015). Utilizing Big Data and Predictive Analytics to Manage Supply Chain Risk*. Journal of Business Forecasting*, Winter 2014-2015, 11-17.

Hanseth, O., & Braa, K. (2000). Who's in control: Designers, managers-or technology? Infrastructures at norsk hydro. *From control to drift: The dynamics of corporate information infrastructures,* 41-55.

Holmstrom, J. & Stalder, F. (2001) Drifting technologies and multi-purpose networks: the case of the Swedish cashcard. *Information and Organisation*, 11:187-206.

Latour, B. (1987). Science in action: how to follow scientists and engineers through society. Harvard *University Press.*

Latour, B. (1991). Technology is society made durable, A sociology of monsters: essays on power, technology and domination, J. Law (ed.), *Routledge, London*, 103-131.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). “The Parable of Google Flu: Traps in Big Data Analysis*.” Science* 343 (6176): 1203–1205.

Lewis, M.W. (1998). Iterative triangulation: A theory development process using existing case studies. *Journal of Operations Management*, 16, 455–469.

Lucas, H. C. & Spitler, V.K. (1999), Technology Use and Performance: A Field Study of Broker Workstations. *Decision Sciences*, 30: 291–311.

Martin Frické. (2015). Big Data and Its Epistemology, *Journal Of The Association For Information Science And Technology*, 66(4):651–661.

Orlikowski, Wanda J. (2005). Material Works: Exploring the Situated Entanglement of Technological Performativity and Human Agency. *Scandinavian Journal of Information Systems*. Vol. 17: Issue 1, Article 6.

Rose, J., Jones, M. & Truex, D. (2005). Socio -Theoretic Accounts of IS: The Problem of Agency. Scandinavian Journal of Information Systems, (17:1): 133-152.

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision sciences*, *27*(3), 451-481.

Venkatesh, V., Speier, C., & Morris, M. G. (2002). User acceptance enablers in individual decision making about technology: Toward an integrated model. *Decision Sciences*, *33*(2), 297-316.

Yanpei Chen , Sara Alspaugh, & Randy Katz (2012). Interactive analytical processing in big data systems: a cross-industry study of MapReduce workloads, *Proceedings of the VLDB Endowment*, v.5 n.12.

DECISION SCIENCES INSTITUTE
Context-Based Sentiment Analysis

Guneeti Sharma
University of Massachusetts Boston
Email: guneeti.sharma001@umb.edu

Josephine M. Namayanja
University of Massachusetts Boston
Email: josephine.namayanja@umb.edu

Abhishesh Adhikari
University of Massachusetts Boston
Email: abhi@workership.com

ABSTRACT

Sentiment analysis is an approach in natural language processing (NLP) that identifies the emotional tone behind the body of the text. However, the challenge is to extract a single sentiment from large context-sensitive text. This study presents a new hybrid approach to predict text-level sentiments based on generating a context-rich text feature vector by combining single words and word sequences to form a multi-dimensional word embedding structure using the Word2Vec model. The proposed approach improves on the traditional bag of words and TF-IDF models by efficiently detecting sentiment expressions in large textual phrases that contain various contexts of conversations.

Keywords: Bigram, Context-Based Sentiment, Deep Learning, Opinion Mining, Predictive Modelling, Unigram, Word2Vec

INTRODUCTION

Sentiment analysis is the interpretation and classification of emotions or opinions. For example, sentiments can be classified into positive, negative and neutral. Recently, sentiment analysis, also referred to as opinion mining has gained attention in the areas of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information [1]. Specifically, sentiment analysis is utilized discovering hot search keywords which facilitates search engine optimization (SEO) [1], social media analytics to improve marketing strategies, analyze organizational harmony through gathering employee feedback [1], service and product review and management [2], to mention a few.

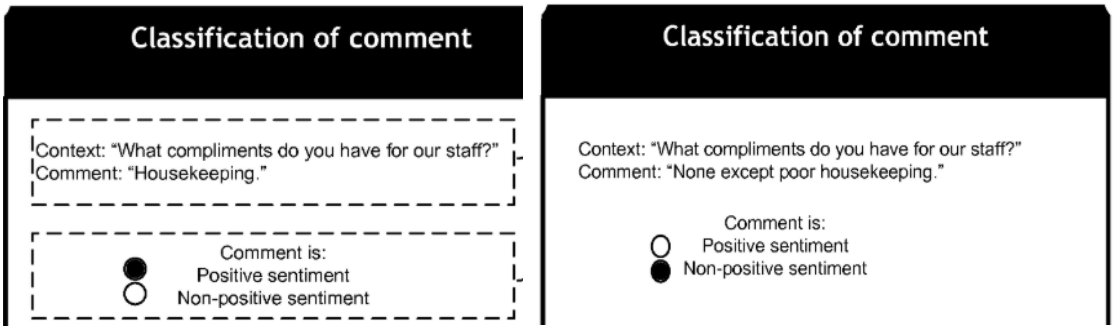
Advancements in pattern analysis enable automated detection of sentiments from massive unstructured data. However, determining sentiment is very subjective. For example, keyword processing only identifies the sentiment reflected in a single word, and thus typically fails to provide all of the elements necessary to understand the complete context in text. While natural language processing uses machine learning to efficiently analyze large unstructured text, the inherent complexity of language makes it difficult to ensure algorithms accurately analyze tone and context. Factors that pose such limitations

include grammatical nuances, implied meaning from facial expressions and body language, misspellings, ambiguity, and regional or cultural variations in language [3].

There are several elements to take into consideration such as, the object, the attributes, the opinion holder, the opinion orientation, and the opinion strength [4]. While humans are generally better equipped to identify all five elements needed to accurately interpret the opinions expressed in a piece of text, manual processing presents its own challenges, primarily in regard to speed and scale. Additionally, subjective interpretations of opinions can lead to varying results. For example, a study by [4] found that humans only agree on interpretation 79% of the time. Such conflicts pose as a limitation in the objective analysis of sentiments. On the other hand, the lack of labeled data tailored to specific domains prohibits wide and effective utilization of sentiment analysis methods. Existing models often have poor adaptability between domains or different text genres because they often rely on domain specific features from their training data [4]. Furthermore, with a finite number of words in existing lexicons challenges arise when extracting sentiment from very dynamic environments. Secondly, sentiment lexicons tend to assign a fixed sentiment orientation and score to words, irrespective of how these words are used in the text.

In order to obtain complete, accurate and actionable information from a piece of text, it is important to not only identify the aforementioned elements individually, but to also understand how they work together to provide the full context of the text and the sentiment associated it. By examining context, one is also able to detect sarcasm and unique domain language in text. In Figure 1, we give an example of applying context in sentiment analysis.

Figure 1: Examining Context in Sentiment Analysis



Source: [4]

Therefore, in this study, we propose a hybrid approach to sentiment analysis, which combines words, word sequences and word embeddings to generate a large text feature set as input to detect context-based sentiment. Our objective is to measure context sensitivity, the determination being made with reference to a set of text features. We apply our proposed method on both short and large text and across datasets from different domains to validate our context-based approach to sentiment analysis. Our proposed model is trained and tested using a traditional ensemble classifier, specifically random forest, although other classifiers can be applied.

The rest of the paper is organized as follows, in section 2 we discuss related work, in section 3 we present our methodology, in section 4 we present our experimental results, in section 5 we discuss a summary of our findings and present our conclusions and future work.

LITERATURE REVIEW

Sentiment classification on Twitter has attracted increasing research in recent years. Most existing work focuses on feature engineering according to the tweet content itself. [8] It proposes a context based neural network model for Twitter sentiment analysis, by incorporating contextualized features from relevant Tweets into the model in the form of word embedding vectors. Most recent literature also indicates that sentiment analysis on social media platforms such as Twitter is based on the idea that the sentiment is a function of an incoming tweet. However, tweets are filtered through streams of posts, so that a wider context, such as a topic, is always available. [5] investigates the contextual information of tweets where they model the polarity detection problem as a sequential classification task over streams of tweets. In addition, a Markovian formulation of the Support Vector Machine discriminative model is employed to assign the sentiment polarity to entire sequences. Outcomes from their study prove that sequential tagging effectively embodies evidence about the contexts and is able to reach a relative increment in detection accuracy of around 20% based on F1 measure

On the other hand, [6] proposes and approach called SentiCircles, which takes into account the co-occurrence patterns of words in different contexts in tweets to capture their semantics and update their pre-assigned strength and polarity in sentiment lexicons accordingly. Their approach allows for the detection of sentiment both at the entity-level and tweet-level and evaluated on three datasets [13] [14] [15] using three different sentiment lexicons to derive word prior sentiments. They argue that their approach significantly outperforms the baseline methods in accuracy and *F*-measure for entity-level subjectivity (neutral vs. polar) and polarity (positive vs. negative) detections. Furthermore, for tweet-level sentiment detection, the proposed approach performs better than the state-of-the-art SentiStrength by 4–5% in accuracy in two datasets but falls marginally behind by 1% in *F*-measure in the third dataset.

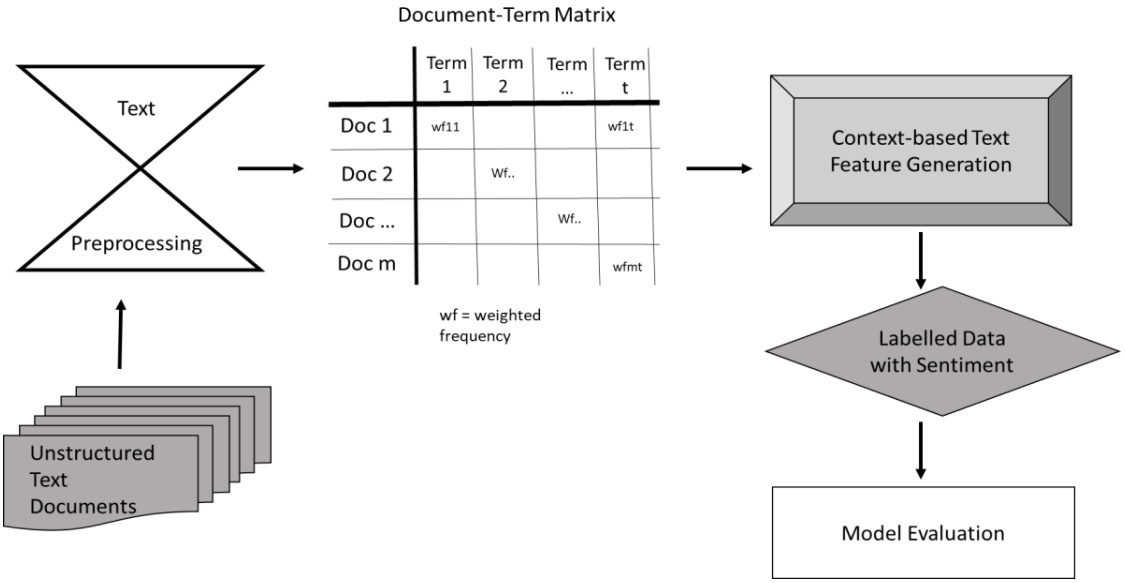
It's been noted that a predominant body in literature focuses on feature modeling of short text such as tweets, this poses limitations on the effectiveness of the sentiment analysis. This is because such methods do not leverage contextual information regarding the author/source or the tweet to improve the performance of text ambiguity such as sarcasm detection. [7] It compares sarcastic utterances in twitter to utterances that express positive or negative attitudes without sarcasm. The sarcasm detection problem is modeled as a sequential classification task over text and its contextual information.

While the aforementioned studies account for context in sentiment analysis, our hybrid approach for context-based sentiment analysis shows that multiple textual embeddings coupled with multiple pooling functions and sentiment lexicons offer rich sources of feature information, which leads to significant improvement in the accuracy of sentiment analysis.

METHODOLOGY

In this section, we present our methods as used in this study. Figure 2 provides an overview of our proposed approach.

Figure 2: Overall Approach



Next, we discuss the steps in detail.

Text Preprocessing

In this study, we utilize labelled unstructured text data from various data sources [8, 10]. We then apply data cleaning where we first convert the text from “uft8” to “Latin” in order to take the special words into consideration. Next, we are removing special characters such as ‘&’, ‘*’, ‘@’, so on. All hyperlinks, URLs, encodings, as well as special mark-up language commonly used in social media platforms such as Twitter are removed. We convert all text to lowercase, remove white spaces and numbers. Special emoticons are disabled in the text. We also remove common words such as ‘and’, ‘is’, ‘the’ which are considered stop words. Lastly, we apply lemmatization, where words are reduced to the root.

We then transform the cleaned unstructured text into structured form thus generating a document-term matrix (DTM) which is used to generate context-based feature sets for modelling context-based sentiment.

Context-Based Feature Generation

Ours is a hybrid model to sentiment analysis, which combines words, word sequences and word embeddings to generate a large text feature set as input to detect context-based

sentiment. For this, we combine 3 text-based feature components to create a comprehensive and context-rich text feature set [8]. Specifically, we combine the following:

Unigrams: A unigram model can be treated as the combination of several one-state finite automata. Here each individual word is a token [9]. The unigram model is based on the Bag-Of-Word (BOW) approach, which is the most basic model for extracting features from text. For this study, we also apply term frequency-inverse document frequency (TF-IDF) which applies a weighted approach to determine the frequency of words.

Bigrams: A bigram, which is also referred to as a digram, is a sequence of two adjacent elements from a string of tokens, which are typically letters, syllables, or words [9]. A bigram is also considered as an n-gram where n=2. Similar to the unigram model, the bigram employs a BOW approach except that instead of single terms, it take two adjacent words. The frequency distribution of every bigram in a string is commonly used for simple statistical analysis of text in many applications, including in computational linguistics, cryptography, speech recognition, and so on [9].

Word2vec: Word2Vec is a word embedding model that utilizes a neural network based approach to generate numerical feature vectors for words in a way that maintains the relative meanings of the words in the mapped vectors. It provides a reliable method to reduce the features in the data, which is one of the biggest challenges in traditional NLP models such as the BOW model where the vector representation is really sparse due to high dimension data [11]. In this study, we use the Word2Vec model to form a multi-dimensional word embedding structure in order to expand on the context of words and thus increase sentiment accuracy. Our Word2Vec models are generated for unigram and bigram models respectively.

In this study, we propose to combine features from the unigram, bigram and Word2Vec to form a context-based feature set to form a hybrid model for sentiment analysis.

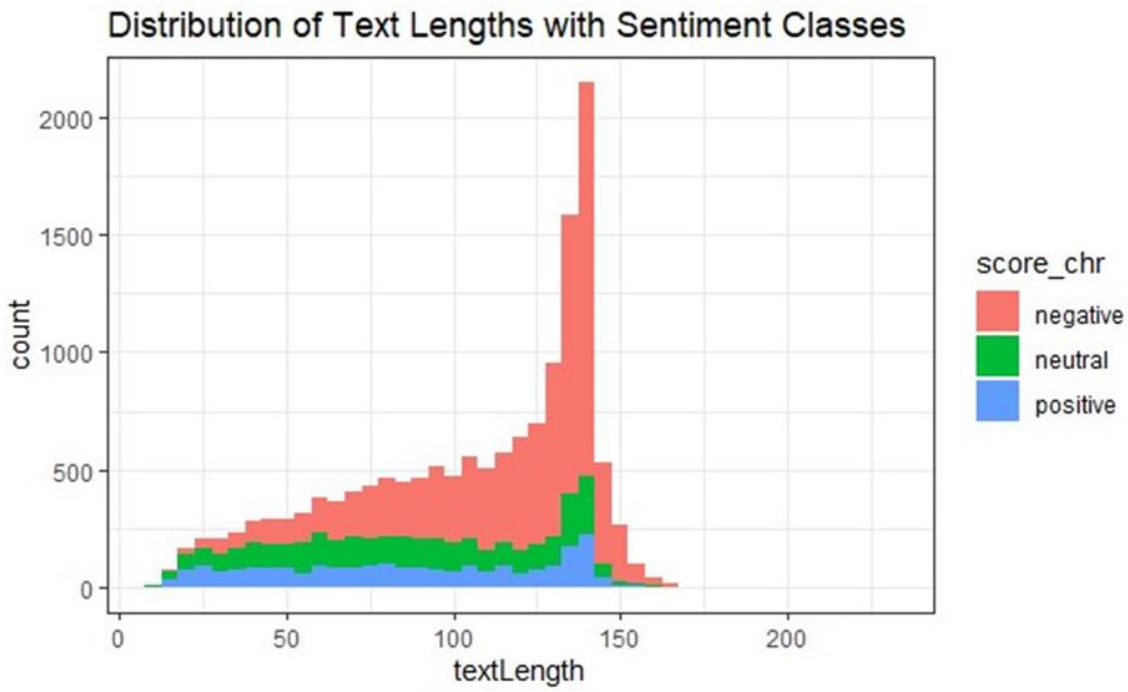
Labelled Data with Sentiment

Our study explores 3 sentiment categories, which include; positive, neutral and negative. The sentiment categories in question are selected because they are popularly applied in sentiment analysis literature. However, for future work, we would like to explore other modes of sentiment such as anger, sadness, and happiness among others.

In this study, we apply our proposed hybrid model on labelled datasets, for which the sentiment is already assigned. Sentiment labels are either determined manually where individuals are selected to provide their opinion on a piece of text. A manual approach to sentiment scoring often is considered a subjective approach and limited by biases from individuals. Alternatively, sentiment labels can be automatically determined based on the overall sentiment score of the words in each document. In the latter scenario, there exist libraries which contain of a score for each term. For example, scores range from 1 to -1 where > 0 is positive, 0 is neutral and < 1 is negative. It should be taken into consideration that the overall sentiment score may also be influenced by the length of the document, which is essentially the number of terms in a document. In Figure 3, we provide an illustration of the sentiment distribution versus the text (document) length.

Figure 3 shows that for this given corpus, there is a significant increase in the negative sentiment across a large number of documents, particularly as the text size (length) increases.

Figure 3: Distribution of Text Length across Sentiment Classes



In the future, we would like to apply our model on unlabeled datasets, where we will explore sentiment scoring techniques to assign sentiment labels.
Nevertheless, we utilize the labelled datasets used in this study to train our model based on the context-based feature set and evaluate the proposed hybrid model across multiple datasets.

Model Evaluation
The word vectors were trained first by using the traditional skip-gram model, followed by character n-gram information. We modify the Word2Vec model by training the embeddings associated with the bigrams, and then mapping each word in the vocabulary to its constituent bigram. We fixed the number of bigrams for each word to be 20 in order to facilitate matrix operations in our model. If a word had less than 20 bigrams, we applied zero padding. Otherwise, we considered only the first 20 bigrams if it had more. These mappings were computed and stored in order to speed up the process of training the actual embeddings.

In the model implementation, we consider the word embedding matrices corresponding to the unigram and bigrams respectively. Based on these mappings, the word embedding vectors of unigrams and bigrams were summed with the word-only embedding to form the required word vector representation. Given the large number of dimensions from

combining multiple word embedding features, we apply feature reduction using singular value decomposition (SVD).

In this study, we apply supervised learning on the labelled instances. Given that the datasets are imbalanced where certain classes have significantly larger instances, we apply random undersampling on the majority class which also minimizes bias towards a single class. It should be noted that we also applied random oversampling, however, this resulted in overfitting of the model. Our models are trained (60%) and tested (40%) using random forest, which is an ensemble classifier, although other classifiers can be applied. We evaluate our models based on:

- Accuracy, which measures the number of correctly classified instances [12].
- Sensitivity, which measures how the model is to detect all instances of a given class. It is also referred to as the recall or true positive rate [12].
- Specificity, which measures the preciseness of the assignment of instances to a given class [12].

RESULTS

In this section, we discuss our findings. Our experiments are implemented in R Studio, and specifically with R 3.0.1.

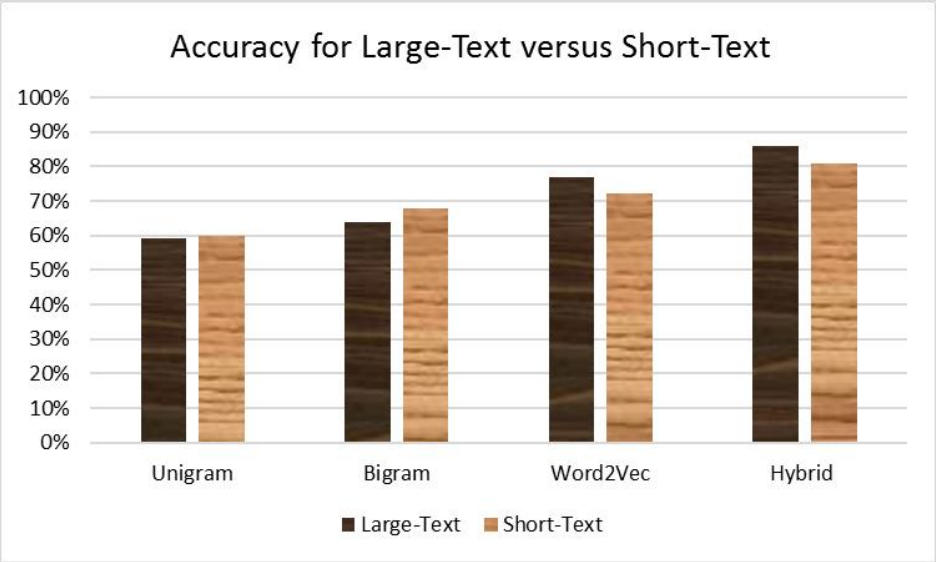
In order to validate our proposed methods, we explore 3 public real-world datasets, specifically;

a) Movies reviews [14]
b) Airline reviews [13]
c) Twitter Tweets [15]

Evaluation of Accuracy on Large-Text versus Short-Text

Here we evaluate the accuracy of hybrid model on short and long-text and compare it against other feature sets discussed in our methods section. To minimize redundancy, we present results on the twitter dataset.

Figure 4: Evaluation of Accuracy for Large-Text versus Short-Text

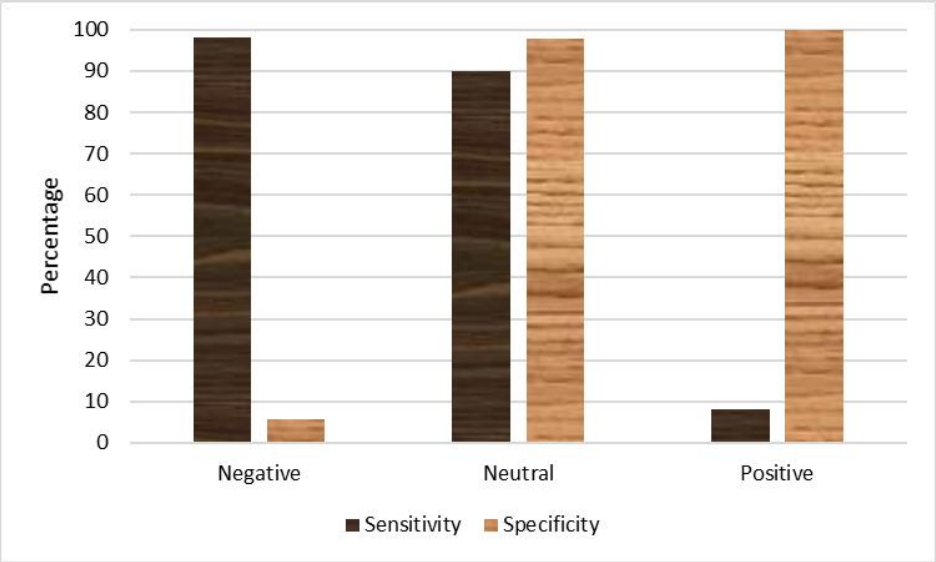


Our findings in Figure 4 indicate that the proposed hybrid model outperforms all other models as indicated by an accuracy >= 80%. On the other hand, we observe that the unigram model has the worst performance with < 60% accuracy overall. Interestingly, with large-text, the Word2Vec and Hybrid models also perform better compared to short-text. Our findings confirm that with more contextual-based features such as hybrid and word2vec, the prediction outcomes on sentiments increase significantly.

Evaluation of Sensitivity and Specificity for Hybrid Model across Class Labels

Here we further evaluate the sensitivity and specificity of the proposed model, which is the hybrid model across the specific predicted class labels, that is negative, neutral and positive using the Twitter dataset.

Figure 5: Evaluation of Sensitivity and Specificity for Hybrid Model across Class Labels



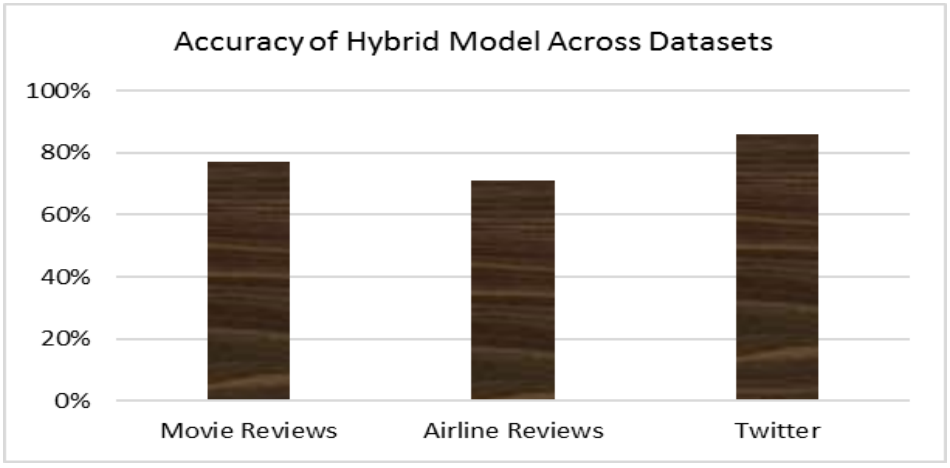
Our findings in Figure 5 indicate that the true positive rate for the negative is very high as indicated by a high sensitivity rate, and a very low specificity or recall rate. On the other hand, we observe reversed behavior in the positive class with low sensitivity and high specificity. The neutral class indicates relatively high sensitivity and specificity.

Evaluation of Hybrid Context-based Sentiment Model Across Datasets

Here we evaluate the accuracy of the proposed model, which is the hybrid model across 3 datasets to test our model's robustness.

Our findings in Figure 6 indicate that Twitter shows higher prediction accuracy > 80% compared to other datasets. On the other hand, Airline reviews show the lowest accuracy. Such variations can be attributed to the common language applied in the twitter comments and thus availability of sentiment libraries to match several of the words in Twitter comments.

Figure 6: Evaluation of Accuracy for Hybrid Model across Multiple Datasets



DISCUSSION AND CONCLUSIONS

In this study, we have proposed a hybrid model to detect predict sentiments by utilizing a context-based approach. Our proposed model applies a context-rich feature set generated from integrating multiple words, word sequences and word embedding structures. The proposed hybrid model takes into account the importance of context in objectively applying sentiment analysis. Additionally, it also employs efficient approach that leverages deep learning methods making it suitable for large unstructured text data. Our findings show that our hybrid model outperforms all other models with an accuracy > 80%. We also demonstrate that our model is applicable to both short and large text and is adaptable to multiple datasets from various domains in predicting meaningful sentiments. For our future work, it would be interesting to embed topic modelling to further refine the context based on latent thematic structures contained in the data.

REFERENCES

1. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

3. Kanakaraj, M., & Guddeti, R. M. R. (2015, February). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015) (pp. 169-170). IEEE.

4. Fang, J. and Ko, A., Medallia Inc, 2018. Context-based sentiment analysis. U.S. Patent 10,037,491.

5. Vanzo, A., Croce, D., & Basili, R. (2014, August). A context-based model for sentiment analysis in twitter. In Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers (pp. 2345-2354).

6. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.

7. Wang, Z., Wu, Z., Wang, R. and Ren, Y., 2015, November. Twitter sarcasm detection exploiting a context-based model. In international conference on web information systems engineering (pp. 77-91). Springer, Cham.

8. Ren, Y., Zhang, Y., Zhang, M., & Ji, D. (2016, February). Context-sensitive twitter sentiment classification using neural network. In Thirtieth AAAI Conference on Artificial Intelligence.

9. Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Information Fusion, 52, 344-356.

10. Jiang, H. Sentiment Analysis on Imbalanced Airline Data.

11. Khandelwal, A., Swami, S., Akhtar, S. S., & Shrivastava, M. (2017, September). Classification Of Spanish Election Tweets (COSET) 2017: Classifying Tweets Using Character and Word Level Features. In IberEval@ SEPLN (pp. 49-54).

12. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

13. Airline sentiment dataset url: https://www.kaggle.com/crowdflower/twitter-airline-sentiment

14. Sentiment dataset on Movie Reviews url: https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews>

15. Sentiment dataset on Twitter Tweets url: https://www.kaggle.com/kazanova/sentiment140

DECISION SCIENCES INSTITUTE
Drivers of Big Data Capability

Canchu Lin
SUNY Farmingdale
Email: linc@farmingdale.edu

Anand Kunnathur
University of Toledo
Email: Anand.Kunnathur@utoledo.edu

ABSTRACT

Prior and current research has articulated the importance of developing big data related capabilities and further explored performance benefits associated with them. This study aims to investigate plausible factors that contribute to the development of big data capability. Drawing on multiple theories, this study builds a model that consists of new product development, knowledge cocreation, and relationship building as antecedents of big data capability. Possible findings and their contributions are discussed with respect to theory and practice.

KEYWORDS: Big Data capability, New product development, Knowledge cocreation, and Relationship building

INTRODUCTION

The popular and business press has well captured the role of big data in creating business value (e.g., see Barton & Court, 2012; Carr, 2013; Pigni, Piccoli, & Watson, 2016; Sanders, 2016). To extract this value, firms must develop big data analytics (Grover, Chiang, Liang, & Zhang, 2018; Mikalef, Boura, Lekakos, & Krogstie, 2019; Wamba et al., 2017), deploy supporting technologies and infrastructure (Wamba et al., 2017), and invest in a combination of human, financial, and intangible resources as foundations sustaining their analytics (Mikalef et al., 2019). Through diagnostic analysis of big data cases and theoretical inference, researchers (e.g., Chen, Preston, & Swink, 2015; Grover et al., 2018; Gupta & George, 2016; Wamba et al., 2017) argued that it is important for firms to develop big data related capabilities in order to enhance their performance. Thus, prior and current research is primarily concerned with exploring performance benefits associated with these capabilities. However, just because they bring about these benefits, it is equally important for us to investigate what helps to foster the growth of big data capabilities.

To address this research gap, this study is to examine whether firm performance, more specifically, new product development (NPD), together with two environment management concepts, supply chain knowledge cocreation (KC) and relationship building (RB), contribute to developing big data capability (BDC), a generic term encapsulating all big data related capabilities. These four concepts will be defined respectively in the next section. With this research objective, the current study aims to make three contributions to the big data research literature: updating our understanding of the relationship between firm performance and big data capability, revealing the importance of building an environment conducive to the development of BDC, and empirically supporting the view that BDC development is a result of learning from other dynamic capabilities.

The rest of the paper is structured as follows. We will review the literatures of big data research and the other three concepts and then make some propositions regarding their relationships. Lastly, we will discuss how this study contributes to theory and practice.

LITERATURE REVIEW AND PROPOSITIONS

We will draw on multiple theories including the resource-based view (RBV) (Barney, 1991), the dynamic capabilities (Teece, 2007; Teece, Pisano, & Shuen, 1997), strategic orientations (Day & Wensley, 1983; Day, 1994), and co-development of multiple dynamic capabilities (Bingham, Heimeriks, Schijven, & Gates, 2015) to introduce the four concepts involved in this study and their corresponding relationships, which constitute the research model of this study.

Big Data Capability

Exploring possible factors of organizational success and sustainability, strategy research has pointed to resources (Barney, 1991), knowledge (Grant, 1996), and organizational capabilities (Eisenhardt & Martin, 2000; Helfat & Peteraf, 2003; Winter, 2000; Zollo & Winter, 2002). In our thriving “Age of Data” (Mikalef et al., 2019), extensive use of electronic devices in the workplace as well as private life creates huge amounts of data, now commonly known as big data. Big data is undoubtedly an important resource that can be used to generate knowledge. Yet, knowledge generation relies on organizational capabilities (Gold, Malhotra, & Segars, 2001). Further, organizational capabilities assume usage and deployment of resources and knowledge (Teece et al., 1997).

Among organizational capabilities, it is dynamic capabilities that help organizations to gain and maintain a competitive advantage (Teece et al., 1997; Teece, 2007), as they are path-dependent and firm-specific (Collins, 1994). Overall, dynamic capabilities are congruent with organizational strategic direction (Teece, Peteraf, & Leih, 2016). Recent research shows that organizational big data practices exhibit such characteristics of dynamic capabilities.
For example, Pigni et al. (2016) found that a full package of resources, skills, competencies, and cultural values involved in big data practices in organizations can best be characterized as a collection of dataset, toolset, skillset, and mindset.

First, big data is a dataset. More specifically, big data refers to large volumes of various types of data (McAfee & Brynjolfsson, 2012) placed in different organizational information systems, and unconnected in structure (Daveport & Patil, 2012). Similarly, big data also flows out as real-time data streams (Pigni et al., 2016). Besides the raw data, other types of resources used in big data practices include hardware, software, and complex IT infrastructure that enable collection, storage, transformation, and analysis of big data. This is the toolset. Further, big data practice requires the attainment of data analytics skills, a knowledge making competency (Chen et al., 2015). This constitutes the skillset. Even though this is not an exclusive list, these features of big data are enough to show that big data is not a sole technology but a multiplicity of interactive elements. Most important, there is a mindset that helps to hold these interactive elements together (Pigni et al., 2016), something comparable to a spirit. This mindset underlying the big data multiplicity is close to what is commonly understood as organizational strategy.

After analyzing interview data collected from seven large US firms and other cases, Lin, Kunnathur, and Li (2020) found that those characteristics have been clearly revealed in big data practices in organizations. Given these characteristics, and in alignment with the dynamic capabilities perspective, a predominant view emerged in the literature is that big data

undertakings must be conceptualized as a dynamic capability (Braganza et al., 2017; Chen et al., 2015; Grover et al., 2018; Gupta & George, 2016; Mikalef et al., 2019; Wamba et al., 2017), which is further defined as "a firm's dynamic capability of identifying sources where large volumes of various kinds of data flow out in high speed and collecting, storing, and analyzing such big data for the purpose of accomplishing the firm's strategic as well as operational goals" (Lin et al., 2020, p. 216).

New Product Development Contributing to Big Data Capability Development

Success stories of organizations implementing big data initiatives have been carried out in the popular press. Some of them are cases of NPD. For example, benefited from big data analysis that powerfully revealed viewers' tastes such as favorite actors and actresses, Netflix created and marketed a new TV series, *House of Cards*, that brought them millions of dollar revenue (Carr, 2013). Similarly, exploiting the potential business values of real-time flow of digital data streams emanated from big data motivated the creation of Uber (Pigni et al., 2016). These cases can easily lead us to reason that BDC contributes to NPD. Empirical research has seemingly supported this reasoning. For example, Mikalef et al. (2019) empirically demonstrated that big data analytics capability facilitates firm innovation. NPD is one exhibition of firm innovation, as it is a process of generating something new, which is the core idea of innovation.

However, these same cases can also be used to make a different argument that because these firms are highly innovation oriented (as revealed in their preoccupation with developing new products), they are most likely to develop BDC, which, they see, will greatly enable their NPD. This strikingly different argument has theoretical groundings. As the literatures of strategic management, entrepreneurship, and marketing strongly suggest, over time, a firm develops a mindset or mentality that guides its strategy formulation and application, and decision making on creation and deployment of critical resources (Bettis & Prahalad, 1995; Prahalad & Bettis, 1986). This mindset or mentality is generically named as strategic orientation (Day & Wensley, 1983; Day, 1994). One such strategic orientation is called entrepreneurship orientation, which refers to a firm's preference and display of innovativeness, proactiveness, and risk-taking for the purpose of entering new markets (Covin & Slevin, 1989). Lin and Kunnathur (2019) already found that entrepreneurship orientation positively contributes to BDC.

NPD, in many ways, demonstrates such entrepreneurship orientation. NPD is a performance concept, which is defined as an ability as well as a process of innovatively translating a business idea into a manufacturable product that will generate business value, involving a series of activities such as assessing market opportunities and technical possibilities, building and testing a prototype, and launching it at the market (Doll, Hong, & Nahm, 2010; Kourfteros, Cheng, & Lai, 2007; Pavlou & Al Sawy, 2006; Perols, Zimmermann, & Kortmann, 2013; Schoenherr, Griffith, & Chandra, 2014). Based on its definition, NPD can best represent entrepreneurship orientation.

First, NPD involves product innovation in the fullest sense (Song, Montoya-Weiss, & Schmidt, 1997; Szymanski, Kroff, & Troy, 2007). Next, launching a new product in a market means that the firm anticipates customer needs and acceptance. In this sense, it is proactive. Further, NPD can be a failure (Heidenreich, Kraemer, & Handrich, 2016; Ogawa & Piller, 2006). This is especially so when the product belongs to a new or emerging product category and thus customers do not understand it enough (Feiereisen, Wong, & Broderick, 2008). Apparently, NPD involves a lot of risk of failure. Riskiness increases with strategic emphasis and product

complexity inherent in NPD (Sharma, Saboo, & Kumar, 2018). Indeed, entrepreneurship orientation is positively related to NPD (Chen, Lin, & Tsai, 2020; Davis, Morris, & Allen, 1991). Given its affinity to entrepreneurship orientation, we have ample reason to argue that NPD will be positively related to BDC.

Additionally, the literature on customer knowledge management in the process of NPD tends to enhance our argument. It has been well recognized that NPD is also a process of gaining and managing knowledge about customers (Cooper, 2014; Zhang, Zhou, Lu, & Chang, 2017). More specifically, it is critical to NPD that a firm uses effective approaches, mechanisms, and devices to collect, store, access, and analyze data about customers (Fidel, Schlesinger, & Cervera, 2015; Olson, 2018). Firms must extract information about customer preferences and requirements for products from their raw data (Joshi & Sharma, 2014; Shimomura, Nemoto, Ishii, & Nakamura, 2018), and then further refine them as firm knowledge about customer consumption. To do that, firms can rely on techniques such as data mining, which, according to Zhan, Tan, and Huo's (2019) review, is extensively used in NPD. As the customer knowledge management part of NPD is so heavily data-driven, firms that engage themselves in NPD should have a high proclivity toward BDC. This short discussion of the strategic orientation and more specifically of entrepreneurship orientation literature and the literature on customer knowledge management in NPD leads us to make the following proposition:

P1: NPD promotes BDC.

Synergistic Effect of Knowledge Cocreation and Relationship Building on Big Data Capability

Turbulent environment calls for the execution of organizational capabilities (Girod & Whittington, 2017; Karna, Richter, & Riesenkampff, 2015), especially dynamic capabilities (Teece et al., 1997). Today's business environment is marked with constant changes caused by technology, competitors, regulatory events, economic and political conditions, and other factors (Tushman & O'Reilly, 1996). To keep their competitive advantage sustainable, firms have to develop an ability to respond, manage and even benefit from these revolutionary changes (Tushman & O'Reilly, 1996). To do that, firms must develop and maintain positive relationships with other firms. To start with, firms must establish such strong relationships with their supply chain partners. This strongly suggests that environment management, more specifically, relationship building with supply chain partners, benefits the development of BDC.

Additionally, according to the theory of resources dependency, firms find themselves in need of resources for their development and operations, and therefore must interact with other firms for the purpose of complementing each other in resources (Barringer & Harrison, 2000; Scott, 1987). More specifically, firms must develop interorganizational relationships with other firms so as to reduce resource dependence (Mitchell & Singh, 1996). Therefore, in the case of developing BDC, firms must demonstrate a capability of reaching out to other firms such as suppliers and customers. This is termed as relationship building that is defined as the ability of developing a relational tie with suppliers and customers, which is characterized by trust, commitment, collaboration, and long-term orientation (Griffith, Harvey, & Lusch, 2006; Johnston, McCutcheon, Stuart, & Kerwood, 2004; Monczka, Petersen, Handfield, & Ragatz, 1998; Tangpong, Michalisin, & Melcher, 2008).

BDC, once developed, can create knowledge for firms, such as knowledge about customers, and knowledge about markets. However, the process of developing BDC itself relies on

knowledge as well. The theory of resource dependency also helps to illuminate this idea, as knowledge is an important type of resource. Our earlier description and definition of BDC imply that this capability is characterized by the ability to produce and transfer knowledge, which is reified in the ability to identify sources of data, process them into information, transform processed information into knowledge, and then communicate knowledge at the right time and to the right party. Thus, another capability that a firm should demonstrate while developing BDC is one of creating knowledge. The RB capability paves the way for developing the knowledge creating capability, as it enables firms to collaborate with other firms to co-create knowledge. In this study, knowledge cocreation is defined as a process in which explicit and tacit knowledge regarding the market, customers, and technologies is collaboratively produced through the four sub-processes of socialization, externalization, combination, and internalization in the supply chain.

Further, the dynamic capabilities research literature shows that multiple dynamic capabilities can co-develop in firms. Ethiraj, Kale, Krishnan, and Singh (2005) found that multiple capabilities are developed as a result of: a) tacit accumulation of experience embedded in routines and learning by doing (Nelson & Winter, 1982; Winter, 1990), and b) deliberate investments in organizational structure to make constant improvements in those routines and processes (Zollo & Winter, 2002). These findings suggest that firms can develop multiple capabilities through concurrent learning that is made possible by codification of experience facilitated by organizational structure (Bingham, Heimeriks, Schijven, & Gates, 2015). Following previous studies that treat performance concepts as dynamic capabilities (see, for example, Grewal and Tansuhaj (2001) treating marketing orientation and strategic flexibility as dynamic capabilities in marketing research), we can view RB and KC as two such dynamic capabilities. In line of this theoretical discourse on capabilities, we argue that KC and RB capabilities, once well established, can facilitate the development of BDC, as firms can apply their learning experiences accumulated from those two cases to that of BDC.

This short review of these three threads of theoretical literature reveals that RB and KC have a contributing effect on BDC. Therefore, we propose the following:

P2: RB facilitates the development of BDC.
P3: KC facilitates the development of BDC.

Moderating Effect of Relationship Building and Knowledge Co-creation Capabilities

According to Teece (2007), dynamic capabilities must have the three functions of sensing, seizing, and reconfiguring. The sensing function is to be alert to threats as well as opportunities, seizing to deploy and utilize resources and assets to address those opportunities, and reconfiguring to transform, redeploy, and redevelop organizational competencies for continuous renewal. Viewed from the dynamic capabilities perspective, dynamic capabilities emerge as both means and outcome of responding to and addressing the turbulent business environment (Girod & Whittington, 2017; Karna, Richter, & Riesenkampff, 2015), Teece et al., 1997). Developing new products is a typical response to the turbulent environment, which calls for dynamic capabilities such as BDC. Specifically, changes that constitute the turbulent environment usually foretell new market trends that call for NPD. BDC serves the sensing role in identifying market requirements and potentials so that firms can enjoy first-mover advantages. Then it enacts seizing by transforming the requirements, potentials, and opportunities into concrete products.

Next, coping with changes in the turbulent environment requires exercising the reconfiguring function of dynamic capabilities. Reconfiguring is sparked by the need of the firm to take action to renew and reorganize its resources and competences to address unique needs stemming from the changing environment (Teece et al., 1997). In the same way it enacts organizational sensing and seizing, BDC serves the reconfiguring function as well. In the business sense, understanding environment changes is, to a large extent, to gain knowledge of competition and its causes. BDC involves use of big data management technologies and mastery of analytics procedures and methods to collect and analyze information (buried in big data) about competitors. Then knowledge of competition generated through the operations of BDC serves to guide the firm for how to reorganize and reconfigure its resources, assets, and competencies so as to maintain its competitive advantage over its competitors. Given these benefits, while engaged in NPD, firms are highly motivated to develop BDC.

While benefiting NPD, the sensing and seizing functions of dynamic capabilities can also enhance strategic alliance development (Schilke & Goerzen, 2010). Forming strategic alliances, especially in supply chains, is an effective way of obtaining required resources such as knowledge that are lacking in the firm (Das & Teng, 2000), but is highly needed in NPD. Thus, to summarize, the theorizing of the role of the environment in developing dynamic capabilities in the dynamic capabilities literature strongly suggests that in developing new products as an active way to address the environment, firms are likely to develop BDC, and developing relationships with their supply chain partners facilitates this process.

In addition, the resource dependency theory can also be applied to NPD. While developing new products, firms require acquisition of external resources. Knowledge is an important resource to NPD in many ways. For example, market knowledge, which includes knowledge about prospective customers and competitors, as well as other market aspects to a given product type (De Luca & Atuahene-Gima, 2007), is crucial at each stage of NPD (Dabrowski, 2019). Thus, a market knowledge competence will facilitate NPD (Atuahene-Gima & Wei, 2011; Claudy, Peterson, & Pagell, 2016; Li & Calantone, 1998). Knowledge about consumer tastes, likes, and needs is a prerequisite to NPD. Similarly, consumer knowledge about the new product is critical to adoption of the new product (Ackermann, Teichert, & Truong, 2018). These research findings regarding knowledge's importance to NPD suggest that the KC capability would enhance NPD's effect on BDC. Besides, NPD requires exploiting external knowledge sources (Ferreras-Mendez, Newell, Fernandez-Mesa, & Alegre, 2015). This is where the RB capability comes into play. As both capabilities are enhancements of NPD, we can expect that they will positively moderate the relationship between NPD and BDC. Thus, the following propositions can be formulated:

H4: RB capability positively moderates NPD's effect on BDC.
H5: KC positively moderates NPD's effect on BDC.

CONCLUSION

Numerous reports of big data creating business value in the popular press has motivated scholarly exploration of this issue. In this newly opened research frontier, more and more participating researchers (e.g., Braganza et al., 2017; Chen et al. 2015; Grover et al., 2018; Gupta & George, 2016; Mikalef et al., 2019; Wamba et al., 2017) articulated a view that firms should develop big data related dynamic capabilities if they want to secure success in their big data practices. This view must be supported with empirical evidence. Thus, prior and current big data research has primarily focused on investigating performance benefits associated with big data capabilities. However, as a point of departure, this study makes a shift in research direction

by exploring plausible factors that may contribute to the development of big data capabilities. In doing that, this study reviewed the theories of dynamic capabilities, strategic orientation, and resource dependency with respect to how they are relevant to and effective in guiding the introduction of the four concepts and their corresponding relationships that constitute the research model of this study. As a result of reviewing the multiple threads of literature, we articulated five propositions regarding the relationships of the four concepts.

With this research design, this study aims to make three contributions. First, it points to a new direction in big data research by switching our attention from primarily on seeking to understand what contributions big data can make to the firm to simultaneously investigating what will be helpful to developing BDC. This study will substantiate such efforts by showing whether NPD will contribute to development of big data capabilities. Second, this study will extend our understanding of what factors help to develop BDC to whether contextual factors facilitate that process. More specifically, this study will show whether supply chain knowledge co-creation and RB, besides NPD, will contribute to development of BDC, but more importantly, whether they moderate the impact of NPD on BDC. As knowledge co-creation and RB are two organizational capabilities, the findings of this study will help to demonstrate a theoretical point about dynamic capabilities articulated in the literature that multiple capabilities can be developed, and they combine to show a synergistic effect on developing dynamic capabilities. Third, this study will inform big data practitioners of what could help them to effectively develop big data capabilities. Further, it could help them in their decision making on whether they should invest in big data projects with findings informing them on an assessment of their strategies, resources, strengths, and weaknesses.

REFERENCES

References will be available upon request.

DECISION SCIENCES INSTITUTE

Spatial Characterization of Rare Disease Events: A Case Study of the Ebola Virus Disease in West Africa

Esha Kapoor
University of Massachusetts, Boston
Email: esha.kapoor001@umb.edu

Josephine Namayanja
University of Massachusetts, Boston
Email: Josephine.Namayanja@umb.edu

ABSTRACT

This paper aims to characterize rare disease events such as the Ebola Virus Disease (EVD) outbreak in West Africa. Specifically, we propose to utilize cluster analysis, an unsupervised technique that groups data instances based on similarity to analyze confirmed cases and deaths associated with EVD across macro and micro-spatial regions. Our study also examines the role of socio-economic factors in the EVD outbreak based on spatial regions. Overall, this study poses implications in understanding the importance of characterizing the rapid spread of diseases in order to highlight potential control measures to curb future disease spreads.

KEYWORDS: Clustering, K-Means, Sum of Squared Errors, Ebola, Outbreak, Sierra Leone, Guinea, Liberia, Confirmed cases, West Africa, Variation, Deaths, Spatial

INTRODUCTION

The Ebola Virus Disease (EVD) first surfaced in 1976 near the Ebola River in the Democratic Republic of Congo [12]. Since then, the virus has developed intermittently and has infected thousands of people worldwide and killing a lot of them. Specifically, countries in the region of West Africa have been greatly impacted by EVD where 11,324 people died during the period of 2014-2016, thus marking it as the largest and longest EVD outbreak in medical history [13]. EVD is a highly contagious disease with a mortality rate of 50% and is particularly is most common in Sub-Saharan Africa [23]. One of the key factors identified in the significant outbreak of EVD in the region of West Africa is community spread facilitated by people-to-people interactions both intra-national within countries and inter-national across countries in the region.

This research aims to characterize the EVD outbreak across spatial regions in West Africa. For this, we propose to utilize cluster analysis, an unsupervised technique that groups data instances based on similarity. Clustering is widely utilized in various applications such as recommendation engines, market segmentation, social network analysis, search result grouping, and anomaly detection, among others. In healthcare, clustering is widely applied in medical imaging and image segmentation, disease detection, prediction of outbreaks and clinical research [3], to mention a few. The accelerated growth in data-driven spatial modeling over the last three decades has contributed to significant advances in the convoluted modeling approaches and substantial improvements in data availability and computational capacity [1]. The application of spatial pattern analysis in characterizing disease behavior poses potential by providing valuable clues to disease

ecology, such as understanding the direction and distance of disease spread, as well as importance, and proximity of the sources of a virus) [9].

In this paper, our aim is to study similarities and differences in the number of confirmed cases and deaths associated to EVD in highly impacted countries in West Africa, specifically in Guinea, Sierra Leone, and Liberia during the critical 12-month time period between 2014 - 2015. Hence, we conduct a macro-spatial region analysis where we study EVD confirmed cases and deaths across the countries in question in order to identify any similarities as well as dissimilarities across countries. Similarly, we drill down into each country to study EVD behavior in the counties and towns within each country, which we refer to as micro-spatial region analysis.

Further, according to World Bank [28], the countries in question are classified as developing economies based on their poor economic performance, high poverty level, lack of critical infrastructure, and high inequality of wealth and access to quality education. Therefore, we extend this study to understand the role of socio-economic factors in EVD outbreak both in the macro-spatial regions where we examine the countries and within micro-spatial regions where we examine intra-country resources. Our assumption is that inadequacies in socio-economic resources such as healthcare centers and skilled workers affected timely response and treatment availability particularly in remote areas. More so, the illiteracy rate in these countries, also in remote areas impacted awareness of best practices when dealing with EVD outbreak such as limiting social interactions or communal engagements.

The rest of the paper is organized as follows, in section 2 we discuss related work in the literature review, in section 3 we present our methodology, in section 4 we present our experimental results, in section 5 we discuss a summary of our findings and present our conclusion in section 6.

LITERATURE REVIEW

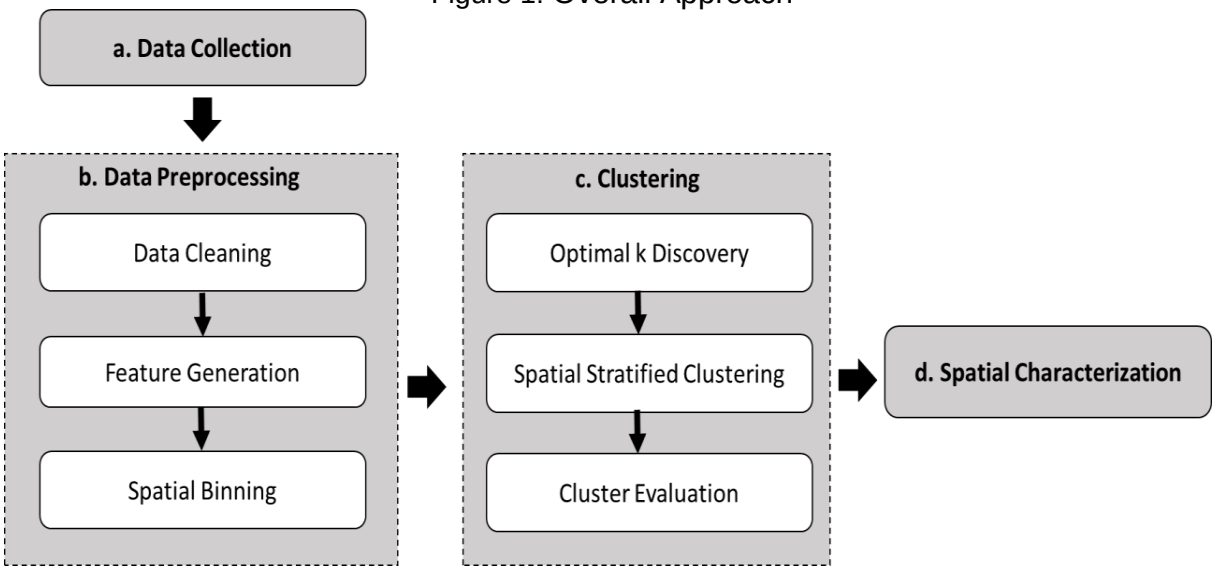
Spatiotemporal dynamics play a crucial role in understanding the importance of characterizing rapid spread of diseases to highlight potential control measures to curb future disease spreads. A study by [7] proposes Bayesian model inference to incorporate individual-level spatial information with other epidemiological data of community-based cases and to unequivocally deduce the distribution of the cases generated by each infected individual. Their research shows that super spreaders play a crucial role in supporting the transmission of epidemic diseases among the community often within a relatively short spatial distance (with a median value of 2.51 km) and thus being responsible for a significant proportion (~61%) of the infections compared to cases contained under supervised clinical care.

A study by [6] compared several models to characterize the spatial network of EVD and estimated the effects of geographical covariates on transmission during the peak spread period. According to their study outcomes, the best model was a generalized gravity model, which utilizes network analysis to model nodes as geopolitical administrative units and weighted links as the strength of potential infection pathways. Their findings also indicate that the probability of transmission between locations depended on distance, population density and international border closures between highly impacted neighboring countries. In [2], a distance- and area-based spatial statistical methods have been proposed to study disease spread. Findings in this study reveal that clusters of infected cases appear in close proximity. Similarly, our study aims to understand epidemic spread across spatial regions by applying distance-based measures to evaluate the similarity in disease behavior over time.

METHODOLOGY

The purpose of the research is to investigate the factors that led to the abnormally high incidence of Ebola virus in Guinea, Liberia, and Sierra Leone. In Figure 1, we provide an overview of our approach.

Figure 1: Overall Approach



Data Collection

For this study, we collect structured data on cases and deaths associated with EVD from two sources, specifically, i) the Humanitarian Data Exchange (HDX) [16] and ii) World Health Organization (WHO) [24]. The dataset consists of five attributes and 58,600 instances captured across six countries in West Africa. In this study, we select data from 3 countries: Guinea, Liberia, and Sierra Leone, which were significantly affected by EVD. Our study period is 12 months, from March 2014 to March 2015. Further, the data collected is cumulative on a day-to-day basis.
In Table 1, we provide a description of the attributes as used in this study.

Table 1: Summary of Attributes

ATTRIBUTE	DATA TYPE	DESCRIPTION
Country	Nominal	Country Name
Localite	Nominal	Regions, cities, and counties per country
Category	Nominal	New cases, confirmed cases, suspected cases, cases, probable cases, and deaths.
Value	Discrete	Number of EVD incidences based on category.
Date	Interval	Recorded date of EVD incidence

Furthermore, for the attribute ‘Category’, we define the possible values in detail as follows:

- New Cases: Total number of new cases by each day.
- Confirmed Cases: Cumulative cases that were test positive for Ebola Virus Disease (EVD).

- Suspected Cases: Cumulative cases that were in touch with EVD positive person or had a travel history of affected region or had symptoms, but not yet tested.
- Probable Cases: Cumulative cases that were tested for EVD but had inconclusive results.
- Cases: Sum of probable, suspected and confirmed cases.
- Deaths: Number of people who died from EVD.

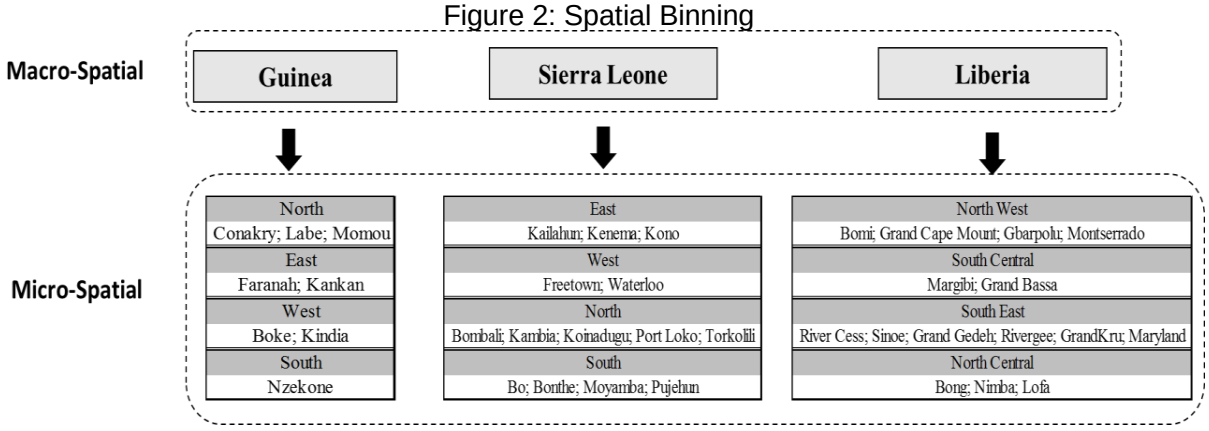
Data Preprocessing

According to WHO, there were 10,148 cases and 10,048 deaths associated with EVD worldwide, from March 2014 to March 2015, for which 99.93% of the cases were recorded in West Africa. As mentioned previously, we select three countries for this study, which include Guinea (46% cases), Liberia (23% cases), and Sierra Leone (29% cases). We consider these as macro-spatial regions. On the other hand, only 1.98% of cases were recorded in Mali, Nigeria, and Senegal combined, and thus, these countries were eliminated from this study.

Next, we conduct data cleaning, where we identify and eliminate any records associated with inconsistent data values. Some of the reasons highlighted for the missing or inconsistent data are human error, equipment error, incorrect measurements, undocumented information, or misinformation during the EVD incident reporting process. We also perform feature generation based on the “category” attribute using one-hot encoding. Specifically, each data value for the category attribute is transformed into a data attribute. Thus, the resulting feature set consists of; new cases, confirmed cases, suspected cases, cases, probable causes, and deaths. To minimize data redundancy, we focus on confirmed cases and deaths.

Finally, we apply spatial binning. First, we examine the confirmed cases and deaths per country, which we refer to as macro-spatial regions. Then we group counties per country, which are described as part of the “localite” attribute into what we refer to as micro-spatial regions. In order to minimize data redundancy, we eliminate towns, cities, and regions per country. Therefore, each micro-spatial region represents a spatial bin capturing confirmed cases and deaths, respectively, over time. In Figure 2, we illustrate the macro and micro-spatial regions where we map counties to regions per country.

Figure 2: Spatial Binning



It should be taken into consideration that the number of counties per region varies. For this study, such variations are essential in characterizing EVD behavior across spatial regions.

Furthermore, given that EVD cases are captured cumulatively over time, the cases recorded in each country and region vary. Some regions have missing case counts based on the progression

of EVD over time. Therefore, to maintain consistency, we apply temporal reconciliation for case counts across all spatial regions. For example, if Guinea reported its first confirmed case on March 28th, 2014, then on April 10th, 2014, confirmed cases were reported in Liberia while Sierra Leone had its first confirmed case on May 25th. To maintain consistency, we take March 28th, 2014, as the first date and record 0 for Liberia and Sierra Leone, respectively, until the dates correspond to their first confirmed cases.

On the other hand, where no cases are reported on subsequent dates following the first case(s), we apply the value of the preceding date to show consistent case/death count for that given time period. For example, if Guinea reported 5 cases on March 28th, 2014, and 7 confirmed cases where reported on April 1st, 2014, we apply 5 confirmed cases for March 29th to March 31st, respectively, to indicate no change in case count. We apply the same temporal reconciliation for deaths and across all macro- and micro-spatial regions.

Clustering

In this study, we utilize cluster analysis using k-means clustering to characterize the similarities and variations in EVD cases and deaths in West Africa. K-means clustering is the most popular and straightforward unsupervised method that groups data distances into clusters based on distance metrics such as Euclidean distance [4]. To group data instances into clusters, the optimal number of clusters k is pre-defined. In this study, we apply the elbow method [5] to determine k, which is based on evaluating the sum of squared errors (SSE) of clusters. SSE is defined as the sum of squared differences between each data observation (xi) and a mean of all the observations ($\bar{x}$). It is a measure of variation within and across clusters. A lower indicates higher similarity (lesser discrepancy) within a cluster. In this study, we determine k for each macro-spatial region (i.e. per country) in order to account for the spatial variations. Figure 3 shows the optimal k discovery for each country.

Figure 3(a): Optimal k Discovery for Guinea

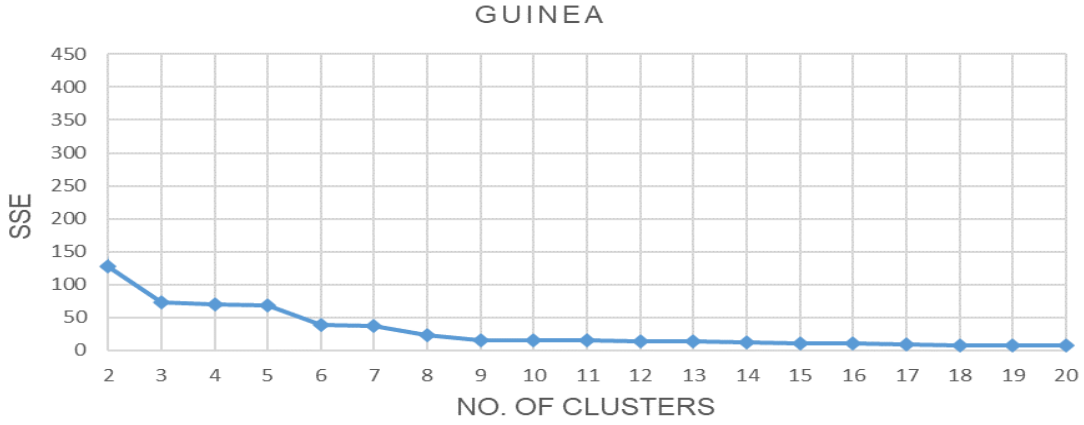


Figure 3(b): Optimal k Discovery for Liberia

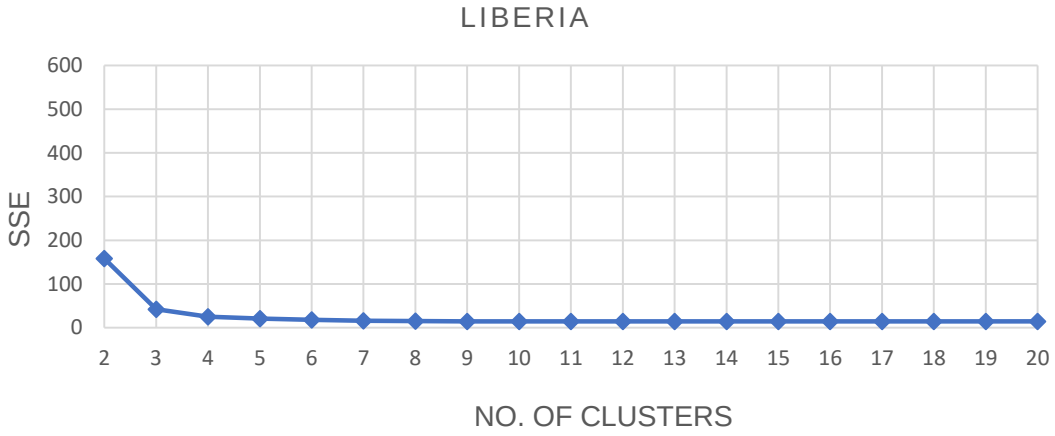
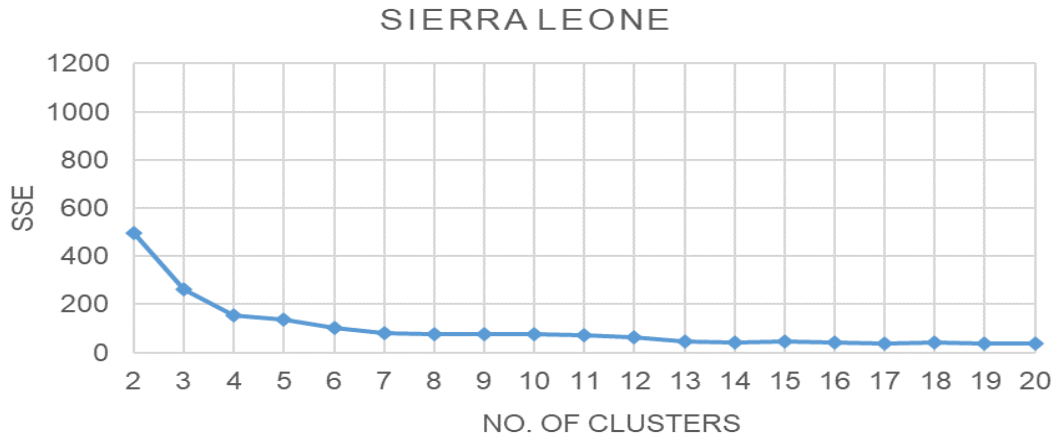


Figure 3(c): Optimal k Discovery for Sierra Leone



According to Figures 3(a), 3(b) and 3(c) optimal *k* values are summarized in Table 2.

Table 2: Summary of Optimal *k* Per Macro-Spatial Region (Country)

MACRO-SPATIAL REGION	OPTIMAL K
Guinea	3
Liberia	3
Sierra Leone	4

Using the optimal k, we then apply k-means clustering on EVD confirmed cases and deaths respectively within each micro-spatial region per country. Our objective is to characterize the similarities in the confirmed cases and deaths across spatial regions.

Spatial Characterization

We extend our analysis, to study the socio-economic factors associated to the macro- and micro-spatial regions. Our objective is to further explain the role of socio-economic factors such as education, health amenities and population in the spread of rare disease events such as EVD. For this, we examine various sources such as Government of Sierra Leone Ministry of Health [15], Liberia Institute of Statistics & Geo Information Services (LISGIS) [18], Ministry of Plan and Economic Development in Guinea [20], Statistics Sierra Leone (Stats SL) [21], World Health Organization (26), World Health Organization: Global Health Observatory data repository [29],

and United Nations [30]. In Table 3, we provide a summary of socio-economic factors examined in this study. The data collected corresponds to the EVD period of 2014-2015 to properly investigate possible associations with the outbreak in those countries at that time.

Table 3: Description of Appended Attributes

SOCIO-ECONOMIC FACTOR	DATA TYPE	DESCRIPTION
Population	Discrete	Number of people residing in a county/region
Unemployment Rate	Ratio	Percentage of population unemployed
Literacy Rate	Ratio	Percentage of population able to read and write
Number of physicians	Discrete	Number of medical physicians per 10,000 habitants

EXPERIMENTAL RESULTS

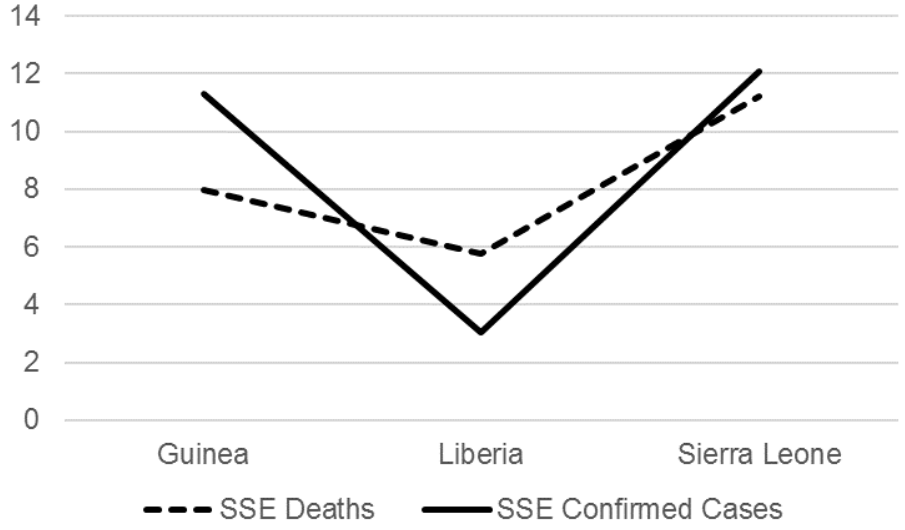
In this section, we provide a comprehensive analysis of the findings. All clustering experiments were implemented using Weka [27].

Evaluation of Clusters for Macro-Spatial Regions

Our findings in Figure 4 show high similarity in confirmed cases of EVD in Sierra Leone and Liberia as indicated by lower SSE compared to deaths, which show higher variation as indicated by higher SSE. On the other hand, the number of confirmed cases varies in Guinea as indicated by the higher SSE and higher similarity in deaths as indicated by lower SSE.

Overall, higher variation in both EVD confirmed cases and deaths is observed in Sierra Leone as indicated by higher SSE, while low variation in EVD behavior in Liberia as portrayed by the low SSE values. This further indicates that confirmed cases and deaths for EVD were very similar in Liberia over time across the country compared to other countries.

Figure 4: SSE Analysis Per Macro-Spatial Region (Country)



Evaluation of Clusters for Micro-Spatial Regions

Our findings in Figure 5 (a –c), we observe a general trend of higher similarity among confirmed cases across different regions in each country as indicated by lower SSE values compared to deaths where there is high diversity across regions as indicated by higher SSE.

Figure 5 (a): SSE Analysis per Micro-Spatial Region in Guinea

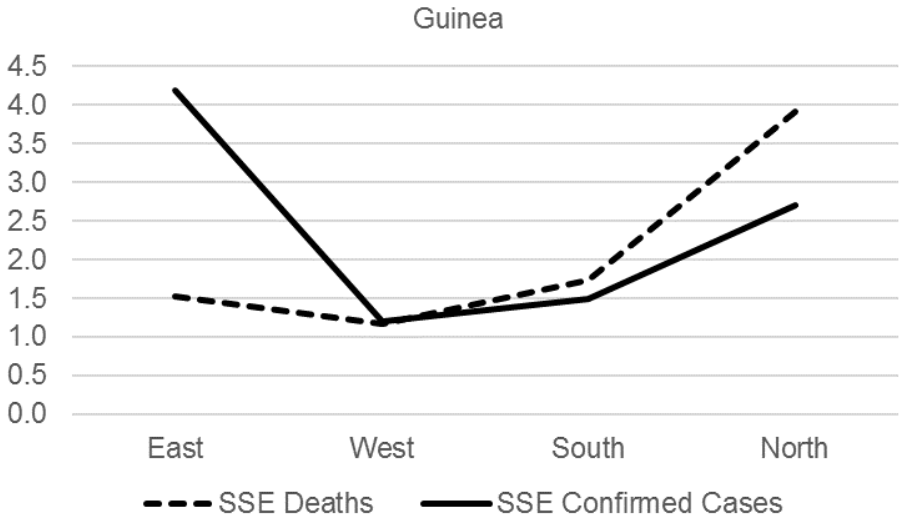


Figure 5 (b): SSE Analysis per Micro-Spatial Region in Sierra Leone

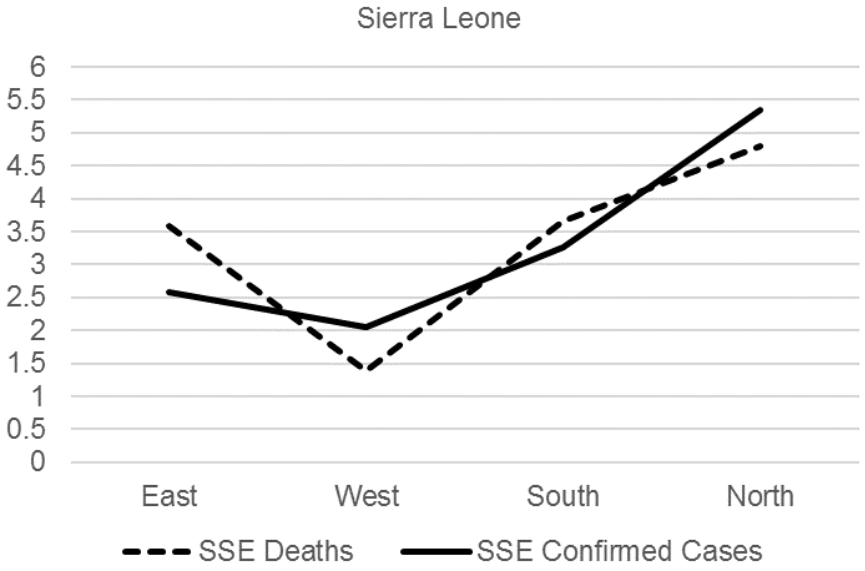
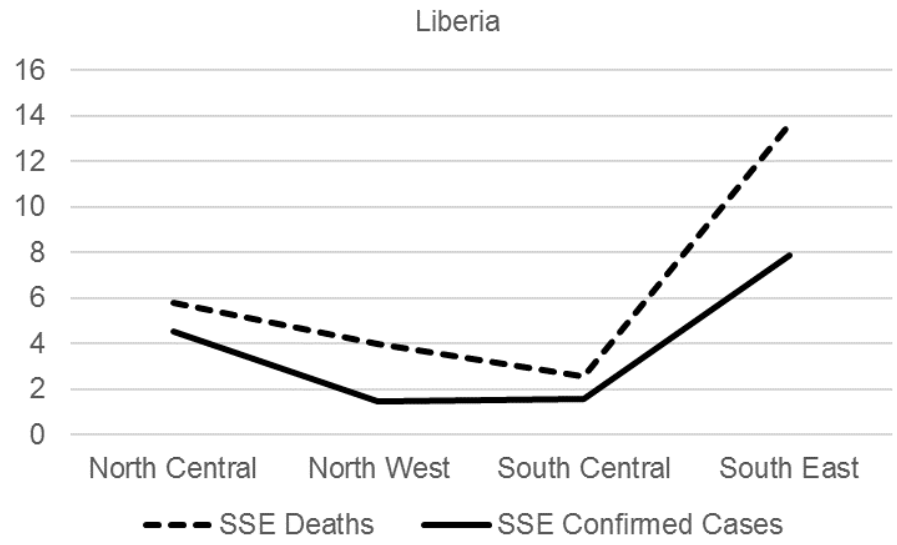


Figure 5 (c): SSE Analysis per Micro-Spatial Region in Liberia



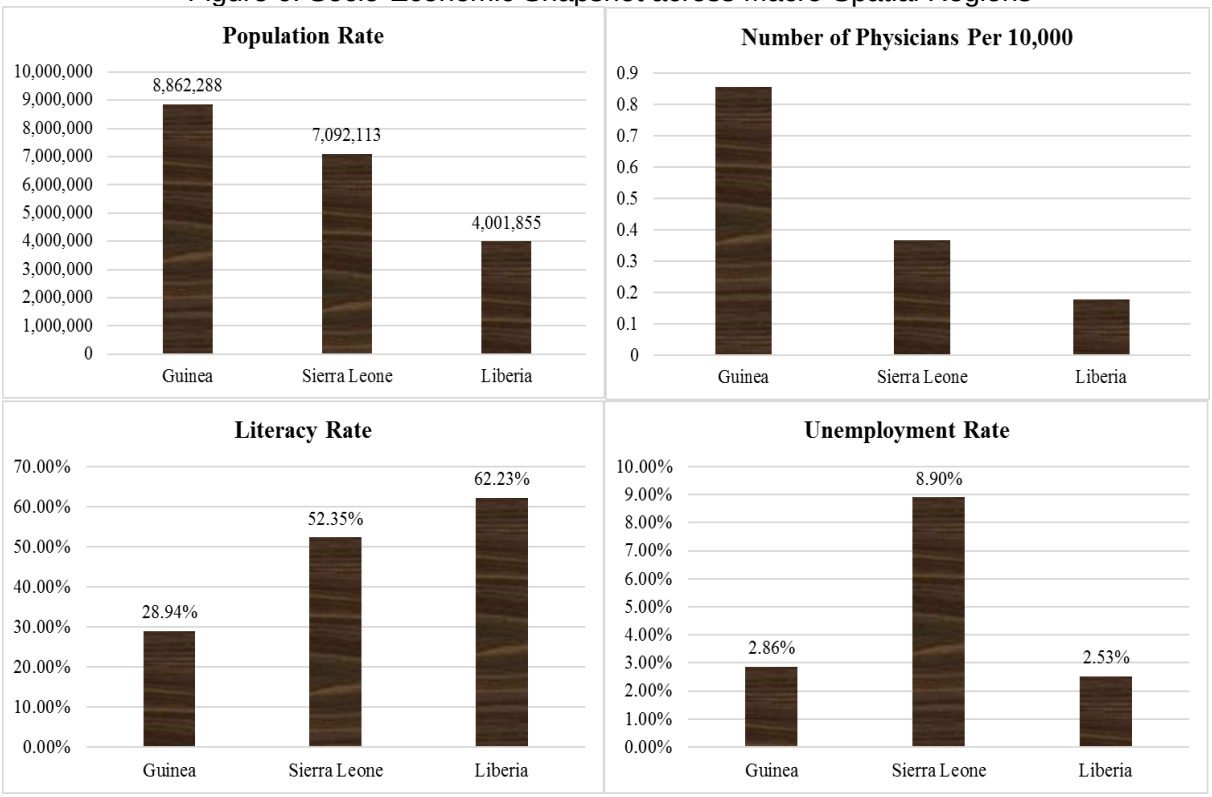
Spatial Characterization based on Social Economic Factors

In this section, we examine socio-economic factors specifically population rate, literacy rate, unemployment rate, and number of physicians per 10,000. These factors are argued to contribute to the significant outbreak of EVD. We examine these socio-economic factors at the macro and micro-spatial bin levels respectively.

Analysis of Socio-Economic Factors based on Macro-Spatial Regions

According to [8,15,17,18,20,21,26], Guinea, Liberia, and Sierra Leone are described as developing economies. In Figure 6, we provide a snapshot of each country based on socio-economic factors.

Figure 6: Socio-Economic Snapshot across Macro-Spatial Regions



According to Figure 6, Guinea has the highest population, which is closely followed by Sierra Leone and Liberia records the least population. In terms of number of physicians per 10,000 habitants, Guinea is highly ranked compared to Sierra Leone and Liberia. For literacy rates, Liberia has the highest literacy rate, followed by Sierra Leone and lastly Guinea. Concerning unemployment, Sierra Leone reports the highest unemployment rate of approximately 3 times higher than Guinea and Liberia. In general, we observe similarities in socio-economic factors across all 3 countries. All 3 countries share a common border, and thus share cultural similarities such as language. This also facilitates cross-migration between the countries. Because of their spatial connectivity, all three countries suffer from negative implications of a prolonged civil war that significantly affected social infrastructure such as limited education centers, leading to high illiteracy rate among young adults. Further, inadequate health care facilities and personnel indicate a ratio of 1:2 doctors per 100,000 persons [25].

Analysis of Socio-Economic Factors based on Micro-Spatial Regions

In this section, we examine the socio-economic factors in each micro-spatial region (bin) within each country in the year 2014, with the exception of number of physicians for which no external data was available.

Table 4 shows that in Guinea, the East region has the highest population, which can be considered as a contributor to the disparity in the behavior of EVD confirmed cases in the East of Guinea as shown in Figure 5(a). Interestingly, there is also high similarity in the behavior of EVD deaths, which highlights consistency of deaths in the region over time. On the other hand, the North region indicates the highest literacy rate and unemployment rate. Figure 5(a) also shows that the North has very high diversity in EVD deaths. The West region, which has the highest

number of healthcare centers in the country according to [20, 26] indicates high similarity in both EVD confirmed cases and deaths. It can be argued that the EVD confirmed cases and deaths followed a similar trend in the West compared to other regions.

Table 4: Analysis of Socio-Economic Factors for Micro-Spatial Regions in Guinea			
MICRO-SPATIAL REGION	POPULATION RATE	LITERACY RATE	UNEMPLOYMENT RATE
South	1,578,068	27.30%	1.40%
East	2,914,091	20.25%	2.30%
West	2,644,483	31.20%	1.45%
North	1,725,646	37.00%	6.30%

Table 5 shows that in Sierra Leone, the North region has the highest population which also matches the highest diversity in the behavior in both EVD confirmed cases and deaths as shown in Figure 5(b). On the other hand, higher literacy rate and unemployment are observed in the West region, where higher similarity in the behavior in both EVD confirmed cases and deaths as shown in Figure 5(b).

Table 5: Analysis of Socio-Economic Factors for Micro-Spatial Regions in Sierra Leone			
MICRO-SPATIAL REGION	POPULATION RATE	LITERACY RATE	UNEMPLOYMENT RATE
East	1,642,370	47.50%	5.90%
North	2,508,201	42.20%	6.50%
South	1,441,308	45.20%	7.90%
West	1,500,234	74.50%	15.30%

Table 6 shows that in Liberia, North Central region has a significant population and the lowest unemployment rate. Based on the EVD analysis, the region also has very diverse behavior in EVD deaths and confirmed cases. On the other hand, South Central, which has the lowest number of healthcare centers compared to other regions in the country according to [18, 26] also has the highest similarity in EVD deaths as shown in Figure 5(c). Such limitations in health amenities can be viewed as a contributor to the similar trend in deaths in the region.

Table 6: Analysis of Socio-Economic Factors for Micro-Spatial Regions in Liberia			
MICRO-SPATIAL REGION	POPULATION RATE	LITERACY RATE	UNEMPLOYMENT RATE
North West	1,626,275	65.65%	3.30%
South Central	496,825	65.90%	3.10%
South East	644,372	60.75%	3.00%
North Central	1,234,383	56.60%	0.70%

DISCUSSION

The emergency response in West Africa during the 2014-15 outbreak indicated that the countries were not prepared for the crisis, despite past EVD outbreaks for four decades. Sierra Leone for instance, had the most confirmed cases as well as deaths although the country, was the latest to report a first case of the outbreak. The inefficiency of emergency response teams such as slow response times for ambulance and burial teams resulted in dead bodies being placed in the community for as long as eight days which significantly increased the spread of EVD [25]. The inadequacy of healthcare centers for example in Guinea and Liberia highlights an urgent need to build more healthcare centers across all regions of the countries.

Given that all 3 countries are connected by national borders, restrictive practices such as travel bans, and increasing awareness among the communities and overall population plays a role in containing the spread of such aggressive rare diseases such as EVD. Additionally, a coordinated communication strategy across governments of all the three countries is necessary to enforce restrictive measures as well as promote community awareness especially in remote areas.

CONCLUSION

In this paper we outlined a spatial-based study to understand the behavior of rare disease events across macro and micro-spatial regions through cluster analysis. We outlined some preliminary results using Ebola Virus Disease in 3 neighboring countries in West Africa, specifically Guinea, Sierra Leone, and Liberia. In this paper, we have also examined the role of socio-economic factors in the outbreak of such rare disease events. However, this is an ongoing study, which aims to extend this work by examining temporal features in disease spread to get a comprehensive analysis of space and time particularly in rapid disease spreads. We also aim to extend this study to other rare diseases such as COVID-19 to determine spatial and temporal behavioral similarities. Moreover, future implications of this study will help key stakeholders such as governments and local communities to identify emergency response strategies and best practices to create awareness on highly contagious, rapid spreading rare diseases.

REFERENCES

1. G Chowell & R Rothenberg, (2018), Spatial infectious disease epidemiology: on the cusp: BMC Medicine volume 16, Article number: 192.

2. Henne, Workneh, and Rush, (2012), Spatial Patterns and Spread of Potato Zebra Chip Disease in the Texas Panhandle, APS Publications, Vol 96, ISSN:0191-2917.

3. Islam, Hasan, Wang, Germack, and Alam (2018), A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining, NCBI nlm PMID: 29882866

4. Jain, Murthy, and Flynn, (1999), Data Clustering: a review, ACM Journals, Vol 31, No. 3.

5. Janeja & Namayanja, (2011). Subspace discovery for disease management: A case study in metabolic syndrome. International Journal of Computational Models and Algorithms in Medicine (IJCMAM), 2(1), 38-59.

6. Kramer, Pulliam, Alexander, Park, Rohani and Drake, (2016), Spatial spread of the West Africa Ebola epidemic, NCBI nlm PMID: 27853607.

7. Lau, Dalziel, Funk, McClelland, Tiffany, Riley, Jessica. Metcalf, and Bryan T, (2017), Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic, PNAS 114 (9) 2337-2342.

8. Rudge & Bosc,(2009),Sustainability factors for water points in Liberia,34th WEDC Intern ational Conference, Addis Ababa, Ethiopia, 2009,WATER, SANITATION AND HYGIENE : SUSTAINABLE DEVELOPMENT AND MULTISECTORAL APPROACHES.

9. Thresh, J. M, (1976), Gradients of plant virus diseases, Ann. Appl. Biol. 82:381-406.

10. BBC News, (2016), Ebola: Mapping the outbreak <https://www.bbc.com/news/world-africa-28755033>

11. Center for Disease Control (CDC), (2016), 2014 Ebola Outbreak in West Africa Epidemic Curves,<https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/cumulative-cases-graphs.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvhf%2Febola%2Foutbreaks%2F2014-west-africa%2Fcumulative-cases-graphs.html>
12. Center for Disease Control (CDC), (2018), History of Ebola Virus Disease <https://www.cdc.gov/vhf/ebola/history/summaries.html>

13. Center for Disease Control (CDC), (2019), 2014-2016 Ebola Outbreak in West Africa< https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.htmll>

14. Dihuni/ Voucher Cloud (2020), Every Day Big Data Statistics – 2.5 Quintillion Bytes of Data Created Daily < https://www.dihuni.com/2020/04/10/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/>

15. Government of Sierra Leone Ministry of Health (2016), Annual Health Performance Report, < https://www.afro.who.int/publications/sierra-leone-health-sector-performance-report-2016>

16. Human Data Xchange (HDX), (2015), Sub-national time series data on Ebola cases and deaths in Guinea, Liberia, Sierra Leone, Nigeria, Senegal and Mali since March 2014, <https://data.humdata.org/dataset/rowca-ebola-cases>

17. International Trade Center (2014), Guinea Infrastructure, <http://www.intracen.org/country/guinea/Infrastructure/>

18. Liberia Institute of Statistics & Geo Information Services (LISGIS) (2016), Household Income and Expenditure Survey of 2014,< https://www.lisgis.net/pg_img/Liberia%20Statistical%20Abstract%20FINAL.pdf>

19. 12.Médecins Sans Frontiers, Ebola 2014-2015 Facts & Figures, <https://www.msf.org/sites/msf.org/files/ebola_accountability_report_low_res.pdf>

20. National Institute of Statistics, Guinea, Ministry of Planning and Economic Development of Guinea, Annual report (2016), <http://www.stat-guinee.org/index.php/publications-ins/89-publications-annuelles>

21. Statistics Sierra Leone (Stats SL) (2015), Population and Housing Census, < https://www.statistics.sl/images/StatisticsSL/Documents/final-results_-2015_population_and_housing_census.pdf>

22. World Economic Forum (2019), Africa's approaches to crisis management are unique when it comes to dealing with disease <https://www.weforum.org/agenda/2019/08/how-to-manage-in-a-crisis-lessons-from-the-west-africa-ebola-outbreak/>

23. World Health Organization (WHO), (2020), Ebola Virus Disease, < https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease>

24. World Health Organization (WHO), (2014), Ebola Response Roadmap Situation Report, <https://apps.who.int/iris/bitstream/handle/10665/146763/roadmapsitrep_31Dec14_eng.pdf?sequence=1>

25. World Health Organization (WHO), (2015), Factors that contributed to undetected spread of the Ebola virus and impeded rapid containment, <https://www.who.int/csr/disease/ebola/one-year-report/factors/en/>

26. World Health Organization (WHO), (2014), Global Atlas of Medical Devices, Liberia < https://www.who.int/medical_devices/countries/lbr.pdf?ua=1>

27. WEKA, data analysis tool, <https://www.cs.waikato.ac.nz/~ml/weka/>

28. World Bank (2014), World Development Report: Risks and Opportunities < http://documents1.worldbank.org/curated/en/914141468325443509/text/WPS6847.txt>

29. World Health Organization: Global Health Observatory data repository: Health Workforce <https://apps.who.int/gho/data/node.main.SDG3C?lang=en>

30. United Nations: Diagnosing Africa's medical brain drain <https://www.un.org/africarenewal/magazine/december-2016-march-2017/diagnosing-africa%E2%80%99s-medical-brain-drain>

DECISION SCIENCES INSTITUTE
Customer Education for a Multi-Channel Service Provider

Sara Babaee
Wilfrid Laurier University
Email: sbabaee@wlu.ca

Mojtaba Araghi
Wilfrid Laurier University
Email: maraghi@wlu.ca

Ignacio Castillo
Wilfrid Laurier University
Email: icastillo@wlu.ca

ABSTRACT

This research is aimed at providing a formal methodology to assess investment in increasing customer education in service for a service environment requiring a high level of customer participation for success. The considered service provider offers the service in two channels: traditional, face to face contact with employees, and online, access to the firm mobile application or website. Our model links investment on customer education in the online channel to likely future shopping and purchase patterns and hence to revenue flow. We analyzed that under what conditions, it is not profitable to increase customer participation using educational tools.

KEYWORDS: marketing-operations interface, customer churn management, multi-channel service, customer education, complex and prolonged service

INTRODUCTION

Providing service through multiple channels is enabled by various service interfaces, such as physical stores, websites, and mobile applications (apps for short). These multiple channels can have both competitive and complementary effects. Different channels of a firm can be competitive as any enhancement in one channel can shift the demand from the other channels or complementary, as enhancing one channel can increase the overall perceived quality of the firm. Thus, understanding the trade-offs that service providers face when deciding to enhance a channel is essential.

During the last decade, providing service through online platforms has become one of the most crucial channels for multi-channel service providers as they are very successful in attracting new customers. However, these platforms are usually not reliable in converting them to repeat customers after the first use. A study on the adoption of digital coaching platform among teenagers showed that while teenagers often download related apps, they stop using them after a short period (Kettunen & Kari, 2018). Another survey conducted recently in the United States showed that while 87 percent of respondents of age 18-to-29 years old are open to using an app that gives fitness instructions to them, only 12 percent of them will end up using the app regularly (“Usage of health apps to receive fitness instructions among US adults by age 2017 | Survey,” 2017). Therefore, in addition to the cross-channel trade-offs, service providers should consider various churn management strategies to improve the loyalty of online channel users.

While providing multi-channel service is the new standard in many industries, in this paper, we focus on a particular type of service, commonly named as *complex and prolonged* services (Spanjol et al., 2015; Temerak, Winklhofer, & Hibbert, 2018). The distinctive feature of complex and prolonged services is that the value is co-created by the service provider and customers: the firm provides guidelines and structures, but customers are the leading players of the service process as they need to pursue the instructions and perform a series of exercises and/or activities to achieve a goal. All forms of education, fitness training, health care activities, and weight loss programs are examples of complex and prolonged services (Jo Bitner, Faranda, Hubbert, & Zeithaml, 1997).

The online channel has been extensively used in complex and prolonged service environments, as evidenced by the proliferation of health and fitness mobile apps available in the Google play store and Apple App store ("mHealth apps, App Store 2018 | Statistic," n.d.). In these co-creation service environments, customers' level of expertise is not only essential in improving the progress toward desired service outcomes, but also in affecting the relationship between customers and the firm (Temerak et al., 2018). Thus, educating customers to make them proficient in the field can be considered as a customer retention strategy.

As a real-world example, consider a weight loss program that offers the service in two channels. In the traditional channel, customers are in contact with a personal expert coach who will create their appropriate diet plan considering their needs and preferences. A similar service is provided through the app, where technology offers a diet plan based on the input from the customer. Clearly, these two channels have different attractiveness for customers and generate different profits per customer.

In addition to the weekly diet plans, in both channels, customers receive some nutrition education, which helps them in tailoring the diet plan to their needs and preferences. This education can improve a customer's knowledge up to a level that affects her relationship with the firm, specifically her choice of service channel or even her churn rate as she feels confident to plan her own diet.

While the firm decides on the level of educational effort in both channels, it is usually more difficult to modify the effort for the traditional channel due to more established industry standards, as well as platform limitations. Therefore, in this paper, we focus on the trade-offs that a complex and prolonged service provider faces when trying to enhance its digital channel with educational features.

We propose a mathematical model that analyses the customer flow through the system. We model the system as a two-channel network (traditional and digital) with base customers with two levels of knowledge (novice and expert) and state-dependent routing. The model captures novice customers' conversion to expert customers, and by linking the customer education to the firm's revenue flow, we characterize the conditions under which it is profitable to enhance the digital channel with educational features.

The rest of this paper is organized as follows. First, we briefly review the related literature. We then develop our mathematical programming model. We analyze the profit function and derive the optimal education strategy in the numerical example section. Finally, in the last section, we present our concluding remarks.

LITERATURE REVIEW

Our work is primarily related to three streams of research, multi-channel service optimization, customer churn management, and customer education. We will discuss each of these streams in this section.

Multi-channel Service Optimization

Previous research discussed multi-channel service management from the perspective of customer's channel selection (Chen, Kaya, & Özer, 2008; J. Li, Konuş, Langerak, & Weggeman, 2017), multi-channel strategy implementation (Montaguti, Neslin, & Valentini, 2016) , and resource allocation between the channels (Xiao, Chen, & Chen, 2009). A wealth of related research has studied how the combination of traditional and new channels influence the service provider's performance through cross channel effects, complementary effects (Homburg, Vollmayr, & Hahn, 2014; Wallace, Giese, & Johnson, 2004) and competitive effects (Bernstein, Song, & Zheng, 2009; Huang, Lu, & Ba, 2016; Wang & Goldfarb, 2017). Huang et al. present an empirical study on the cross-channel effects between the web and mobile shopping channels for an e-commerce company. The results verified the existence of both complementary and competitive effects. However, the positive complementary effect of introducing the new mobile channel outperformed its negative competitive effect (Huang et al., 2016). Wang and Goldfarb discuss that for a brick-and-click retailer, offline and online channels are competitive when viewed as retail channels and complementary when viewed as marketing channels (Wang & Goldfarb, 2017). While cross channel effects are mainly studied empirically, our paper analytically discusses how improving the quality of one channel affects the firm's other service channels.

Moreover, as it is shown in a comprehensive review of studies in the multi-channel retail industry (H. Li, Lobschat, & Verhoef, 2018), the research has been mainly focused on retail services (Oh, Teo, & Sambamurthy, 2012) and other service environments have rarely been studied. One exception is the work of Akçura and Ozdemir, who investigate the multi-channel optimization problem for expert services. Their research discusses whether the expert service providers, such as lawyers, designers, and financial advisers, should charge separate prices for the in-person and the online service, or they should provide the online service as a supplement to their traditional in-person one (Akçura & Ozdemir, 2017). Our research is focused on the complex and prolonged services, which is fundamentally different from online sales, as it includes the customer involvement in the service process and its impact on the service outcome (Bellos & Kavadias, 2019; Roels, 2014).

Customer Churn Management

Customer churn or defection is a crucial concern for firms in various industries and can dramatically affect a firm's financial performance. The majority of papers on the customer churn management has either focused on single-channel services (Ascarza & Hardie, 2013) (Ng & Chung, 2012) or studied the impact of different factors in multi-channel services on customer retention empirically (Boehm, 2008). Chun and Zhang analytically model the dynamics of multi-channel customer relationships in a non-contractual retail setting (Chang & Zhang, 2016). Our model extends the customer retention activities, in the form of customer education, to multi-channel complex prolonged service providers.

Customer Education

Competing hypotheses exist on whether educating customers have a positive or negative effect on customer retention (Simon J. Bell, Auh, & Eisingerich, 2017). While some studies show a positive relationship between customers expertise and their loyalty to the firm (S. J. Bell, Auh, & Smalley, 2005; Suh, Greene, Israilov, & Rho, 2015), some other studies have found no significant relationship (Simon J. Bell & Eisingerich, 2007) or even negative relationship (Wirtz & Mattila, 2003) between customer's level of expertise and their loyalty.

Complex and prolonged service providers are generally reluctant to invest on customer education as they are afraid it provides customers with enough knowledge to produce the service by themselves, blame the firm when things go wrong, and switch to the competitors (Eisingerich & Bell, 2008; Hilton, Hughes, Little, & Marandi, 2013). However, some studies show that customer education can be a useful tool to build a stable and trusting relationship with customers (Burton, 2002; Retana, Forman, & Wu, 2016). Building upon these studies, we link the investment in customer education to the customers' retention and the corresponding firm's profit.

To summarize, the central contribution of our work is integrating the above three streams of literature by mathematically modelling the impacts of implementing a customer education strategy on customers' flow and cross channel effects in a complex and prolonged multi-channel service provider.

MODEL AND PROBLEM FORMULATION

We consider a firm that offers a complex and prolonged service through two separate channels: (i) traditional channel where in-person service, education and support is provided for customers, and (ii) digital channel where customers have access to the firm's mobile app (or website) which offers the service, alongside online learning process. Due to different provided services, channels have distinct attractiveness for customers and profitability.

New customers arrive at the system through each channel following a stationary Poisson process with rate λ_i for channel $i \in \{t: traditional, d: digital\}$. While receiving the service, customers are educated to increase their knowledge in the field. We consider two levels of customers' expertise: expert customers who are properly applying all the service guidelines and effectively deriving the value by adapting the service to their preference, and novice customers who only follow the instructions and cannot improvise or design a plan. We assume all new customers are novice at their arrival but can grow to the expert level. After each period of interaction with the firm, a novice customer receiving service and education in the traditional channel becomes expert with probability $\boldsymbol{\varepsilon_t}$, and a novice customer who receives some education while receiving service in the digital channel becomes expert with probability $\boldsymbol{\varepsilon_d}$. All expert customers remain expert until they eventually leave the firm.

One obvious caveat of considering only two levels of expertise is that we cannot fully capture the learning process, and we observe a memoryless process where the education history of a novice customer is neglected. This "recency effect" is commonly assumed in models that link customers' behaviour to their past interaction with the firm (e.g., (Afèche, Araghi, & Baron, 2017; Ho, Park, & Zhou, 2006)). Capturing the education history of each customer is feasible in our model by considering the intermediary steps of customers' knowledge and flows of customers among them, but it is outside the scope of this paper.

We consider a contractual setting where at the start of each period (week or month, for instance) base customers either renew their subscription to the traditional or digital channel or leave the firm. We assume that the probability of each of these observable decisions depends on the characteristics of the channel as well as the customer's current level of expertise. At the start of each renewal period, a novice (expert) customer will leave the firm with the probability of $\mu_n(\mu_e)$, sign a contract for the traditional channel with the probability of r_{tn} (r_{te}) or sign a contract for the digital channel with the probability of r_{dn} (r_{de}). The firm's profit of serving a customer is independent of the customer's knowledge level. We let w_t and w_d represent the profit per contract period for the traditional and digital channel, respectively.

Figure 1 shows a schematic of customer flow through the system. A summary of the notation used in our model is presented in Table 1.

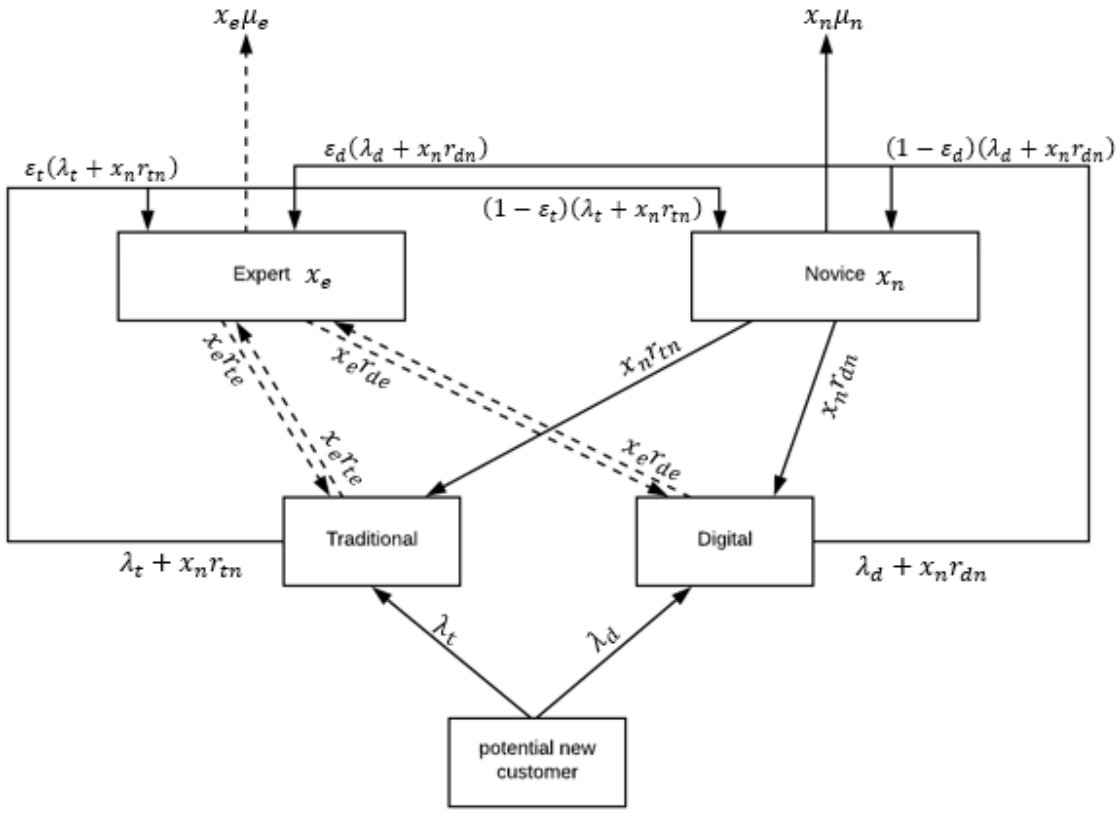


Figure 1: The flow of New and Novice (solid black lines), and Expert (dashed lines) customers through the system

Table 1: Summary of Notation

x_n, x_e	Size of novice or expert customer base
μ_n, μ_e	The probability that a novice or expert customer chooses to terminate his/her relationship with the firm
r_{in}, r_{ie}	The probability that a novice or expert customer chooses to sign a contract for using channel $i \in \{t: traditional, d: digital\}$
λ_i	The arrival rate of new customers to channel $i \in \{t: traditional, d: digital\}$
ε_i	The probability that a novice customer becomes expert after receiving service in channel $i \in \{t: traditional, d: digital\}$
w_i	Profit per contract of channel $i \in \{t: traditional, d: digital\}$

The rate of becoming expert after receiving service depends on the quality and level of education the firm has considered in the traditional or digital channels. As we discussed in the introduction, the traditional channel usually follows widespread industry standards. Therefore, in our model, we treat ε_t as given and focus on optimizing ε_d for a firm that is considering starting or improving its digital channel. We assume the firm needs to spend $F(\varepsilon_d)$ each period to provide an online education environment that elevates ε_d percent of novice customers to experts.

We assume that $F(\varepsilon_d)$ is a non-decreasing convex function of ε_d. Note that $F(\varepsilon_d)$ considers the cost of designing the education strategy, not the operational cost of applying it. Therefore, $F(\varepsilon_d)$ is independent of the number of users, and all operational costs are considered in the profit per user w_t and w_d.

To investigate the long-time effects of educating customers, we consider the system in its steady state. The number of novice customers in the steady-state is:

$$x_n = \frac{\lambda_d(1-\varepsilon_d) + \lambda_t(1-\varepsilon_t)}{\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t} \qquad (1)$$

where the numerator is the rate of novice customers that join the base after receiving service as new customers through digital and traditional channels, and the denominator captures the departure rate from the novice customer base either by leaving the system or growing to expert customers.
Similarly, the size of the expert customer base in the steady-state is:

$$x_e = \frac{\lambda_d\varepsilon_d + \lambda_t\varepsilon_t}{\mu_e} + x_n \times \frac{r_{dn}\varepsilon_d + r_{tn}\varepsilon_d}{\mu_e} \qquad (2)$$

which depends both on the arrival rate of new customers to the system and the size of the novice customer base, as the novice customers will grow to expert customers with a probability of ε_d and ε_t after a period of interaction with the digital and traditional channels, respectively. Note that the departure from the expert customer base happens with rate μ_e as expert customers remain expert while in the system.
Let π be the firm's profit in the steady-state:

$$\pi = w_t(\lambda_t + x_e r_{te} + x_n r_{tn}) + w_d(\lambda_d + x_e r_{de} + x_n r_{dn}) - F(\varepsilon_d) \qquad (3)$$

The first product in (3) shows the total profit of the traditional channel where $(\lambda_t + x_e r_{te} + x_n r_{tn})$ is the average number of people using the traditional channel at each period in the steady state. The second product captures the profit of the digital channel and $F(\varepsilon_d)$ is the cost of reaching the desired education level in the digital channel.

OPTIMAL EDUCATION POLICY

In this section, we solve the profit maximization problem (3) by choosing the optimal value for $\boldsymbol{\varepsilon_d}$, given the dynamics of the system considered in (1) and (2). Before solving this problem, we discuss how the number of users of traditional and digital channels will be affected by $\boldsymbol{\varepsilon_d}$.

Proposition 1: For a fixed new customer arrival rate (λ_t, λ_d) by increasing ε_d, the number of users in traditional and digital channels changes as follows.

1. If $\frac{\mu_e}{\mu_n} \leq \min(\frac{r_{te}}{r_{tn}}, \frac{r_{de}}{r_{dn}})$, then increasing ε_d will increase the number of users in both traditional and digital channels.
2. If $\frac{\mu_e}{\mu_n} \geq \max(\frac{r_{te}}{r_{tn}}, \frac{r_{de}}{r_{dn}})$, then increasing ε_d will decrease the number of users in both traditional and digital channel.
3. If $\min(\frac{r_{te}}{r_{tn}}, \frac{r_{de}}{r_{dn}}) \leq \frac{\mu_e}{\mu_n} \leq \max(\frac{r_{te}}{r_{tn}}, \frac{r_{de}}{r_{dn}})$, then
    a. If $\frac{r_{te}}{r_{de}} \geq \frac{r_{tn}}{r_{dn}}$, then increasing ε_d increases the number of users in the traditional channel while the number of users in the digital channel decrease.

b. Otherwise, increasing ε_d increases the number of users in the digital channel and decreases the number of users in the traditional channel.

Proposition 1 highlights the important consequences of educating online customers in different situations for a multi-channel service provider. First, if the churn rate of expert customers is sufficiently low in comparison with the churn rate of novice customers, by investing in educating customers in the digital channel, more novice customers will grow to the expert customer base and stay longer with the firm. Therefore, the total number of base customers in steady-state will increase, and we observe an increase in the number of users in both channels. Conversely, if expert customers have a relatively high churn rate, investing in educating online customers decreases the total number of base customers and thus reduces the number of users in both channels over time. In the middle of these two extremes, the change in the number of users of digital or traditional channels depends on the relative retention rate of different channels for different customers. If the attractiveness of traditional to digital channel is higher for expert customers (i.e., $\frac{r_{te}}{r_{de}} \geq \frac{r_{tn}}{r_{dn}}$), a higher conversion rate from novice to expert customers increases the number of traditional channel users and decreases the number of digital channel users. The vice versa happens when expert customers prefer the digital channel more than novice customers.
Proposition 1 can be used to align the firm strategic plan regarding customer education to its operational plans. It provides a general understating of how capacity requirements can be affected by enhancing the education in the digital channel, and if the workforce should be increased, decreased or shifted between the channels.

Analysis of The Profit Function

In this section, we study the behavior of the profit function in response to improving the quality of customer education, ε_d. We first derive a customer value metric to compare the value of novice and expert customers in the absence of education, and then determine an optimal education policy that hinges on this metric.
The value of a customer can be defined as her expected lifetime profit for the firm. During each period, an expert (novice) customer generates a profit of w_t by choosing the traditional channel with probability r_{te} (r_{tn}), or profit of w_d by choosing the digital channel with probability r_{de} (r_{dn}). In the absence of education, novice and expert customers do not convert to each other and only leave the corresponding customer base by terminating their relationship with the firm. Therefore, the expected number of periods that an expert (novice) customer stays with the firm (i.e., her sojourn time) in the absence of education is $\frac{1}{\mu_e}$ ($\frac{1}{\mu_n}$). Let φ represent the difference between the lifetime value of an expert and a novice customer in the absence of education:

$$\varphi = \frac{w_t r_{te} + w_d r_{de}}{\mu_e} - \frac{w_t r_{tn} + w_d r_{dn}}{\mu_n} \quad (4)$$

A negative value for φ means that a novice customer (even if she remains novice for the rest of her relationship with the firm) is more valuable than an expert one. This may happen when the expert level of knowledge makes customers feel self-sufficient and thus increasing their churn rate, μ_e. A lower value of expert customers may also happen when the higher level of knowledge encourages expert customers to choose a less expensive channel and thus bringing down the average profit of the firm. Under these scenarios, the profit of the education is not aligned for the firm and the customers. Therefore, while the firm provides customers in the

digital channel with the service they have registered for, it does not have enough incentives to educate them in the field. Note that here we assume the arrival rate of new customers is exogenous and independent of the quality of education provided by the firm. In the next section, we extend the model such that word-of-mouth (WOM) regarding the quality of education affects the new customer arrival rate. Considering the WOM effect, the boost in the arrival rate of new customers may justify investing in novice customers' education, even if expert customers are less valuable for the firm.

Educating novice customers is profitable when the lifetime value of an expert customer is higher than the one of a novice customer who remains novice, i.e., when $\varphi \geq 0$. To find the optimal level of education, however, we need to compare the marginal cost of education with its marginal profit.

As we discussed in the model section, to educate and convert ε_d percent of novice customers to expert ones after using the digital channel, the firm needs to spend $F(\varepsilon_d)$ in education where $F(\varepsilon_d)$ is a non-decreasing convex function of ε_d. Let $F'(\varepsilon_d)$ and $R'(\varepsilon_d)$ define the marginal cost and revenue (profit of serving customers excluding the education cost) of education in the digital channel, respectively, where $F'(\varepsilon_d)$ depends on the firm's specific education cost function, and $R'(\varepsilon_d)$ follows from (1), (2), and (3):

$$R'(\varepsilon_d) = \frac{\mu_n\left(\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)\right)\varphi}{(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^2} \quad (5)$$

It can be seen from (5) that the investment in digital education has a diminishing return as the marginal revenue of education decreases in ε_d. It is intuitive that the marginal value of educating novice customers increases in the difference in the lifetime value of expert and novice customers, φ. $R'(\varepsilon_d)$ also increases in the average sojourn time of novice customers, i.e., the reciprocal of the rate at which novice customers leave the novice base $(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^2$. Proposition 2 illustrates how the optimal educational strategy depends on the difference between the value of novice and expert customers φ, marginal revenue $R'(\varepsilon_d)$, and marginal cost $F'(\varepsilon_d)$.

Proposition 2: to maximize the profit function (3), the firm should choose the following education level in the digital channel:

1. If $\varphi \geq 0$, the profit function is concave in ε_d and the optimal education level is ${\varepsilon_d}^* = sup\left\{0 \leq \varepsilon \leq 1: R'(\varepsilon) > F'(\varepsilon)\right\}$ if such ε exists and ${\varepsilon_d}^* = 0$ otherwise.
2. Otherwise, the profit is decreasing in the education of novice customers in the digital channel; thus, the firm should not invest in it, ${\varepsilon_d}^* = 0$.

If $\varphi \leq 0$, novice customers are at least as valuable as expert ones, and the firm's revenue does not increase by educating them $R'(\varepsilon_d) \leq 0$. Therefore, as long as the cost function $F(\varepsilon_d)$ is non-decreasing, the profit decreases in ε_d, and the optimal solution is ${\varepsilon_d}^* = 0$.

When the expert customers are more valuable $\varphi \geq 0$, however, the revenue increases by education $R'(\varepsilon_d) \geq 0$. If the marginal revenue of educating novice customers is less than its corresponding cost at $\varepsilon_d = 0$, it would not be profitable even to start educating customers and $\varepsilon_d = 0$. Otherwise, the firm should improve education up to a level where the marginal cost of educating customers equals the marginal increase in revenue. Note that if the marginal revenue remains larger than the marginal cost of education at the extreme level of $\varepsilon_d = 1$, it is optimal to invest in novice customer education so much that all of them become expert just after using the

digital channel once. However, it does not happen in reality, as the marginal cost of education would be extremely high at that level of education.

Optimal Education in the Presence of Word-of-Mouth

The quality of educating customers may generate a positive WOM effect that increases the firm's ability to attract new customers, especially through the digital channel. In this section, we extend our model to study the changes in the structure of the optimal education level in the presence of WOM.
We consider λ_d as the minimum arrival rate of new customers to the digital channel, and model the arrival rate at the presence of positive WOM as $\lambda_d + k\varepsilon_d$, where parameter k captures the WOM intensity. The corresponding size of the novice and expert customer bases, as well as the profit function, can simply be modelled by substituting λ_d with $\lambda_d + k\varepsilon_d$ in equations 1, 2, and 3. However, the effect of WOM on the marginal revenue of the firm, R'_{WOM} is more complicated, as it is provided in the proof of Proposition 3.

Proposition 3: In the presence of WOM, to maximize the profit function (3), the firm should choose the following marketing level in the digital channel:

1. If $\varphi < 0$ and $k \geq \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$ or $\varphi \geq 0$ and $k < \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$, the profit function is concave in ε_d and the optimal education level in the digital channel is ${\varepsilon_d}^* = sup\left\{0 \leq \varepsilon \leq 1: R'_{WOM}(\varepsilon) > F'(\varepsilon)\right\}$ if such ε exists and ${\varepsilon_d}^* = 0$ otherwise.
2. Otherwise, the revenue of the firm becomes convex in ε_d, thus, the profit function may increase or decrease more than once depending on the shape of the cost function.

Note that, in the absence of WOM when $\varphi < 0$, the optimal solution is not to educate novice customers in the digital channel. However, when the intensity of WOM, k, is sufficiently large, it may be optimal to grow more valuable novice customers to less valuable expert customers to attract more new customers into the digital channel in return. At lower levels of k, the behaviour of the profit function remains as before for $\varphi \geq 0$. When none of these conditions hold, the revenue of the firm becomes convex in ε_d as the WOM effect is not significant enough to make up for the lost revenue in case of negative φ, or it dominates the diminishing return on education in case of positive φ. As a result, we cannot establish the structure of the optimal solution without knowing the cost function.

Numerical Example:
To reinforce our findings in propositions 2 and 3, we use a numerical example to illustrate how the optimal education level in the digital channel changes with the education cost and WOM. We consider a convex education cost $F(\varepsilon_d) = \frac{a}{1-\varepsilon_d}$. The education cost has a lower bound a and grows to infinity as ε_d approaches 1. This cost function can be justified by noting that we are focusing on the cost of designing an education system. Unlike an operational cost that usually grows linearly in the amount of usage, improving the quality becomes more difficult as it gets better. Particularly, we can assume that it is virtually impossible to design an online system that converts all novice customers to an expert after just a single period of interaction.

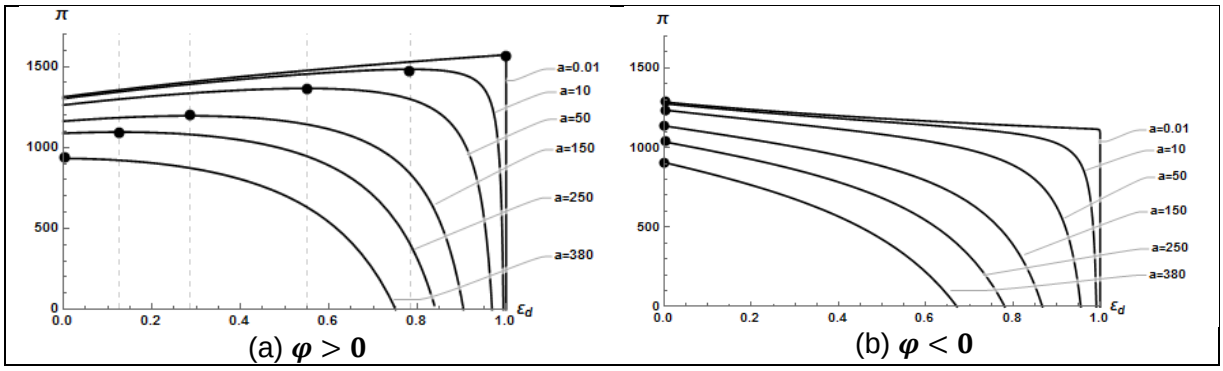


Figure 2: Profit function in the absence of WOM

Figure 2 shows the firm's profit as a function of the education level at the digital channel $\boldsymbol{\varepsilon_d}$, for different values of the education cost's lower bound a. We set $\frac{r_{tn}}{r_{dn}} = \frac{1}{2}$, $\frac{r_{te}}{r_{de}} = \frac{1}{3}$, $\varepsilon_t = 0.7$, $w_t = 50$, $w_d = 10$, λ_t=10, and λ_d=20. In Figure 2(a), we assume Novice customers leave the system with a higher probability ($\mu_n = 0.6$, $\mu_e = 0.4$), and educating customers can be profitable for the firm $\varphi \geq 0$. It follows from Proposition 2 that the profit function is concave, and the optimal education level is decreasing in the education cost. The optimal education level is almost 1 when the education cost is negligible, $a = 0.01$, and it decreases as a increases from 0.0 to 10, 50, 150, 250, and 380. Note that for any value of $a > R'(0) = 360.57$, the marginal cost remains larger than the marginal profit for any education level, thus by part 1 of Proposition 2, ${\varepsilon_d}^* = 0$. In contrast, in Figure 2(b) we consider a system with $\varphi < 0$ by setting $\mu_n = 0.5$, $\mu_e = 0.6$. It follows from the second part of Proposition 2 that ${\varepsilon_d}^* = 0$ for all values of a.

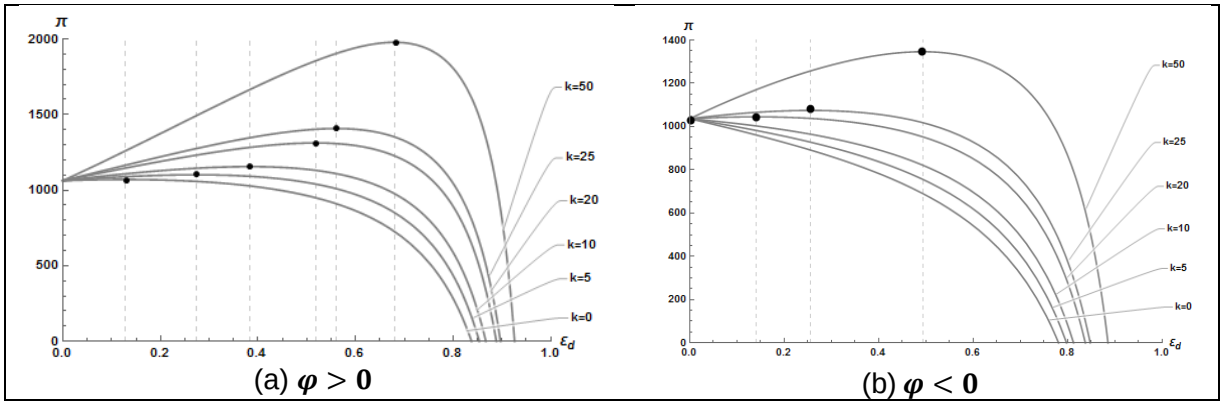


Figure 3: Profit function in the presence of WOM

In Figure 3, we study how positive WOM affects the optimal education level in the digital channel. We use the same setup as in Figure 2, fix a at 250 where without WOM, the optimal education level was ${\varepsilon_d}^* = 0.13$ when $\varphi \geq 0$, and ${\varepsilon_d}^* = 0$ when $\varphi < 0$. The threshold of WOM intensity defined in Proposition 3 is 8 for positive φ, and 11 for negative φ. To study how the profit function behaves before and after the threshold, in Figure 3, we set k to 0, 5, 10, 20, 25, and 50. It is seen in Figure 3(a) that when $\varphi \geq 0$, increasing the intensity of WOM from 0 to 50, increases the optimal education level exponentially. This happens as not only investing in elevating customers to expert ones is profitable for the firm; it also attracts more new customers to the firm. In contrast, in Figure 3(b), where the expert customers are not as profitable as

novice ones, improving the education quality is not optimal when the WOM intensity is relatively small ($k \in \{0,5,10\}$). However, as k increases, the profit of absorbing new novice customers dominates the loss of educating them to novice ones, and $\boldsymbol{\varepsilon_d}^*$ grows to around 0.5.

CONCLUDING REMARKS

we study the profit-maximizing education strategy for a multi-channel complex and prolonged service provider with repeat customers whose behaviour depends on their level of knowledge in the corresponding field. By educating customers, the firm affects customers' choice of channel and retention and consequently, their flow in the system and the generated profit.

We derive a metric that compares the lifetime value of novice and expert customers based on their level of expertise and further prescribes optimal customer education policy that hinges on this metric. We show that when positive word-of-mouth regarding the quality of education in the digital channel is considered, the firm may choose to educate customers even if novice customers are more profitable for the firm. This research can help the firms to better understand how marketing-focused policies such as educating customers affect operational aspects of the system and study their long term effect on the firm's profit.

Our study focuses on two levels of customer's expertise. By extending the analysis to include multiple levels of knowledge, we can model the full history of education that each customer receives and its effects on her behaviour.

In this research, we study customer's behaviour as a function of their knowledge, since it is the most relevant factor in the value co-creation environments. However, our proposed model can be implemented for other multi-channel services with different underlying factors, such as satisfaction and commitment, which can be directly or indirectly managed over time by the firm's investment.

APPENDIX

Proposition 1:
Proof: let x_t and x_d show the number of users in a steady state for the traditional and digital channel, respectively.

$$x_t = \lambda_t + x_e r_{te} + x_n r_{tn} \quad (6)$$

$$x_d = \lambda_d + x_e r_{de} + x_n r_{dn} \quad (7)$$

Differentiating x_t and x_d with respect to ε_d yields

$$\frac{\partial x_T}{\partial \varepsilon_d} = \frac{-[\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)] \times (\mu_e r_{tn} - \mu_n r_{te})}{\mu_e(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^2} \quad (8)$$

$$\frac{\partial x_D}{\partial \varepsilon_d} = \frac{-[\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)] \times (\mu_e r_{dn} - \mu_n r_{de})}{\mu_e(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^2} \quad (9)$$

It follows from (8) that the number of users in the traditional channel will increase by ε_d if $(\mu_e r_{tn} - \mu_n r_{te}) < 0$ and decrease if $(\mu_e r_{tn} - \mu_n r_{te}) \geq 0$.
Similarly, the number of users in the digital channel will increase if $(\mu_e r_{dn} - \mu_n r_{de}) < 0$ and decrease if $(\mu_e r_{dn} - \mu_n r_{de}) \geq 0$. ■

Proposition 2:
Proof: Based on (3) the first and second-order derivative of φ with respect to ε_d will be as follows.

$$\frac{\partial \pi}{\partial \varepsilon_d} = \frac{\mu_n\big(\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)\big)\varphi}{(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^2} - F'(\varepsilon_d) \quad (10)$$

$$\frac{\partial^2 \pi}{(\partial \varepsilon_d)^2} = \frac{-2\mu_n r_{dn}\big(\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)\big)\varphi}{(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t)^3} - F''(\varepsilon_d) \quad (11)$$

The first part of proposition 2, follows from the concavity of profit function in ε_d. Based on (11), when $\varphi \geq 0$, the second-order derivative of profit is negative, assuming that the cost function is non-decreasing convex. Therefore, the profit function is concave in ε_d. when $\varphi < 0$, the first-order derivative of profit function with respect to ε_d is negative, assuming that the cost function is non-decreasing, i.e, $F'(\varepsilon_d) \geq 0$. Therefore, the profit is strictly decreasing in ε_d and $\boldsymbol{\varepsilon_d}^* = \mathbf{0}$. ■

Proposition 3:
Proof: Substituting λ_d with $\lambda_d + k\varepsilon_d$ in equation 1, 2, and 3 we get

$$\frac{\partial \pi^{wom}}{\partial \varepsilon_d} = \frac{\mu_n\left(\lambda_d\left(\mu_n + r_{dn} + r_{tn}\varepsilon_t\right) + \lambda_t r_{dn}(1-\varepsilon_t)\right)\varphi}{\left(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t\right)^2} + \frac{k}{\mu_e\left(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t\right)^2} \times \left\{\left(\mu_n + r_{tn}\varepsilon_t\right)\left[\mu_e\left(w_t r_{tn} + w_d(1-r_{tn})\right) + r_{tn}\varepsilon_t\left(w_t r_{te} + w_d(1-r_{te})\right)\right] + \left[{\varepsilon_d}^2 r_{dn} + 2\varepsilon_d\left(\mu_n + r_{tn}\varepsilon_t\right)\right] \times \left[\mu_n(w_t r_{te} + w_d r_{de}) - \mu_e w_t r_{tn} + r_{dn}(w_t r_{te} + w_d r_{de})\right]\right\} - F'(\varepsilon_d) \quad (15)$$

$$\frac{\partial^2 \pi^{wom}}{(\partial \varepsilon_d)^2} = \frac{2\mu_n\left[k(\mu_n + r_{tn}\varepsilon_t)(\mu_n + r_{dn} + r_{tn}\varepsilon_t) - r_{dn}\left(\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)\right)\right] \times \varphi}{\left(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t\right)^3} - F''(\varepsilon_d) \quad (16)$$

Similar to the proof of proposition 2, the first part of proposition 3 follows from the concavity of the profit function. According to 16 and the convexity of cost function, if $\varphi < 0$ and $k \geq \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$ or $\varphi \geq 0$ and $k < \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$, the second-order derivative of profit function is negative. Therefore, the profit is concave in ε_d and the first part of proposition 3 follows.

If $\varphi < 0$ and $k < \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$ or $\varphi \geq 0$ and $k \geq \frac{r_{dn}}{\mu_n + r_{tn}\varepsilon_t}\left(\lambda_d + \lambda_t \frac{r_{dn}(1-\varepsilon_t)}{\mu_n + r_{dn} + r_{tn}\varepsilon_t}\right)$, we observe that

$$R''_{wom} = \frac{2\mu_n\left[k(\mu_n + r_{tn}\varepsilon_t)(\mu_n + r_{dn} + r_{tn}\varepsilon_t) - r_{dn}\left(\lambda_d(\mu_n + r_{dn} + r_{tn}\varepsilon_t) + \lambda_t r_{dn}(1-\varepsilon_t)\right)\right] \times \varphi}{\left(\mu_n + r_{dn}\varepsilon_d + r_{tn}\varepsilon_t\right)^3} > 0$$

Therefore, $\frac{\partial^2 \pi^{wom}}{(\partial \varepsilon_d)^2}$ is not always positive or negative, and there might be multiple roots in the first-order conditions of 14 depending on the shape of the cost function. ■

REFERENCES

Afèche, P., Araghi, M., & Baron, O. (2017). Customer Acquisition, Retention, and Service Access Quality: Optimal Advertising, Capacity Level, and Capacity Allocation. *Manufacturing & Service Operations Management*, *19*(4), 674–691. https://doi.org/10.1287/msom.2017.0635

Akçura, M. T., & Ozdemir, Z. D. (2017). A Strategic Analysis of Multi-Channel Expert Services. *Journal of Management Information Systems*, *34*(1), 206–231. https://doi.org/10.1080/07421222.2017.1297637

Ascarza, E., & Hardie, B. G. S. (2013). A Joint Model of Usage and Churn in Contractual Settings. *Marketing Science*, *32*(4), 570–590. https://doi.org/10.1287/mksc.2013.0786

Bell, S. J., Auh, S., & Smalley, K. (2005). Customer Relationship Dynamics: Service Quality and Customer Loyalty in the Context of Varying Levels of Customer Expertise and Switching Costs. *Journal of the Academy of Marketing Science*, *33*(2), 169–183. https://doi.org/10.1177/0092070304269111

Bell, Simon J., Auh, S., & Eisingerich, A. B. (2017). Unraveling the Customer Education Paradox. *Journal of Service Research*, *20*(3), 306–321. https://doi.org/10.1177/1094670517691847

Bell, Simon J., & Eisingerich, A. B. (2007). The paradox of customer education. *European Journal of Marketing*, *41*(5/6), 466–486. https://doi.org/10.1108/03090560710737561

Bellos, I., & Kavadias, S. (2019). When Should Customers Control Service Delivery? Implications for Service Design. *Production and Operations Management*, *28*(4), 890–907. https://doi.org/10.1111/poms.12956

Bernstein, F., Song, J.-S., & Zheng, X. (2009). Free riding in a multi-channel supply chain. *Naval Research Logistics*, *56*(8), 745–765. https://doi.org/10.1002/nav.20379

Boehm, M. (2008). Determining the impact of internet channel use on a customer's lifetime. *Journal of Interactive Marketing*, *22*(3), 2–22. https://doi.org/10.1002/DIR.20114

Burton, D. (2002). Consumer education and service quality: conceptual issues and practical implications. *Journal of Services Marketing*, *16*(2), 125–142. https://doi.org/10.1108/08876040210422673

Chang, C.-W., & Zhang, J. Z. (2016). The Effects of Channel Experiences and Direct Marketing on Customer Retention in Multichannel Settings. *Journal of Interactive Marketing*, *36*, 77–90. https://doi.org/10.1016/j.intmar.2016.05.002

Chen, K.-Y., Kaya, M., & Özer, Ö. (2008). Dual Sales Channel Management with Service Competition. *Manufacturing & Service Operations Management*, *10*(4), 654–675. https://doi.org/10.1287/msom.1070.0177

Eisingerich, A. B., & Bell, S. J. (2008). Perceived Service Quality and Customer Trust. *Journal of Service Research*, *10*(3), 256–268. https://doi.org/10.1177/1094670507310769

Hilton, T., Hughes, T., Little, E., & Marandi, E. (2013). Adopting self-service technology to do more with less. *Journal of Services Marketing*, *27*(1), 3–12. https://doi.org/10.1108/08876041311296338

Ho, T.-H., Park, Y.-H., & Zhou, Y.-P. (2006). Incorporating Satisfaction into Customer Value Analysis: Optimal Investment in Lifetime Value. *Marketing Science*, *25*(3), 260–277. https://doi.org/10.1287/mksc.1050.0158

Homburg, C., Vollmayr, J., & Hahn, A. (2014). Firm Value Creation through Major Channel Expansions: Evidence from an Event Study in the United States, Germany, and China. *Journal of Marketing*, *78*(3), 38–61. https://doi.org/10.1509/jm.12.0179

Huang, L., Lu, X., & Ba, S. (2016). An empirical study of the cross-channel effects between web and mobile shopping channels. *Information & Management*, *53*(2), 265–278.

https://doi.org/10.1016/J.IM.2015.10.006

Jo Bitner, M., Faranda, W. T., Hubbert, A. R., & Zeithaml, V. A. (1997). Customer contributions and roles in service delivery. *International Journal of Service Industry Management*, *8*(3), 193–205. https://doi.org/10.1108/09564239710185398

Kettunen, E., & Kari, T. (2018). Can Sport and Wellness Technology be My Personal Trainer? – Teenagers and Digital Coaching. *Digital Transformation – Meeting the Challenges*, 463–476. https://doi.org/10.18690/978-961-286-170-4.32

Li, H., Lobschat, L., & Verhoef, P. C. (2018). Multichannel Retailing: A Review and Research Agenda. *Foundations and Trends® in Marketing*, *12*(1), 1–79. https://doi.org/10.1561/1700000059

Li, J., Konuş, U., Langerak, F., & Weggeman, M. C. D. P. (2017). Customer Channel Migration and Firm Choice: The Effects of Cross-Channel Competition. *International Journal of Electronic Commerce*, *21*(1), 8–42. https://doi.org/10.1080/10864415.2016.1204186

mHealth apps, App Store 2018 | Statistic. (n.d.). Retrieved April 3, 2019, from https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/

Montaguti, E., Neslin, S. A., & Valentini, S. (2016). Can Marketing Campaigns Induce Multichannel Buying and More Profitable Customers? A Field Experiment. *Marketing Science*, *35*(2), 201–217. https://doi.org/10.1287/mksc.2015.0923

Ng, M. K., & Chung, Y. (2012). Double Mover-Stayer model on customer switching in telecommunications industry. *Naval Research Logistics (NRL)*, *59*(8), 663–674. https://doi.org/10.1002/nav.21511

Oh, L.-B., Teo, H.-H., & Sambamurthy, V. (2012). The effects of retail channel integration through the use of information technologies on firm performance. *Journal of Operations Management*, *30*(5), 368–381. https://doi.org/10.1016/J.JOM.2012.03.001

Retana, G. F., Forman, C., & Wu, D. J. (2016). Proactive Customer Education, Customer Retention, and Demand for Technology Support: Evidence from a Field Experiment. *Manufacturing & Service Operations Management*, *18*(1), 34–50. https://doi.org/10.1287/msom.2015.0547

Roels, G. (2014). Optimal Design of Coproductive Services: Interaction and Work Allocation. *Manufacturing & Service Operations Management*, *16*(4), 578–594. https://doi.org/10.1287/msom.2014.0495

Spanjol, J., Cui, A. S., Nakata, C., Sharp, L. K., Crawford, S. Y., Xiao, Y., & Watson-Manheim, M. B. (2015). Co-Production of Prolonged, Complex, and Negative Services. *Journal of Service Research*, *18*(3), 284–302. https://doi.org/10.1177/1094670515583824

Suh, M., Greene, H., Israilov, B., & Rho, T. (2015). The Impact of Customer Education on Customer Loyalty Through Service Quality. *Services Marketing Quarterly*, *36*(3), 261–280. https://doi.org/10.1080/15332969.2015.1046776

Temerak, M. S., Winklhofer, H., & Hibbert, S. A. (2018). Facilitating customer adherence to complex services through multi-interface interactions: The case of a weight loss service. *Journal of Business Research*, *88*, 265–276. https://doi.org/10.1016/J.JBUSRES.2018.03.029

Usage of health apps to receive fitness instructions among US adults by age 2017 | Survey. (2017). Retrieved April 4, 2019, from Statista, The Statistics Portal website: https://www.statista.com/statistics/699312/us-adults-that-use-apps-to-get-fitness-instructions-by-age/

Wallace, D. W., Giese, J. L., & Johnson, J. L. (2004). Customer retailer loyalty in the context of multiple channel strategies. *Journal of Retailing*, *80*(4), 249–263. https://doi.org/10.1016/J.JRETAI.2004.10.002

Wang, K., & Goldfarb, A. (2017). Can Offline Stores Drive Online Sales? *Journal of Marketing Research*, *54*(5), 706–719. https://doi.org/10.1509/jmr.14.0518

Wirtz, J., & Mattila, A. S. (2003). The effects of consumer expertise on evoked set size and service loyalty. *Journal of Services Marketing*, *17*(7), 649–665. https://doi.org/10.1108/08876040310501223

Xiao, Y., Chen, F. Y., & Chen, J. (2009). Optimal inventory and dynamic admission policies for a retailer of seasonal products with affiliate programs and drop-shipping. *Naval Research Logistics*, *56*(4), 300–317. https://doi.org/10.1002/nav.20346

DECISION SCIENCES INSTITUTE

Going Digital to Conform and to Perform:
Learning Mechanisms Underpinning the Budgetary Decision on Digital Advertising

Seoyoung Kim
University of Georgia
Email: seoyoungkim@uga.edu

Sundar Bharadwaj
University of Georgia
Email: sundar@uga.edu

ABSTRACT

Firms are increasingly allocating their advertising budgets to digital advertising media, but little is known about the drivers of the budgetary decision. Drawing on neo-institutional theory and multiple learning theory lenses, the study delineates six learning mechanisms shaping firms' decisions on digital advertising budgets, aimed at legitimacy and performance. The empirical results from a structural panel vector autoregression model show that firms learn from a wide array of stakeholders to attain legitimacy and performance, and the effects of six learning motivations on the budgetary decision are contingent upon the learning opportunity and ability of firms.

KEYWORDS: Resource Management and Allocation, Strategic Decision Making, Organizational Learning, Strategy Process, and Electronic Market Places

DECISION SCIENCES INSTITUTE
Win the Buy Box: Product Characteristics and Direct Sales Choice

Hao Su
University of Maryland
Email: haosu@umd.edu

Martin E. Dresner
University of Maryland
Email: mdresner@umd.edu

ABSTRACT

The study examines the factors that predict the marketplace operator's decision to directly sell a product. Using data from Amazon, we find that marketplace operators are more likely to offer direct sales, (1) when product/service quality is more difficult to ascertain, (2) when products are offered by well-known brands, and (3) when their brands offer a higher number of products on the marketplace. We further investigate under which scenarios Amazon wins the competition against third-party sellers. The results suggest that Amazon is more likely to win placement in the Buy Box even if it does not have the lowest price.

KEYWORDS: Online marketplaces, Competition, Information asymmetry, Product/Service quality, Brand, Distribution intensity, Antitrust

DECISION SCIENCES INSTITUTE

Use of Discrete Event Simulation to Enhance Failure Mode & Effect Analysis (FMEA)

Adil Abdukeyim
Illinois State University
Email: aabuduk@ilstu.edu

Dr. Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

To improve system performances and process dependability, analyzing the system accurately is an essential step but difficult to achieve and it is even more challenging if the system is complex and dynamic. In this paper, one popular analysis tool called, Failure Mode & Effect Analysis (FMEA), is described for its common usage. An alternative methodology, Discrete Event Simulation, is also introduced and compared with FMEA as a way of overcoming the inherent weaknesses of FMEA.

KEYWORDS: Discrete Event Simulation, Failure Mode & Effect Analysis (FMEA), Performance Improvement, Dependability Analysis

INTRODUCTION

Continuous improvement of products and processes plays a significant role for companies to have competitive advantages in highly challenging markets such as manufacturing, service sector, and healthcare (Doshi & Desai, 2017). Hence, organizations need effective approaches to achieve sustainable and ongoing performance improvement effort to achieve such a goal. One such methodology, Lean Six Sigma, is a comprehensive practice to eliminate the defects and wastes in a system with various improvement tools to address the system related issues.

Among these tools, FMEA (Failure Modes & Effect Analysis) is widely adopted by industries, which is a systematic method designed for dependability analysis (Scipioni & Saccarola & Centazzo & Arena, 2002). It is used to recognize possible failures and their impacts on processes and products (Doshi & Desai, 2017). FMEA was introduced for the first time in 1949 by the US army. During the 1970s, due to its powerful and valid features, its implementation also spreaded into aerospace, automotive, and general manufacturing (Scipioni & Saccarola & Centazzo & Arena, 2002).

FMEA is a team activity that involves, first, studying and evaluating the processes or products (Scipioni & Saccarola & Centazzo & Arena, 2002); then, listing the possible failures and their effects; and, lastly, establishing future actions that could eliminate or reduce potential failures (Hekmatpanah & Shabin & Ravichandran, 2011). To determine the level of potential risks related to the failure of a certain element of the process, the risk priority number (RPN) is calculated by multiplying its occurrence, severity, and detection of a failure. Each element is ranked based on a 1 to 10 scale using its RPN value, and higher the RPN is, the more urgency of the future actions is required (Hekmatpanah & Shabin & Ravichandran, 2011).

FMEA is an effective tool to improve processes by proactively identifying and preventing high risk elements in the system. The case study done by Ravichandran, Hekmatpanah, and Shahin (2011) shows that after implementing FMEA, the net profit of the organization in the study increased significantly. Another study indicates that when FMEA is integrated with HACCP (Hazard Analysis Critical Control Points) system in a food company, it greatly improved the process reliability of the system in the company (Scipioni & Saccarola & Centazzo & Arena, 2002).

PROBLEM STATEMENT

As much as we want to improve processes and improve the system dependability, analyzing a real-life system is difficult, and it is even more challenging if we deal with a complex system (Sumari & Ibrahim & Zakaria & Ab Hamid, 2013). FEMA has been one of those tools for the system analysis. When FMEA is applied as an analytical tool to enhance a system, however, it brings with it some of the inherent issues. During the analysis phase, for example, staff from different departments use heavily their knowledge and past experience in a subjective manner when quantifying the level of risks (Scipioni & Saccarola & Centazzo & Arena, 2002). As a result, the RPN can fluctuate quite a bit depending on whose opinions are valued more. Subjective judgment (Hekmatpanah & Shabin & Ravichandran, 2011) and implicit knowledge (Murphy & Heaney & Perera, 2011) were often involved, making the outcomes of the FMEA analysis not reliable and inconsistent because of its methodological bias (Hekmatpanah & Shabin & Ravichandran, 2011). Therefore, FMEA may become a practice with less confidence by the users, driven more by qualitative analysis even though it has to be objective and quantitative (Murphy & Heaney & Perera, 2011). This indicates that there is a need for FMEA to be more rigorous and reliable as a method in order to appeal to operations managers.

RESEARCH METHODOLOGY

To analyze the system quantitatively, a more general tool, called a discrete event simulation, has been used in various industry applications to model a process for performance improvement. Initial studies using the discrete event simulation could be traced back to the 1950s when the parallel and distributed simulation was operated by null-message protocol (Vanmechelen & Munck & Broeckhove, 2012). Currently, it is often adopted in process-oriented industry sectors such as manufacturing, logistics, mining, and energy, to name a few (Wohlgemuth & Page & Kreutzer, 2006). It provides a better understanding of a real-life system by emulating the system in place. This method represents series of steps at different points of a system process and changes their states in a discrete time manner (Sumari & Ibrahim & Zakaria & Ab Hamid, 2013). It is established based on the events in the system (Vanmechelen & Munck & Broeckhove, 2012). When the state changes, different events will occur. As a result, agents move through the simulation flow which represents the process of the system being simulated. Here, an agent is any object in a system that needs clear representation that needs to be processed. Some examples of the agents are: products, customers, patients, transportation, and any entities or objects being processed in the system. (Banks & Carson, 1986). On the other hand, an event could be viewed as an activity or a task. Such examples are: waiting, payment, treatment, (Sumari & Ibrahim & Zakaria & Ab Hamid, 2013), delivering, and other similar activities.

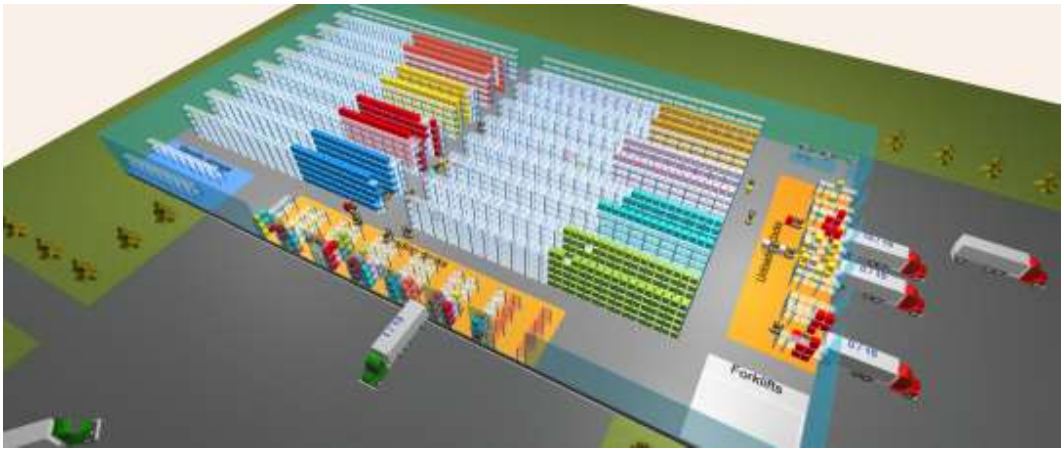
Discrete event simulation is a powerful methodology. A study of an industrial environmental protection issue by Wohlgemuth & Page & Kreutzer (2006), for example, reveals the discrete event simulation enabled the analysis of various aspects of process changes, resulting in the improvement of both tactical and strategic management in protecting the environment. Another example would be a study where the discrete event simulation was used for the healthcare

team to improve resource-utilization practices (Raunak & Osterweil & Wise & Clarke & Henneman, 2009).

CASE STUDY

In this paper, as a way of demonstrating how the dependability analysis capability of FMEA can be enhanced, the discrete event simulation is introduced and directly compared against FMEA. To better illustrate the main differences between FMEA and discrete event simulation, a case involving a distribution center (DC) will be used. A visual snap shot of the DC used is shown in Figure 1 (Note the DC simulation model is an existing example model provided in the AnyLogic simulation software. Only the FMEA was conducted by the researchers independently).

Figure 1. An operational scene of the distribution center used in the case



In this case study, the processes in distribution center involves three main steps. First, the distribution center receives pallets delivered by trucks. Pallets are then unloaded from the trucks using forklifts and placed in the receiving dock area (*the right-hand side of the building in Figure 1*). After this, other forklifts move the pallets to the main storage area. Second, orders from distribution center's clients are assembled from the pallets according to their needs. In this step, enough space for the assembly area near the loading docks is needed (or a backup storage area can be used in case there is not enough space there), and main storage must have the required number of pallets for the orders, which are assembled by the forklifts. Third, after the orders are assembled, a truck is assigned to the loading dock. Then forklifts will load the pallets into the truck and pallets are ready to be delivered to clients.

There are two parts of dependability analysis conducted in the case study. In Part A, FMEA is applied first to analyze the processes to evaluate possible failures in the system, the causes of the failures, effects, RPN, and future actions. After this is finished, its limitations are discussed. In Part B, as comparison, the discrete event simulation method is applied to explore the same processes in the distribution center. As complements to FMEA's drawbacks, the discrete event simulation's features are discussed in this part.

Part A: FMEA Analysis

In this part, a detailed FMEA analysis of the distribution center is described. As a first step, the processes should be studied carefully or the domain experts should lead the analysis in order to distinguish the possible failures. There are three main phases to focus on for the evaluation of various potential failures. As shown in Table 1, the processes of the distribution center are divided into 1) the unloading phase; 2) the assembling phase; and 3) the loading phase. And

each phase relies on a few critical operational resources such as trucks, forklifts, and pallets for the processes to operate as intended.

Table 1: Main phases and related critical resources, and process details

Main Phases / Critical Resources	Phase 1	Phase 2	Phase 3
	Unloading	Assembling	Loading
Unloading trucks Loading trucks Forklifts Pallets Pallet Racks	The distribution center receives pallets delivered by trucks. Pallets are then unloaded from the trucks using forklifts and placed in the receiving dock area. After this, other forklifts move the pallets to the main storage area.	Orders from distribution center's clients are assembled from same or different types of pallets by forklifts accordingly.	After the orders are assembled, a truck is assigned to the loading dock. Then forklifts will load the pallets into the truck.

After the preliminary understanding of the processes in the distribution center, the second step is to evaluate possible failures, the causes of the failures, their effects, and the current controls, and calculate the risk priority number (RPN) related to the critical resources over different phases. As the last step, the action priorities are determined based on the RPN values and suggested future actions are provided to prevent the identified failures from happening, hence the processes could be enhanced over time. Table 2 shows a section of the complete FMEA analysis related to the failures to do with the forklift resource (*The entire FMEA analysis is included in the Appendix*). Under the "Forklifts" category, there are several sub-categories of what could go wrong and its more specific failure items are identified. For example, as highlighted with the red line, under the "Amount of Forklifts", "Less than needed" is identified as one of its possible failures. For this possible failure, "Insufficient understanding of daily need for forklifts" is listed as a cause of the failure. Then, "A: May cause longer operation time. Hence, may delay the orders", "B: May cause stress for employees", and "C: May Shorten forklifts' lifetime" are assessed as the effects of the failure. After establishing these, RPN is calculated to be 96 based on the level of perception of the occurrence, severity, and detection of the effects

of the failure. The RPN value (96) is relatively higher with respect to other failure items, which justifies the Action Priority as “Urgent”. As a result, a Suggested Action, “Try to understand the need for forklifts based on the order flow and adjust the amount”, is recommended to prevent the perceived failure.

Table2: Example of action priority and suggested action

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severenity	Detection	Current Control	RPN	Action Priorty	Suggested Action
Forklifts		More than needed	Insufficient understanding of daily need for forlifts	May increase labor cost, maintenance cost, and purchase fee of forklifts	3	5	3	Inspecting Forklift Utilization	45	Relatively High	Try to understand the need for forklifts based on the order flow, and adjust the amount
	Amount of Forklifts	Less than needed	Insufficient understanding of daily need for forlifts	A:May cause longer opreration time. Hence, may delay the orders. B: May cause stress for employees. C: May Shorten forklifts' lifetime	4	8	3	Inspecting Forklift Utilization	96	Urgent	Try to understand the need for forklifts based on the order flow, and adjust the amount.
	Speed of Forklifts	Overspeed	Driving too fast	May cause safety problem during the operation time.	4	7	2	Visual inspection	56	Relatively High	Train employees regulary and emphasize the both safety, and work efficiency.
		Underspeed	Driving too slow	May cause longer operation time, hence may delay the orders.	4	6	2	Visual inspection	48	Relatively High	Train employees regulary and emphasize the both safety, and work efficiency.
	Maintenance of Forklifts	Damages during the operation	Fail to maintain regularly	A: May cause safety problem during the operation B: May cause longer operation time, hence may delay the orders C: May shorten the forklifts' lifetime D: May increase maintenance cost	3	7	5	Regular maintenance	105	Urgent	Follow the regular maintenance. Also, implement monitering solution for it.
	Forklift Dock	Unproportioned space	Wrong decision over space for forklifts	A: May cause poor utilization of forklifts (if the space is insufficient) B: May increase unnecessary waste of space	2	6	2	Visual inspection	24	Low	Adjust the space for forklift dock based on analyzing the current amount of forklifts and space.

As much as the FMEA is a solid evaluation tool to assess the level of failure for potential issues, it has inherent weaknesses that need to be addressed as a system improvement tool. First, FMEA is a subjective analysis. This may be the vital disadvantage of FMEA. During the FMEA approach, one of the main challenges is that it relies on tacit knowledges and past experiences of the analysis participants. Almost every step involves subjective understanding and opinions. This may result in different or inconsistent outcomes from the FMEA analysis. This could be significant because when calculating the RPN, the subjective opinions and inputs heavily influences how the occurrence, severity, and detection are determined. One may have totally different views on all these elements than others. One may assess the severity of the less forklifts to be 3 and others may consider it to be 9, which could produce a large deviation in the RPN value, which results in quite different action priority and therefore the suggested actions are taken more seriously or less seriously. Since organizations rank their future actions based on action priorities, this type of subjective analysis of FMEA may lead to possible economic loss, and magnify the problems in the system even bigger.

Second, FMEA could possibly miss possible failures all together. This may happen because, during the FMEA, analysis is done without fully visualizing entire processes or really understanding the related processes depending on and affecting each other. There may be

possibilities for the FMEA participant to let some small but important aspects of the process not be considered in the evaluation. For example, a tight operational space for forklifts is one of the elements that could be missed or overlooked, yet it is very crucial. FMEA participants may primarily focus on just the size of the forklift fleet but forget the machines need adequate space to be productive in the distribution center. If the space is insufficient, it may cause longer operation time and safety problems. Not being able to consider every possible failures, seemingly small or large, happens often, which leads to a devastating and unexpected results when its solution is implemented. In addition to that, if any subtle changes happen in the processes, the effects of these changes are hard to detect since there is no direct way to measure the impact of the changes in FMEA. This may also result in another possible failures not considered in the analysis.

In summary, FMEA is a strong operation improvement tool to enhance the system performance by eliminating the potential reasons of the system failure, it is based on the subjective assessment of the failures. However, in reality these failures are not isolated cases but rather they are often entangled each other. One failure, small or large, may have come from another one issue, and may produce subsequent failures down the line of the processes.

Part B: Discrete Event Simulation Analysis

To overcome the inherent weaknesses addressed in the previous section, in this part, the analysis using the discrete event simulation is described as an alternative method to improve the distribution center processes. The simulation model used is provided by AnyLogic simulation software, which is an industry leading simulation platform. The software supports the multi-simulation paradigm, one of them is the discrete event method.

Figure 2 shows the underlying logic of the simulation where the unloading and loading processes represented using the simulation blocks such as cues and delays. The assembly process is embedded into loading process. Also, the “initial filling of the storage” represents placing the pallets in the main storage area. The other two processes, such as “moving from the storage to a moving dock”, and “moving from a standby storage to a moving dock” represent placing the assembled pallets in the loading dock area. Every block in the logic represents an event. For example, the block “truckUnloading” is the event of forklifts unloading the pallets from trucks. The agents (pallets) as the main target of the production at the distribution center go through multiple events which all together represent the entire operation process. What is critically important is that the simulation logic is the basis for explaining the causes and effects of the system in operation.

Along with the logic part, the simulation model has the layout of the distribution center to display the progress of the simulation. The layout is presented in Figure 3 where the main storage area, standby storage, loading and unloading docks are presented. These elements help analyzing the processes while the simulation is running.

Figure 2:Details of the simulation logic of the distribution center

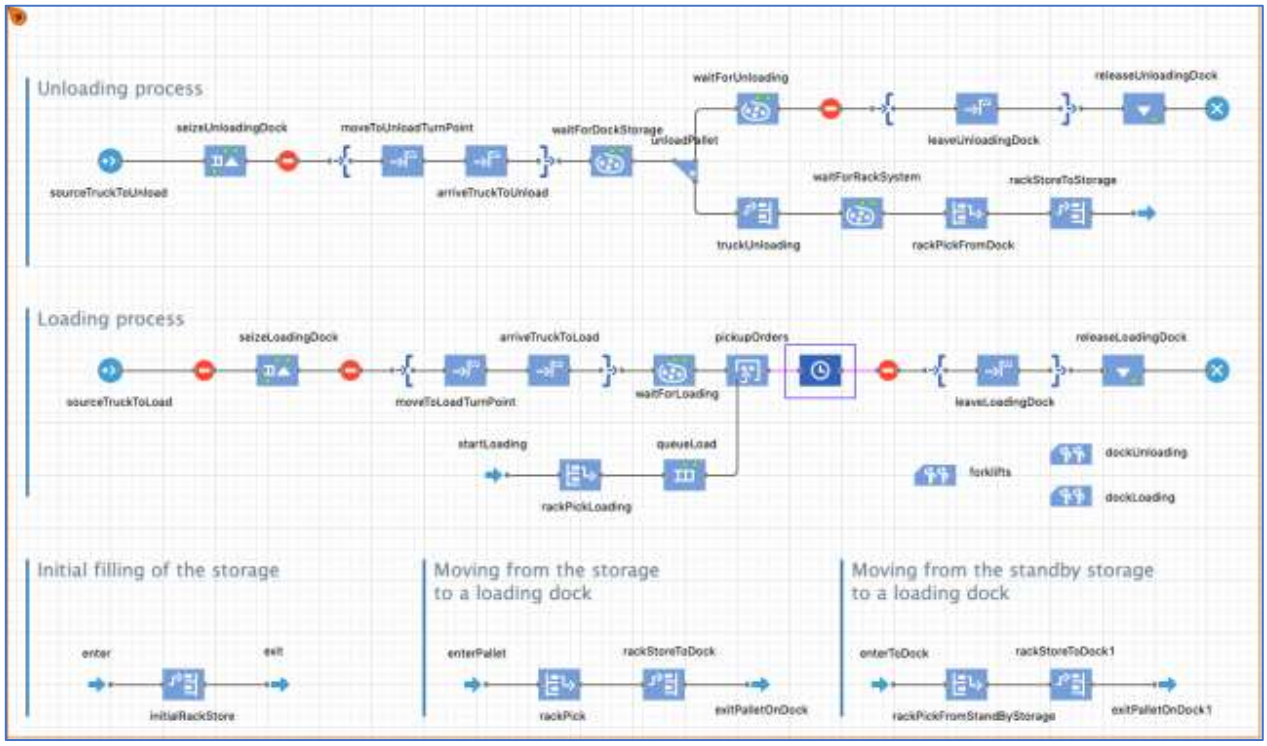
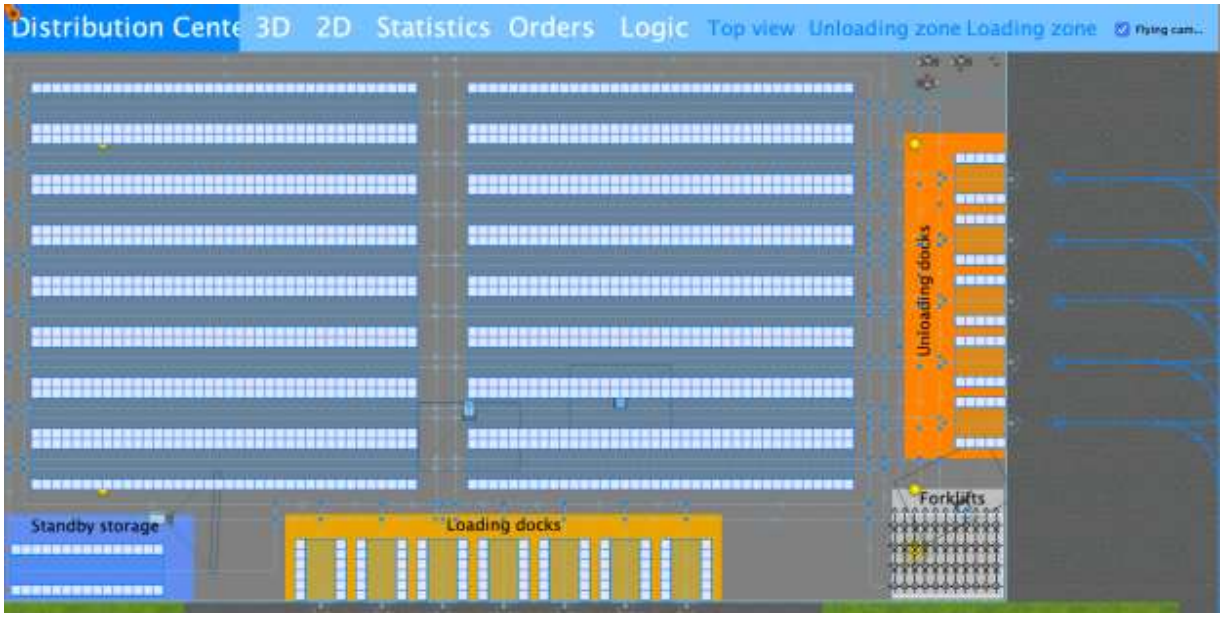


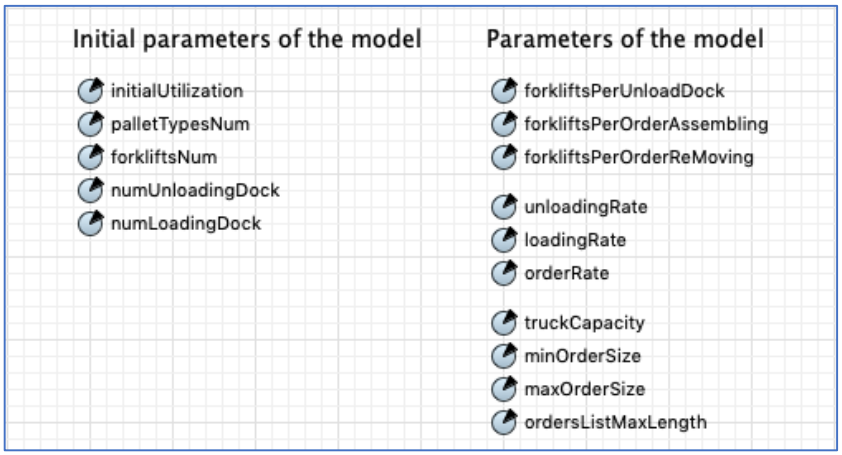
Figure 3: Layout used in the distribution center simulation



The additional key aspect of the discrete event simulation is that it runs on a set of objective data in the form of “parameters”. They are the numerical data that represent different

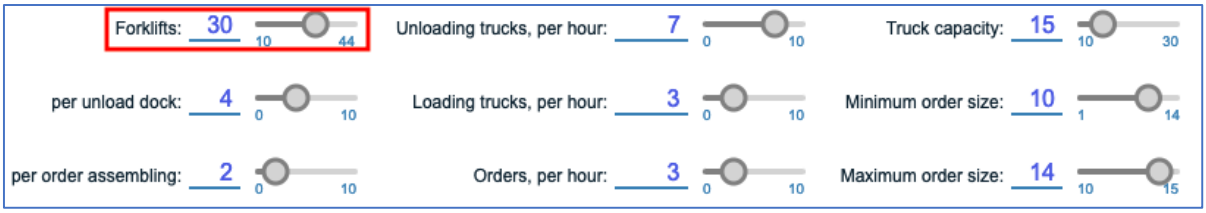
characteristics of the operation. As shown in Figure 4, “unloadingRate” and “maxOrderSize” are some of the parameters by which the operation of the distribution center is controlled. They can be modified during the simulation in real-time by the users to see how a certain parameters affect the system outcomes as shown in Figure 5 below.

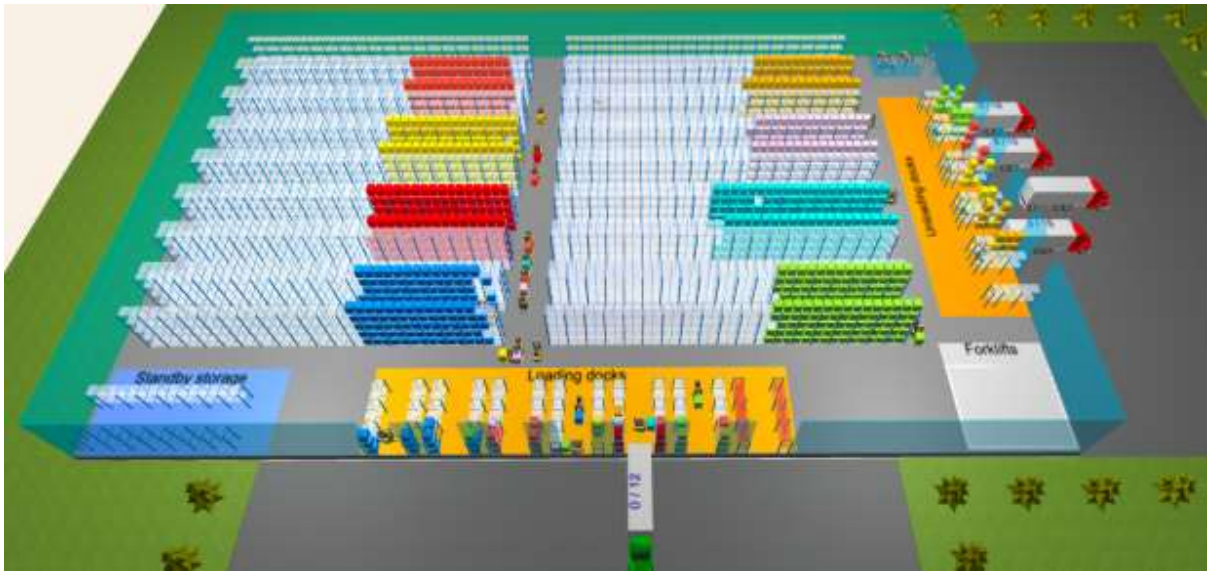
Figure 4: Parameters in the simulation



After the logic, layout, and other related features are defined, the simulation now is ready to run. The resulting operational scene is shown in Figure 6.

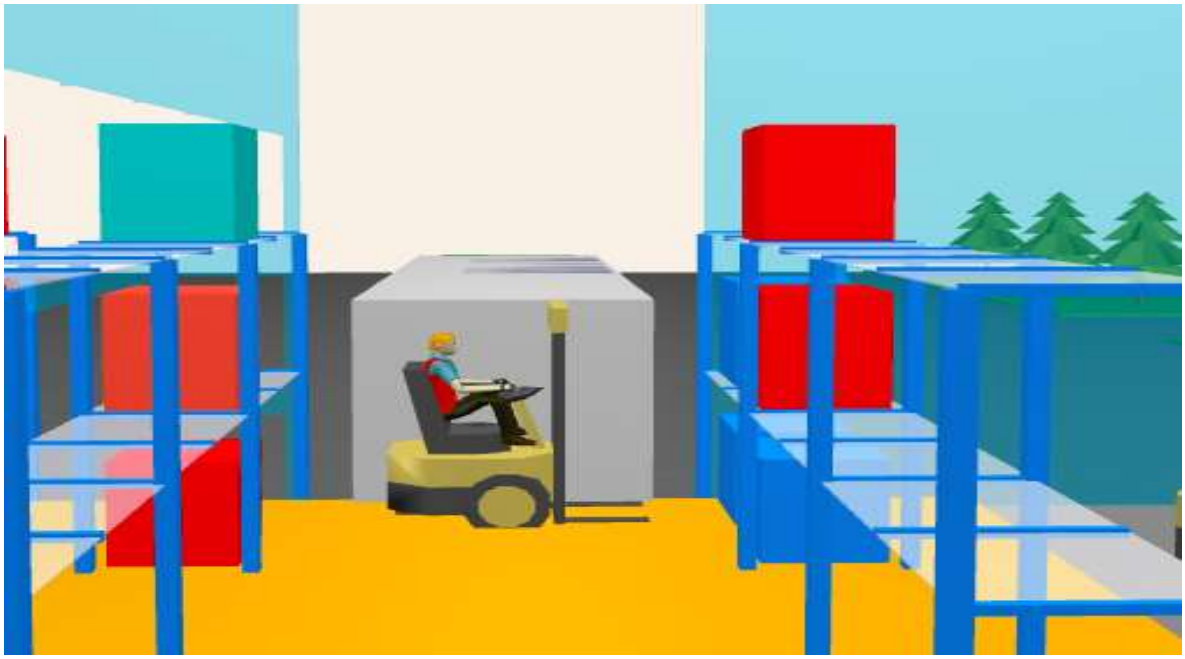
Figure 5: Controls of parameters for the distribution center operation





The simulation displays all the elements in the distribution center. Unloading trucks, loading truck, and forklifts moving in the center can provide an effective visualization to thoroughly evaluate and explore the processes. It is easy to see what is happening in the center. This may lead to better understanding of the center processes at various stages, allowing users to confirm their understanding and to resolve any discrepancies between collaborators. This is certainly an advantage of the discrete event simulation to complement the FMEA analysis where everyone's visualization is not necessarily on the same page. During the simulation, it gives us an opportunity to analyze the processes from different angles and aspects. The simulation offers a virtual reality that the modeled objects and their interactions are displayed. This may maximize the chance for detecting hidden failures that may have not been discussed or considered in the FMEA analysis because not everyone sees the process the same way. For example, let's assume accelerating the loading process is desired in the loading docks area which is shown Figure 6. The visualized operation shown in the discrete event simulation can show us potential safety concerns in the loading dock area, previously not evaluated but possible to affect the well-being and productivity of the workers. However, without actually seeing the same process in FMEA, it is difficult to detect such a flaw.

Figure 6: Close-up view of the loading process



Moreover, the two major drawbacks of FMEA, the subjective analysis and missing possible failures can be prevented effectively in the simulation platform since the discrete event model runs objective data and the series of logical blocks to explain the causes and effects of the system. Specifically, the discrete event simulation requires quantifiable information as numerical data for the simulation inputs, which are linked through the established sequence logic using the modeling blocks. This is what makes an accurate cause-effect analysis possible, not based on one's guesses and subjective perceptions as in FMEA. It is one of the foundational strengths of the simulation. Additionally, since the simulation constantly provides the trends of various key performance indicators during the model execution, the users can develop immediate understanding on how the system is doing under certain conditions, taking immediate measures to improve the processes. For example, Figure 7 shows the dashboards where the statistics on key performance indicators are displayed during simulation when the forklift fleet size is 30. This

shows us how many forklifts would be adequate for the distribution center operation considering the various indicators such as the processing time, the utilization ratio, and the available storage space. Hence future decisions can be made more thoughtfully and realistically and the action priorities can be more rigorously evaluated based on quantitative information, not by individuals' opinions. Furthermore, the impacts of changes for the future actions can be justified easily. For example, as shown in Figure 8, users can tell the difference made on the utilization, the storage space, and processing time when the fleet size shrinks to 10. The effect of the change is also measurable, which can be a strong basis to make a future decisions with confidence in order to prevent future financial loss, safety problems, and bad customer experience.

Figure 7: Statistics on key performance indicators when forklift fleet size = 30

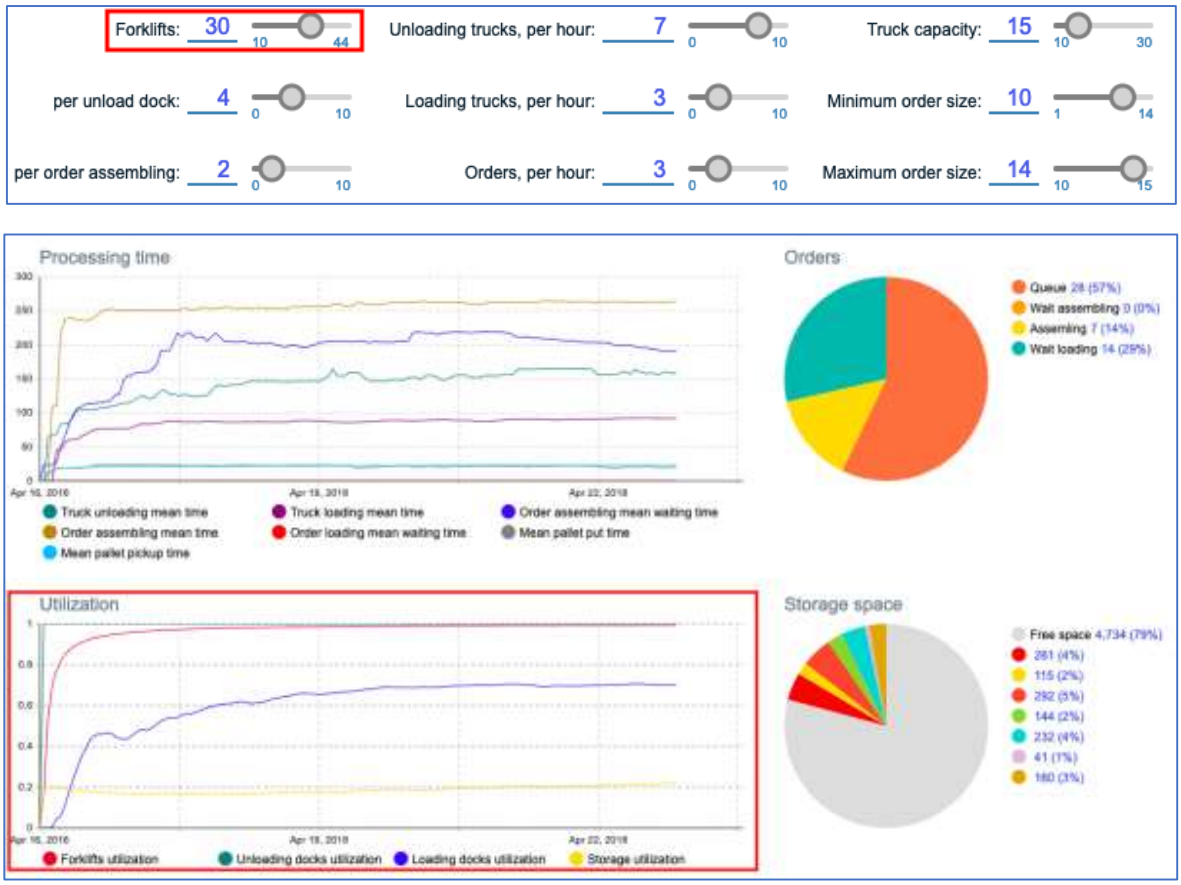


Figure 8: Statistics on key performance indicators when forklift fleet size = 10

CONCLUSION

In this article, through the case of a distribution center operation, the different features of FMEA and the discrete event simulation are described and compared. Although both FMEA and the discrete event simulation are strong methods to enhance the system processes, the discrete event simulation has some strong features that can resolve the weaknesses of FMEA. It is prone to being under the subjective evaluation of the participants and missing possible failures may go unnoticed. The discrete event simulation's strong data-based cause-effect analysis and additional visualization information boosts the possibility of brainstorming more on critical potential failures and credibility on the future actions. Hence it can provide more evidential decisions that may help increase the system performances and reliabilities in the future. In short, through its data-driven approach, the discrete event simulation offers a rigorous analysis capability on cause and effects of a possible failure in the processes, while FMEA relies on implicit knowledge and experience from different participants.

In the end, there are numerous methods to improve the processes and system performances today. FMEA and the discrete event simulation are among the popular tools but they need to be used in unison, rather than one over the other. Though technical, the discrete event simulation is easy to learn and very adaptable to various operation modeling situations. It is a well-known challenge for organizations to select the right tools and methods to remain competitive (Doshi &

Desai, 2017) but bolstering an existing practice (e.g. FMEA) with a complementing method (e.g. discrete event simulation) would be an effective and affordable approach.

REFERENCES

1. Banks, J., J.S. Carson. (1986). WSC '86: Proceedings of the 18th conference on Winter simulation(Vol. 211, pp. 17-23) (Banks, J., J. S. Carson, Author; J. R. Wilson & J. O. Henrikson, Eds.). New York, NY: Association for Computing Machinery. doi:10.1145/318242.318253
2. Doshi, J., & Desai, D. (2017). Application of failure mode & effect analysis (FMEA) for continuous quality improvement - multiple case studies in automobile SMEs. International Journal for Quality Research,11, 345-360. doi:10.18421/IJQR11.02-07
3. Hekmatpanah, M. (2011). The application of cause and effect diagram in the oil industry in Iran: The case of four liter oil canning process of Sepahan Oil Company. African Journal of Business Management,5(26). doi:10.5897/ajbm11.1517
4. Murphy, M., Heaney, G., & Perera, S. (2011). A methodology for evaluating construction innovation constraints through project stakeholder competencies and FMEA. Construction Innovation,11(4), 416-440. doi:10.1108/14714171111175891
5. Raunak, M., Osterweil, L., Wise, A., Clarke, L., & Henneman, P. (2009). Simulating patient flow through an Emergency Department using process-driven discrete event simulation. 2009 ICSE Workshop on Software Engineering in Health Care. doi:10.1109/sehc.2009.5069608
6. Scipioni, A., Saccarola, G., Centazzo, A., & Arena, F. (2002). FMEA methodology design, implementation and integration with HACCP system in a food company. Food Control,13(8), 495-501. doi:10.1016/s0956-7135(02)00029-4
7. Sumari, S., Ibrahim, R., Zakaria, N. H., & Hamid, A. H. (2013). Comparing Three Simulation Model Using Taxonomy: System Dynamic Simulation, Discrete Event Simulation and Agent Based Simulation. International Journal of Management Excellence,1(3), 54. doi:10.17722/ijme.v1i3.9
8. Vanmechelen, K., Munck, S. D., & Broeckhove, J. (2012). Conservative Distributed Discrete Event Simulation on Amazon EC2. 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012). doi:10.1109/ccgrid.2012.73
9. Wohlgemuth, V., Page, B., & Kreutzer, W. (2006). Combining discrete event simulation and material flow analysis in a component-based approach to industrial environmental protection. Environmental Modelling & Software,21(11), 1607-1617. doi:10.1016/j.envsoft.2006.05.015

APPENDIX: Complete FMEA Analysis (DISTRIBUTION CENTER)

Table 3: Complete FMEA analysis on distribution center

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severerity	Detection	Current Control	RPN	Action Priorty	Suggested Action
Forklifts	Amount of Forklifts	More than needed	Insufficient understanding of daily need for forlifts	May increase labor cost, maintenance cost, and purchase fee of forklifts	3	5	3	Inspecting Forklift Utilization	45	Relatively High	Try to understand the need for forklifts based on the order flow, and adjust the amount.
		Less than needed	Insufficient understanding of daily need for forklifts	A:May cause longer opreration time. Hence, may delay the orders. B: May cause stress for employees. C: May Shorten forklifts' lifetime.	4	8	3	Inspecting Forklift Utilization	96	Urgent	Try to understand the need for forklifts based on the order flow, and adjust the amount.
	Speed of Forklifts	Overspeed	Driving too fast	May cause safety problem during the operation time.	4	7	2	Visual inspection	56	Relatively High	Train employees regulary and emphasize the both safety, and work efficiency.
		Underspeed	Driving too slow	May cause longer operation time, hence may delay the orders.	4	6	2	Visual inspection	48	Relatively High	Train employees regulary and emphasize the both safety, and work efficiency.
	Maintenance of Forklifts	Damages during the operation	Fail to maintain regularly	A: May cause safety problem during the operation B: May cause longer operation time, hence may delay the orders C: May shorten the forklifts' lifetime D: May increase maintenance cost	3	7	5	Regular maintenance	105	Urgent	Follow the regular maintenance. Also, implement monitering solution for it.
	Forklift Dock	Unproportioned space	Wrong decision over space for forklifts	A: May cause poor utilization of forklifts (if the space is insufficient) B: May increase unnecessary waste of space	2	6	2	Visual inspection	24	Low	Adjust the space for forklift dock based on analyzing the current amount of forklifts and space.

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severerity	Detection	Current Control	RPN	Action Priorty	Suggested Action
Unloading Docks	Number of Docks	More than needed	Insufficient understanding of order flow	A: May Cause unnecessary waste of space B: May increase unnecessary management cost.	4	5	2	Visual inspection	40	Relatively High	Analyze the order flow, and adjust the number of docks based on it.
		Less than needed	Insufficient understanding of order flow	A: May cause longer unloading time, hence may delay the orders. B: May cause stress to employees due to pressure from unloading process.	4	7	2	Visual inspection	56	Relatively High	Analyze the order flow, and adjust the number of docks based on it.
	Space of Docks	More than needed	Wrong decision on space for docks	A: May cause unnecessary waste of space. B: May increase unnecessary management cost.	2	5	2	Visual inspection	20	Low	Analyze the order flow, pallet flow, and adjust the sapce of docks based on it.
		Less than needed	Wrong decision on space for docks	A: There will not be enough space for palletRacks, hence may cause longer unloading time. B: May cause stress to employees.	3	6	3	Visual inspection	54	Relatively High	Analyze the order flow, pallet flow, and adjust the sapce of docks based on it.
Loading Docks	Number of Docks	More than needed	Insufficient understanding of order flow	A: Causes unnecessary waste of space B: May increase unnecessary management cost.	2	5	2	Visual inspection	20	Low	Analyze the order flow, and adjust the number of docks based on it.
		Less than needed	Insufficient understanding of order flow	A: May cause longer loading time, hence may delay the orders. B: May cause stress to employees due to pressure from loading process.	3	7	2	Visual inspection	42	Relatively High	Analyze the order flow, and adjust the number of docks based on it.
	Space of Docks	More than needed	Wrong decision on space for docks	A: May cause unnecessary waste of space. B: May increase unnecessary management cost.	4	5	2	Visual inspection	40	Relatively High	Analyze the order flow and adjust the sapce of docks based on it.
		Less than needed	Wrong decision on space for docks	A: There will not be enough space for palletRacks, hence may cause longer loading time. B: May cause stress to employees. C: May delay the orders. D: May cause over-reliance on standby storage.	4	7	2	Visual inspection	56	Relatively High	Analyze the order flow and adjust the sapce of docks based on it.

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severerity	Detection	Current Control	RPN	Action Priorty	Suggested Action
Unloading Trucks	Capacity of Unloading Trucks	More than needed	Decided by the manufacturer	A: May cause over-inventory B: May increase management cost due to over-inventory	3	6	3	Manufacturer's responsibility	54	Relatively High	Based on the order flow, give manufacturer feedback to adjust truck capacity.
		Less than needed	Decided by the manufacturer	A: May decrease the work efficiency. B: May delay the orders.	3	7	3	Manufacturer's responsibility	63	High	Based on the order flow, give manufacturer feedback to adjust truck capacity.
	Maintenance of Unloading Trucks	Damages during the delivery	The manufacturer fails to maintain regularly	A: May cause safety problem during the delivery B: May cause longer delivery time, hence may delay the orders	2	7	2	Manufacturer's responsibility	28	Low	Give feedback to the manufacturer
	Amount of Unloading Trucks	Less than needed	Decided by the manufacturer	May delay the orders.	4	7	4	Manufacturer's responsibility	112	Urgent	Based on the order flow, give manufacturer feedback to adjust the amount of trucks they possess.
Loading Trucks	Capacity of Loading Trucks	More than needed	Insufficient understanding of order flow	May decrease the efficiency of every delivery.	3	6	2	N/A	36	Relatively High	Based on the order flow, adjust truck capacity.
		Less than needed	Insufficient understanding of order flow	A: May delay the orders. B: May increase unnecessary truck utilization C: May cause extra purchase of trucks	3	7	2	N/A	42	Relatively High	Based on the order flow, adjust truck capacity or purchase new trucks.
	Amount of Loading Trucks	More than needed	Insufficient understanding of order flow	May ncrease labor cost, maintenance cost, and purchase fee of trucks.	2	5	2	Daily inspection	20	Low	Based on the order flow, adjust amount of trucks.
		Less than needed	Insufficient understanding of order flow	A: May delay the orders. B: May increase unnecessary truck utilization C: May cause extra purchase of trucks	3	7	2	Daily inspection	42	Relatively High	Based on the order flow, adjust amount of trucks.
	Maintenance of Loading Trucks	Damages during the delivery	Fail to maintain regularly	A: May cause safety problem during the delivery B: May cause longer delivery time, hence may delay the orders C: May shorten the trucks' lifetime	3	7	3	Regular maintenance	63	High	Follow the regular maintenance. Also, implement monitering solution.

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severerity	Detection	Current Control	RPN	Action Priorty	Suggested Action
Spaces	Main Storage Area	More than needed	Insufficient understanding of the order flow, and need for inventory.	A: May cause unnecessary waste of space B: May cause unnecessary management cost C: May cause over-inventory	2	6	2	Visual inspection	24	Low	Analyze the order flow and adjust the space based on it.
		Less than needed	Insufficient understanding of the order flow, and need for inventory.	A: May cause low inventory, hence may delay the orders. B: May cause extra delivery fee (from manufacturer)	3	7	3	Visual inspection	63	High	Analyze the order flow and adjust the space based on it.
	Standby Storage	More than needed	A: Insufficient understanding of the order flow B: Insufficient space for Main Storage	A: May cause unnecessary waste of space B: May increase unnecessary management cost	2	5	2	Visual inspection	20	Low	Analyze the order flow, and current main storage. Then adjust the space based on it.
		Less than needed	A: Insufficient understanding of the order flow B: More space for Main Storage than needed.	A: May cause longer loading time, hence may delay the orders.	4	6	2	Visual inspection	48	Relatively High	Analyze the order flow, and current main storage. Then adjust the space based on it.
	Space for Forklift Transportation	More than needed	Insufficient understanding of forklift transportation, and utilization.	A: May cause unnecessary waste of space. B: May increase unnecessary management cost. C: May decrease spaces for storage.	3	5	3	Visual inspection	45	Relatively High	Analyze the daily forklift utilization.Then adjust the space based on it.
		Less than needed	Insufficient understanding of forklift transportation, and utilization.	A: May cause safety problem during the operation. B: May cause longer operation time, hence may delay the orders.	3	7	3	Visual inspection	63	High	Analyze the daily forklift utilization.Then adjust the space based on it.

Category	Sub-category	Possible Failure	Cause of the failure	Effects of the failure	Occurance	Severerity	Detection	Current Control	RPN	Action Priorty	Suggested Action
All PalletRacks	Capacity of PalletRacks	More than needed	A: Insufficient understanding of order flow B: Insufficient understanding of need for inventory	A: May cause unnecessary waste of space. B: May increase unnecessary management cost.	2	5	3	Visual inspection	30	Low	Analyze the order flow, and need for inventory. Then adjust the capacity of PalletRacks.
		Less than needed	A: Insufficient understanding of order flow B: Insufficient understanding of need for inventory	A: May cause low inventory, hence may delay the orders. B: May cause longer unloading and loading time.	2	7	3	Visual inspection	42	Relatively High	Analyze the order flow, and need for inventory. Then adjust the capacity of PalletRacks.
Human Resources	Employees in Distribution Center	Insufficient understanding of all relative operation processes	Lack of training	A: May cause safety problem during the operation. B: May delay the orders. C: May decrease work efficiency.	4	8	3	N/A	96	Urgent	Arrange regular training for employees and managers. Also, issue operation handbook if possible.

DECISION SCIENCES INSTITUTE

Highlighting the Challenges in Analytics Practice of a Non-Profit Organization

Alex Maharjan
Illinois State University
Email: amahar1@ilstu.edu

Dr. Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

Non-profits have wealth of data but certain unique situations are hindering them to utilize their data effectively. Most non-profits want to take steps to leverage their data and generate insights but they face challenges in their adoption efforts and are at risk of losing the analytics initiatives. This study is based on the analytics project experience the researchers had to help a local non-profit organization to generate insights from their data. Major learning from the study is that the non-profit has the inherent data deficiencies and the lack of active culture of analytics use, which negatively affects all downstream activities in the analytics work flow. It is recommended that substantial time and resources should be invested in the early stages of transitioning to the analytics practice.

KEYWORDS: Non-Profit Organizations, Data Analytics, CRISP-DM Model, Analytics Challeges, Analytics Workflow

BACKGROUND: CURRENT STATE OF ANALYTICS USE IN NON-PROFITS

Organizations have been leveraging big data to improve business results. A strong data-driven decision-making approach has led to better output and productivity (Giacomo & Breman, 2016). According to the research conducted by Tambe and Saraf (2014), organizations, with significant data and competing in labor markets, demonstrated higher productivity than an average organization by employing big data technologies. Even non-profit organizations that have employed analytics-driven decision making have reported higher effectiveness in assessing social impact and driving action to improve productivity (Giacomo & Breman, 2016). By employing big data analytics to measure the economic impact of the performance, non-profit organizations have presented credible analyses, which has been positively received by the partners and helped them in funding resources (Pease & Beresford, 2013). According to Wallace (2019), artificial intelligence has helped non-profit organizations to tailor their fundraising campaigns and large non-profit organizations have been using analytics to identify promising potential donors for a long time now.

Big non-profit organizations such as United Nations have employed big data in projects focusing on training and development, but there is little evidence of big data analytics

employed in workplace training and performance improvement projects in other non-profit organizations (Giacomo & Breman, 2016). Non-profit organizations such as various government organizations hold a tremendous quantity of client data but have encountered the challenge of effectively managing, analyzing and implementing insights to improve their performance (Wang et al, 2019). Non-profit organizations have been facing an increasing demand to use data and analytics for quantifying social impact, but despite it, many nonprofits are largely behind in employing analytics. There is a little conversation about actual implementation or success stories about the potential use of big data in a non-profit organization, with healthcare-focused organizations being an exception (IBM, 2017).

Non-profit organizations have cited budget, technology, and talent among the primary barriers for using analytics in operations. Leaders in the non-profit organizations admit they are in hindsight stage of the analytics process. Raising funds for internal investments has been a struggle for non-profits. Barriers are common across sectors and the nature of operating in a non-profit industry amplifies the lack of resources for any endeavors (IBM, 2017).

Non-profits have cited technological problems that have made it difficult for non-profits to capture, structure and understand daunting volumes of available data. Non-profits also are challenged by the affordability of talents to employ and implement analytics (IBM, 2017). At present, even though there is pressure to employ analytics, the status of using analytics is at a preliminary stage for most non-profit organizations.

Major donors demand quantified evidence of the social impact of non-profit organizations. There is a chance of donors might look for other investments if non-profit do not employ analytics to quantify their performance. Lagging to implement analytics could threaten future outcomes and impacts of non-profits (Pease & Beresford, 2013).

Thus, the purpose of the study is to explore the state of analytics process in non-profit organizations and the steps a non-profit organization can follow to implement analytics or upgrade their analytics process to the next level.

STUDY GOALS

Current literature shows non-profit organizations are not able to leverage their data and data analytics properly to generate insights. Barriers include lack of budget, talented personnel and technological problems that are hindering non-profits from implementing analytics-based work practice.

In this article, the researchers present a case study of a non-profit organization in adopting analytics in their workflow. The study aims to understand the present conditions of the non-profit organization in using data for the decision-making process. It is based on the interactions and communications with the non-profit's managers and on the experience in understanding and analyzing their data. The study will also cover present analytics process used in the organization and improvements that could be employed.

The current study aims to:

- Research and understand the state of analytics through experience with a non-profit organization; and
- Provide recommendations to help the analytics adoption and practice for non-profit organizations in similar situations.

STUDY METHODOLOGY

The researchers of this study use the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a guideline to initiate and execute the analytics project of a local non-profit organization. CRISP-DM is a non-proprietary, documented and freely available data analytics workflow model developed by industry leaders with input from various data analytics users. The CRISP-DM breaks down the life cycle of a data mining project and presents a comprehensive data mining model for anyone with process for a data mining project. The model encourages best practices and helps organizations realize better, faster results from data mining by providing a proper structure to their infrastructure. This model organizes data mining process into following six phases: business understanding, data understanding, data preparation, modeling, evaluation and deployment.

Business understanding process involves setting business objectives, assessing the situation, setting project objectives from a business perspective and converting the knowledge into a data mining objective. It is vital for practitioners to fully understand the business model for which the objective is being set.

Data understanding process follows the program objectives set through business understanding. It starts with initial data collection, getting familiar with the data, identifying data problems, exploration of data and verifying data quality. The process requires to check if the data collected meets the project requirements or if additional information is required.

Data preparation process covers the activities to construct the final dataset that will be fed into a model. Tasks in this process include selecting the data, cleaning the data and developing new records as per the project needs. Also, this process includes integrating data from different sources and converting them to proper format for analysis.

Modeling process includes selecting modeling technique, generating test design, building a model and assessing the model to judge the success of application. It may include building different models and evaluating against each other to get the optimum results.

Evaluation process involves evaluating the model on different measures, evaluating the results, confirming the model achieves business objectives etc. Based on the evaluation of the model, the decision might be to initiate further iterations or pursue new data mining project.

Deployment stage involves planning, organizing and deploying the model for end user. It involves monitoring, maintaining and updating the model as needed. Also, the process includes documenting and summarizing all the important experience during the project for knowledge transfer in the future.

CASE STUDY

An organization used in this study is a medium size non-profit in education operating in a mid-western region. Its mission is to provide a safe, inclusive, playful and engaging learning environment for their clients. The organization engages the visitors in experiences that evoke imagination and curiosity through its exhibits, classes and programs.

The organization is in its very early stage of using analytics in the organization. Employees have come to become advocates of using analytics in their organization and integrating analytics in their workflow through participation in the non-profits conferences and getting to know use of analytics in other non-profits. They were interested in understanding how they could initiate the process of using analytics in their own work.

They recently outsourced their data management to online solutions to reduce their learning curve. All the previous data from their legacy system was transferred to the new data management solution. At present, the organization has about a couple of years of complete data recorded in their new database system.

The employees are currently in the process of using data to make more informed decisions. The middle-tier managers have been gathering information about data analytics practice and they had also employed a data analytic intern this past year to understand their data and generate insights. Employees routinely use the online platform to gather data for occasional analysis.

The organization is currently wanting to use a data visualization technique to make sense of the data they have but the employees have limited experiences and resources in adopting analytics technology. Along with this new interest and curiosity, the employees recognize the low level of articulation of the data and lack of specific data in the organization to start an analytics initiative. Thus, the organization recognizes challenges and the new data they need to capture to continue to mine insights from the data. The researchers worked with the staff members for several months as external consultants to help the organization implement simple visual analytics in extracting insights from the business data.

LESSONS LEARNED

The study involved a comprehensive analytics project where the researchers helped the non-profit to take off the ground to adopt the analytics in their everyday business operations and to understand how the analytics knowledge transfer process could be accelerated. Training on visual analytics was provided multiple times and the participants learned various easy-to-use but impactful analytics techniques. They realized that the technical barriers in analytics implementation would not be as insurmountable as they thought. Along the way, the analytics consultation service was provided to give them specific business insights through a series of meetings. However, at the end of the project, it was not so much the analytics technical barriers that slowed down their plan of adopting the analytics practice. The lessons learned are summarized below.

Lesson #1: Dirty Data but Who Made it Dirty and Who Can Fix it?

The non-profit implemented a cloud-based database platform in recent years to streamline and consolidate their old fragmented databases. The new database is a software solution designed for a certain type of non-profit organizations such as the institution in this study. The organization achieved its goal of having an efficient and uniform data management system. However, the success of transitioning to the modern data management system led to unexpected impediments in their analytics adoption effort.

Firstly, the new database's highly specialized features allowed the non-profit to conduct only the usual query operations for simple reporting purposes. This conversely means that it was not designed for people who want to extract data to further process and analyze them in depth. For example, the features to download the data from the system were developed too narrow with limited capabilities and some of the data extraction process was not optimized to support the creation of complete datasets. The system without the schema to database, the interrelationships of the tables and the variables in the database are some of the specific examples that consumed much of the researchers' effort in executing the analytics project in the early phase.

Secondly, the new data management system was implemented by combining multiple legacy databases scattered around the organization. In doing so, the structure and definitions of one database were not maintained when mixed with another database. As a result, by the time the new database was in place in the cloud the data was stored in a state-of-the-art technology but its contents lacked the coherent meanings and consistency. Consequences are, for example, in the new cloud database the researchers encountered duplicate variables, variables with same names but with totally different meanings, undefined variables, database tables with no relationship to others, missing data, and other typical types of dirty data with high frequency. This situation directly would affect the level of data quality, which are vital for any analytics projects to be successful. This issue forced the researchers to confine themselves only to a small subset of data for their analysis.

Lesson #2: Which is First, Chicken or Egg?

For analytics projects to deliver business values they have to start with a burning business question and subsequently provide the solution to that question. However, most of time, it would not be done directly from the question to the solution. Rather, an incremental answer to the initial question triggers other related questions, each of them in turn leads to another sets of incremental answers. These answers, of course, initiate another round of questions to be answered. At the end of this iterative questioning/ answering process, the cumulative answers become a richer and more convincing solution to the original business problem. The driving force of leading and maintaining this process would be the business acumen of the parties involved in the analytics project, along with the technical expertise in using analytics. However, in the current analytics-hyped environment, non-technical users of analytics expect a break-through analytics solutions to their business questions. Certainly, this was the case with the non-profit organization in this study. Due to their limited capabilities with analytics, it was not

surprising that the managers tried to share their business information related to the data to the researchers in one meeting and would expect the insights at another meeting, not realizing their contribution would be as much important as the analytics technology power. For example, a few dashboards and storyboards were created from the dataset in order to give the mangers a set of views on how the business was doing, which allowed them to compare performances of various departments. It helped them to view more relationships between the visual solutions and they were able to generate more related questions. The problem, however, was that the newly generated thoughts and ideas were not actively pursued by those business experts to move on to the next round of finding another insights. In other words, the researchers experienced the divide line where a client provides a business question and the consultant provide solution through a formal communication channel. In that relationship, one party does not have what the other party possesses. Unless they work together as a coherent team or one party has both capabilities, it would not be possible for their analytics projects to gain synergistic iterative process to find incremental solutions.

IMPLICATIONS

In a sense, non-profit organizations face a more uphill battle the for-profit organizations in terms of being able to take advantage of the analytics technology. Most of the non-profits already operate their businesses with tight budgets and high expectations from stakeholders. Even when non-profits are trying to be productive by updating their technology such as migrating their databases to the latest data management system, this move could in fact backfire on them because it could create more headaches in terms of data quality and consistency later. These non-profits certainly have more volumes of data but in fact they have more wastes in the data as well, which creates unexpected inefficiencies and troubles when the analytics practice is adopted. Especially small and medium size non-profits do not have luxury to have dedicated IT staff to respond to the demand to provide data of decent quality.

Additionally, no dedicated resources and leadership team for the adoption of analytics practice could be devastating for non-profits because their existing staff members are not exposed to the analytics-centered organizational shake-up as their for-profit counterparts. If they have to constantly rely on occasional interactions with external consultants, they are likely to continue to be on the path where they are not self-generating the business insights from the data on their own.

CONCLUSION

Non-profits have wealth of data but the certain non-desirable situations such as data deficiencies and lack of active culture of analytics use, are hindering them in utilizing their data effectively. Non-profits are slowly taking steps in adopting analytics in their business decision by leveraging the value hidden in the data into the business insights. However, initiating the analytics adoption process without realizing the inherent hindrances could expose them to the high risk of potential failure with their analytics initiatives. After all, even well-funded for-profit organizations experience more analytics adoption failure than success. It is critically important for non-profits who consider using analytics actively to support business decision to carefully plan the transition and secure

buy-ins in and out of the organizations to reduce the potential setbacks in their attempt to be a beneficiary of the analytics practice.

REFERENCES

Giacumo, L. A., & Bremen, J. (2016). Emerging evidence on the use of big data and analytics in workplace learning: a systematic literature review. *Quarterly Review of Distance Education*, *17*(4), 21.

Leap before you lag. (2017). *IBM Institute for Business Value*. Retrieved from https://www.ibm.com/downloads/cas/Y27MY7RN

Pease, G., & Beresford, B. (2013). Big data analysis drives not-for-profit performance. *T+ D Magazine*.

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, *5*(4), 13-22.

Tambe, H., & Saraf, O. (2014, October). How analytics helps companies improve talent ROI. *Baseline,* 1.

Wang, C. S., Lin, S. L., Chou, T. H., & Li, B. Y. (2019). An integrated data analytics process to optimize data governance of non-profit organization. *Computers in Human Behavior*, *101*, 495-505.

WALLACE, N. (2019). A.I. Is Ready for Fundraising: Are Fundraisers Ready for A .I.? *Chronicle of Philanthropy*, *31*(5), 8–14. Retrieved from https://search-ebscohost-com.libproxy.lib.ilstu.edu/login.aspx?direct=true&db=a9h&AN=135196319&site=ehost-live&scope=site

DECISION SCIENCES INSTITUTE

Growth of the Legal Cannabis Industry: Role of Coevolutionary Factors

Shalini Bhawal
University of North Texas
Email: shalini.bhawal@unt.edu

Manjula S. Salimath
University of North Texas
Email: Manjula.salimath@unt.edu

ABSTRACT

The rapidly growing legal Cannabis industry offers strategic opportunity. As firms incorporate cannabis into their products and R&D, implication for business include disruption of traditional industries such as pharmaceutical and consumer goods. Yet little is known about the contributing factors or antecedents to legal cannabis industry growth. To address this gap, we present the role of coevolutionary factors, couched within an external enabler framework. Specifically, propositions that link institutional support, trade associations, and discontinuous change to the rise of the U. S. cannabis industry. The paper contributes to understanding the coevolutionary dynamics of the growth of the legal cannabis industry.

KEYWORDS: cannabis industry, coevolution, institutional support, trade associations, discontinuous change

Full text and references available upon request

DECISION SCIENCES INSTITUTE
Adjusting Forecasts with Seasonal Data

Rick Hesse, Pepperdine University, Union College (KY)
Email: rickhesse@aol.com

ABSTRACT

Three Excel templates are shown using seasonal data (20 quarters). The first performs a Quick & Dirty check for seasonality. The second uses Solver Moving Averages (SMA) to determine optimal indices for each quarter and optimal parameter(s) for 2-period and 3-period SMA and forecasts for period 21. The third template uses the data from the first template and Regression quartic forecasting, re-seasonalizing the forecast. The last two templates minimize RMSE error but also show MAD and MAPE. In this case, SMAs are superior to Regression and there is no need to minimize MAD or MAPE to substantially improve true r^2.

KEYWORDS: Seasonality, Time Series Data, Solver Moving Averages, Regression, True r^2

INTRODUCTION

Since the introduction of Solver into spreadsheets over 30 years ago, the ability to identify and adjust seasonal data has allowed the development of optimizing algorithms to minimize forecasting error (RMSE, MAD or MAPE) in both Solver Moving Averages (SMA) and Regression formulas. First an Excel template is used to identify seasonality of an actual data set. A second Excel template (SMA) uses this seasonal data and Solver to determine optimal weighted moving average and seasonal indices. The third Excel template (REGRESSION) uses the seasonally adjusted data from the first template. This data exhibits a quartic curve which looks like a 4^{th} order polynomial for regression but has much worse error than using SMAs.

SEASONGRAPHS

This first Excel template, SEASONGRAPHS, is an attempt to identify seasonality if it exists. Seasonal indices are computed and seasonally adjusted (de-seasonalized) data generated for use with Solver Moving Averages (SMA). Later it will be used for Regression. This first template does not require the Solver.

The Data

The Foodservice Company, which feeds fire fighters and other emergency workers in California, is shown in **Figure 1** (Google Image). The company at that time was the 5th largest privately owned broadline food distributor on the West coast. Since the company is privately held, its accounting Fiscal Year starts in July (Quarter 1 is July 1 to September 30). Approximate sales figures are given in $000s for five years or 20 quarters.

Figure 1. California Firefighters Taking a Break

The time series data and graph are shown in **Figure 2** by each quarter for 5 years with Quarter 4 and 1 being the highest of the four, and Quarter 3 always the lowest and Quarter 2 always the second lowest.

Seasonality Numerical Rule of Thumb

Quick & Dirty seasonal indices in cells **D4:D7** in **Figure 2** are computed by taking the average sales in each row in column **K** and dividing by the grand average in cell **K8**. These will add up to 400%. A message in cells **A8:B8** is shown if the difference between the highest and lowest index is $\geq$ 30% that there is seasonality, and there is no seasonality if the difference (Δ) is $\leq$ 15%. Otherwise the message is that "there may be seasonality". This is just a quick and dirty "rule of thumb".

Graphical Seasonality Indicators

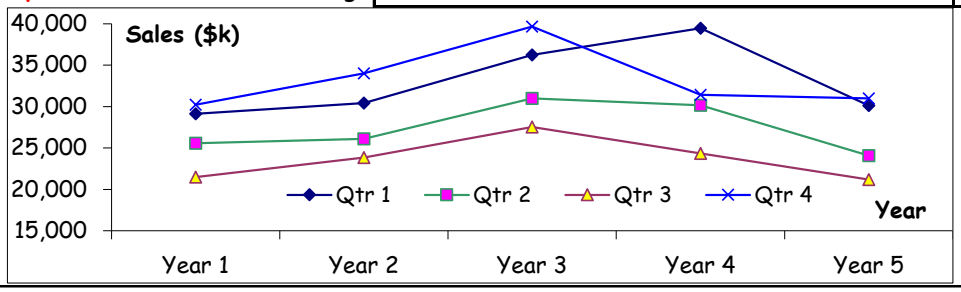
The line graphs of the sales for each quarter are also shown in **Figure 2**, and we can see that the colored lines for all 4 quarters follow a similar pattern, with Quarter 3's line always the lowest revenue and Quarter 2's always the second lowest. Quarter 4's line is the highest except at Year 4, so there is some consistent separation of these lines which confirms our numerical analysis. Both numerical and graphical analysis allow for interpretation. This is both an art and science. It helps to be familiar with what the numbers represent.

Figure 2. ***Foodservice Company Sales for 5 Years***

SEASONGRAPHS

Foodservice Co. Sales ($k)

				Raw Indices	Sales ($k)	Year 1	Year 2	Year 3	Year 4	Year 5	Average
Max	113.33%	1		112.68%	Qtr 1	29,120.6	30,443.9	36,230.1	39,462.8	30,129.8	33,077.4
		2		93.32%	Qtr 2	25,581.9	26,123.8	31,004.7	30,172.0	24,080.3	27,392.5
Min	80.67%	3		80.67%	Qtr 3	21,487.9	23,832.5	27,544.4	24,349.6	21,187.9	23,680.4
Δ	32.66%	4		113.33%	Qtr 4	30,238.9	34,018.8	39,658.6	31,420.4	30,996.5	33,266.6
Seasonality exists					Average	26,607.3	28,604.8	33,609.5	31,351.2	26,598.6	29,354.3



False Signals of Seasonality

Consider the data and graph shown in **Figure 3** of revenue for a well-known dandruff shampoo which graphically showed seasonality for 2001-2003 but by 2004 and 2005 almost all graphical signs of seasonality had disappeared because by 2005 there was almost no separation between quarterly revenues. Yet the numerical Δ between highest and lowest overall indices

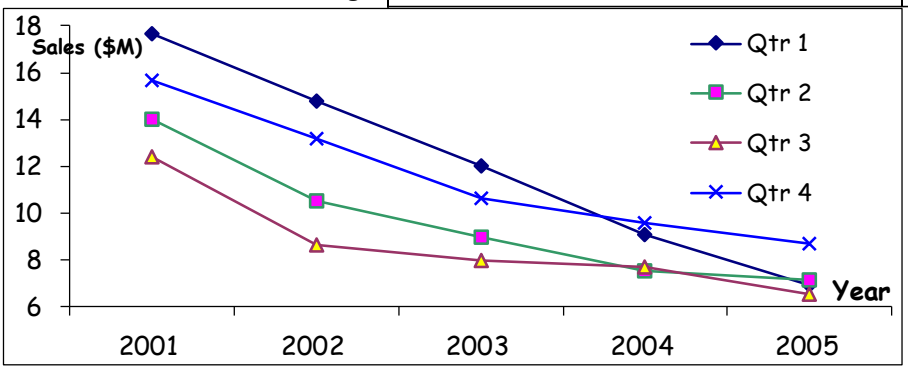
showed 32.99%, slightly higher than the data showed for the Food Service Company revenue. Clearly shampoo product sales were declining over the years and sales were flat lining. Incidentally, the company reformulated their product and sales bounced back as well as seasonality. This was thinly disguised data from a working MBA student at this company who was my Quantitative student at Pepperdine at least 10 years ago.

Figure 3. Data with False Numerical Seasonality Signal

SEASONGRAPHS

Dandruff Shampoo Sales

			Raw Indices	Sales ($M)	2001	2002	2003	2004	2005	Average
Max	115.42%	1	115.42%	Qtr 1	17.7	14.8	12.0	9.1	6.9	12.1
		2	91.84%	Qtr 2	14.0	10.5	9.0	7.5	7.1	9.6
Min	82.43%	3	82.43%	Qtr 3	12.4	8.6	8.0	7.7	6.5	8.6
Δ	32.99%	4	110.30%	Qtr 4	15.7	13.2	10.6	9.6	8.7	11.6
Seasonality exists				Average	14.9	11.8	9.9	8.5	7.3	10.5



One method to deal with truly seasonal data for forecasting for the next quarter (21) has been illustrated previously (Hesse, R, 2018). It used a Solver Moving Average (SMA) template to forecast with the de-seasonalized data, but for this paper, after determining that there is seasonality, we will use three models with two other Excel templates:

- **SEASONAL**: 2-period Solver Moving Averages (**SMA2**) with Indices determined by Solver simultaneously,
- **SEASONAL**: 3-period Solver Moving Averages (**SMA3**) with Indices determined by Solver simultaneously,
- **Regression5**, using the de-seasonalized data to minimize RMSE (Root Mean Squared Error) to fit a quartic, (4th order with 5 parameters) equation.

Then we can determine which method gives the least RMSE error for this data set. We will also track MAD and MAPE error to see which of the three error measures should be optimized.

SEASONAL TEMPLATE

The **SEASONAL** template in **Figure 4** allows the raw data to be input into columns **A** and **B** starting after the headers in row **13**. The template is set up to compute the information about the

Five basic naïve time series models in rows 6→10 (which stays fixed for each data set) and the SMA for a 2-period model (SMA2) in row 5 minimizing RMSE in cell **D5**. The light-yellow colored cells are the Solver variables, with the quarterly indices in **I6:I9** and the parameters for the SMA models in **B11:C11**. In this case, the template is setup for the SMA2 and needs just cell **B11 (P_1,** the % of the change in the last two data values to be added to the last data value**)** to be adjusted along with the four indices. The quarterly indices are set at 100% (no seasonality) rather than use the values computed in **Figure 2**, although they could be used as initial values. The end results are the same and the template seems to be robust regardless of initial values, except P_1 cannot be 0. Also displayed in cells **I11:J12** are the weights on the last two data points derived from the P_1 value. SA Forecast stands for Seasonally Adjusted Forecast in column C and H. Δ is the difference between highest and lowest index and is now 0%. See Appendix for explanation of switching from W_i weights to P_i proportions for SMA3 and SMA3.

Figure 4. Setup for SEASONAL 2-period SMA2 Template

	A	B	C	D	E	F	G	H	I	J
1	Quarterly Seasonal 2-period Solver Moving Average (SMA2) Model									
2		Foodservice Co. Sales ($k)								
3		Minimize	Reseasonalized	Error Measures						
4		R.M.S.E.	Forecast	R.M.S.E.	M.A.D.	M.A.P.E.	df	Seasonality doesn't exist		
5	2-Per Solver Model		26,092.2	6,826.67	6,435.30	22.22%	18	Quarter	Index	
6	N Total Average		29,354.3	5,282.77	4,303.12	15.19%	19	1	100.00%	
7	A Cum Average		29,354.3	5,621.04	4,585.25	15.78%	19	2	100.00%	
8	I 3-Per Smooth		25,421.5	6,079.12	5,265.30	18.11%	17	3	100.00%	
9	V 2-Per Smooth		26,092.2	6,826.67	6,435.30	22.22%	18	4	100.00%	
10	E Next = Last		30,996.5	6,230.63	5,200.98	17.74%	19	Δ = 0.00%	400.0%	
11		-50.00%		-67.0%	-123.7%	-114.1%	r^2		50.0%	W_{-1}
12	n = 20	P_1	P_2		$F_3 = (W_{-1}*Y_2 + W_{-2}*Y_1)$				50.0%	W_{-2}
13	Quarter	Sales	Adjusted	SA Forecast	Quarter	Sales	Adjusted	SA Forecast		
14	1	29,120.6	29,120.6		18	24,080.3	24,080.3	30,775.1		
15	2	25,581.9	25,581.9		19	21,187.9	21,187.9	27,105.0		
16	3	21,487.9	21,487.9	27,351.2	20	30,996.5	30,996.5	22,634.1		
17	4	30,238.9	30,238.9	23,534.9	21	26,092.2		26,092.2		

Because all quarterly indices are 100% in **Figure 4**, initially the original Sales (columns **B**, **F**) are the same as the Adjusted sales (columns **C**, **G**). Notice that because of the great variation in the original data, the naïve Total Average gives the lowest r^2 error measure for all three types of errors (pink shading indicates this) and lowest of the five types of naïve forecasts (which don't need the Solver and stay the same for each data set). The last 3 data points, periods 18-20, are brought over from the bottom of columns **A**, **B**, **C** to also be in cells **E14:G16**. The SA Forecast for Quarter 21 in cell **H17** is then re-seasonalized in cell **F17** and would be $26,092,200.

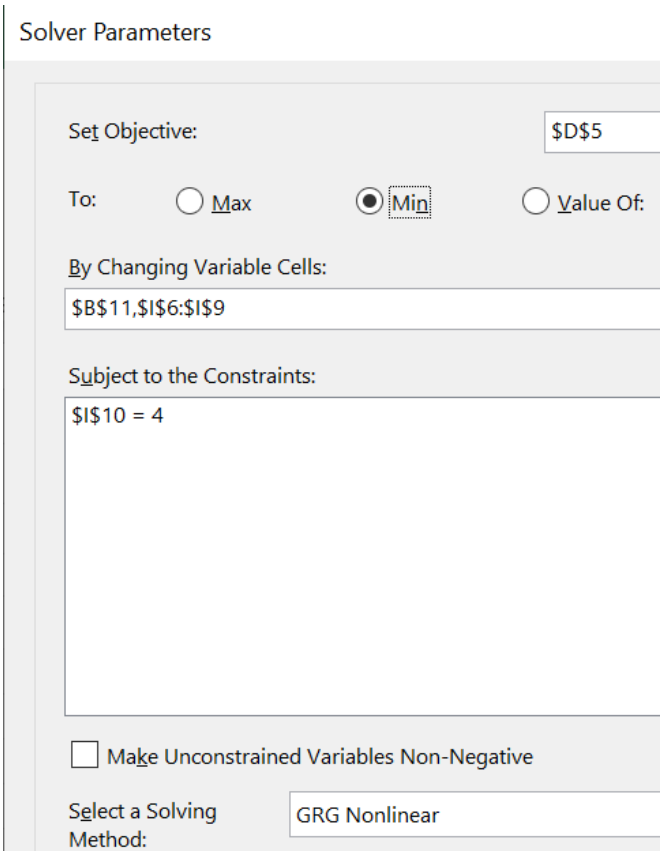
The starting value for cell **B11** is -50%, which takes half the distance between the last 2 data points and adds to the last, or 2-period Smoothing (of 50% of the last 2 periods). Notice that if we convert P_1 to W_{-1} and W_{-2} for weighting the last two data values, they would be 50% each as shown in cells **I11:I12**. All of the r^2 values (1-[Model error/best Naïve model error]^2) are negative! This means the 2-period SMA will not work without using seasonally adjusted data. One last thing to note is that cell **H11** is the difference between the largest and smallest index

and is 0.00% because all indices have been set to 100%. Because cell **H11** is ≤ 15%, the message "**Seasonality doesn't exist**" appears in cells **H4:I4**.
The best naïve model with this initial Solver setup is Y-bar, or the overall average because of the huge amplitude of the sales. The best naïve model is shaded in pink for all 3 error measures.

Solver

Figure 5 shows the Solver Setup. This initial run of the template will just use a 2-period SMA, which only needs P_1% to be optimized along with the 4 quarterly indices, which must add up to 4.00 or 100% times 4 quarters. If an index is >100%, that quarter has more than average quarterly revenue. Because cell **B4** points to cell **D4** (RMSE), the template will minimize RMSE. Note that the "old style" SMA shows what the weights for last period (W_{-1}) and two periods ago (W_{-2}) are 50% each. But the P_1 value tells us we could also just take 50% of the difference in the last two data points and add to the last one. Thus our SMA forecast will be a modification or "tweaking" of Next=Last.

Figure 5. Solver Setup for 2-period SMA



The Objective is set to cell $D$5 or the RMSE error and is to be **Min**imized. The Changing Variables are $B$11 (P1%, the proportion to be added the Next=Last of the difference in the previous two period's de-seasonalized sales) and the other variables in the Changing Variable Cells are $I$6:$I$9, or the four quarterly indices.

There is one constraint in cell $I$10 which forces the indices to add up to 400% or 4.0 for the four indices.

Note that all the variables are allowed to be negative, (but the indices never would be). However, the P_1% certainly can be negative to pick up the trend. In **Figure 4** you will note that P_1% was set to -50%, which computes simple 2-period smoothing or takes half the difference in the two periods to add it to the last period.

The Solver Method is GRG Nonlinear (Gradient Search) because Simplex cannot be used with a nonlinear model. In fact, regardless of which error measure we choose to minimize, each objective function is highly nonlinear because MAD and MAPE use the Absolute Value function, which is why these last two objective functions could not be minimized before computers were invented because calculus cannot compute derivatives algebraically with the absolute (**ABS**) function.

2-PERIOD SMA (SMA2) RESULTS

Figure 6 shows the results of the template optimization. Cell **B11** has changed from -50% to 66.50%, meaning that using the de-seasonalized data in column C, the next period's forecast is taking the previous period and adding 66.50% of the difference in the last two periods. The r^2 values for all three error measures are about 50%, which means a very good fit for any of the error measures (30% or higher is considered acceptable). Note that W_{-1} and W_{-2} have also changed. These true r^2 values are computed by taking model error and comparing it with the best naïve error, not always Total Average. (Hesse, R, 2018). Since RMSE was chosen to be minimized, if we changed to minimize MAD or MAPE, those r^2 values would increase and the current RMSE r^2 of 49.5% would get a bit smaller. This also means it might be worthwhile minimizing either MAPE or MAD in our investigation, but we will hold off on that for a while until we run the 3-period SMA later in this paper and determine if that would be necessary.

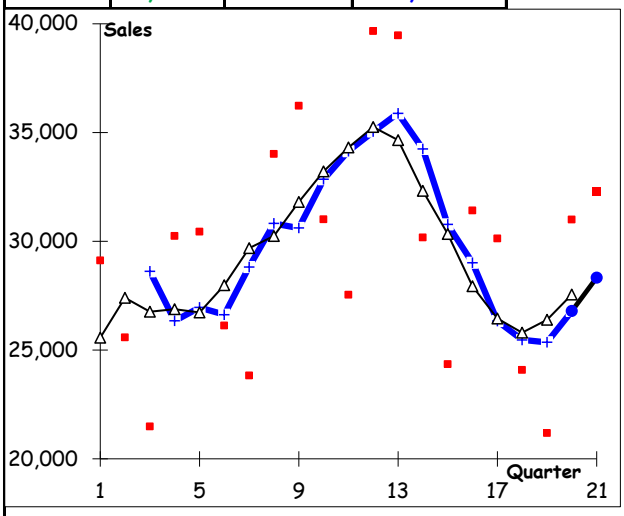
This solution now shows seasonal indices very similar to **Figure 2** and columns **B** and **C** now differ and the values in column **C** are the original data divided by the appropriate seasonal index. Thus, for X = 5, 9, 13, and 17, the values in column **C** are the value in column **B** divided by 113.90%. The Solver works by adjusting each of the 4 or 5 Solver variables backwards and forwards to determine which direction each one should be adjusted using a 5-dimensional search. When the objective function can no longer be reduced, the Solver quickly arrives at an optimal minimum. Column **C** computes the SA (Seasonally Adjusted) forecast by taking the difference between the last 2 column C data points, multiplying it by 66.60% and adding it to the last Adjusted data point. The thick blue line on the graph in **Figure 6** shows the seasonally adjusted forecasts (SA Forecast) and cell **F17** in green is the re-seasonalized forecast.

Looking at other results of this 2-period model, cell **H4:I4** now states "**Seasonality exists**" and that is because cell **H11** is now $\geq$ 30% difference between highest index (113.90%) and lowest (80.27%). We now see that quarters 1 and 4 are higher than average, and quarters 2 and 3 are lower than average. Since the fiscal year is July 1 to June 30 next year, quarters 1 (Summer) and 4 (Spring), are fire and then flood seasons for California. Now the Sales are different than the Adjusted Sales on the template and the color code matches the graph, shown on the graph in **Figure 5**. Incidentally if Δ is between 15% and 30%, the message reads "**Seasonality may exist**". Now W_{-1} = 166.5% and W_{-2} = -66.5% for the weights of the last two data points and these weights always add up to 100% and allowed to be negative. This is what changes moving averages from "smoothing" to "Solver" and optimal.

The red squares on the graph in **Figure 6** represent actual sales in thousands of dollars, while triangles connected by the thin black line represent the seasonally adjusted sales. The vertical spread of the red squares indicates the seasonality, while the shape of the black thin line indicates a "W" shape or a 4^{th} order polynomial (quartic). The blue + signs connected by the wider blue line indicates the seasonalized forecasts (and the last two large blue dots). There is no + or blue line for periods 1 and 2, because we can't forecast until we have 2 real data points. The seasonally adjusted blue dot at the end represents actual data for Quarter 20 and the forecast for Quarter 21 (found in cells **H3:H4**). The seasonal forecast for Quarter 21 is the red dot at the far edge of the graph in the middle and in cells **C5**, and **F17**, sales of $32,264.8k. This is simply SA Forecast for Period 21 in cell **H17** (28,328.4) multiplied by $Index_1$ of 113.90%.

Figure 6. Optimal Solution for 2-Period SMA2

	A	B	C	D	E	F	G	H	I	J
1	Quarterly Seasonal 2-period Solver Moving Average (SMA2) Model									
2		Foodservice Co. Sales ($k)								
3		Minimize	Reseasonalized	Error Measures						
4		R.M.S.E.	Forecast	R.M.S.E.	M.A.D.	M.A.P.E.	df	Seasonality doesn't exist		
5	2-Per Solver Model		32,264.8	963.78	793.78	2.70%	18	Quarter	Index	
6	N Total Average		33,445.8	3,194.04	2,899.42	9.76%	19	1	113.90%	
7	A Cum Average		33,445.8	3,353.41	2,794.88	9.16%	19	2	93.35%	
8	I 3-Per Smooth		30,276.3	2,410.01	2,066.46	6.94%	17	3	80.27%	
9	V 2-Per Smooth		30,724.3	1,860.87	1,552.72	5.21%	18	4	112.48%	
10	E Next = Last		31,385.5	1,356.37	1,183.60	3.99%	19	Δ = 33.62%	400.0%	
11		66.50%		49.5%	55.0%	54.2%	r^2		166.5%	W_{-1}
12	n = 20	P_1	P_2		$F_3 = (W_{-1}*Y_2 + W_{-2}*Y_1)$				-66.5%	W_{-2}
13	Quarter	Sales	Adjusted	SA Forecast	Quarter	Sales	Adjusted	SA Forecast		
14	1	29,120.6	25,567.8		18	24,080.3	25,795.9	25,470.1		
15	2	25,581.9	27,404.5		19	21,187.9	26,395.3	25,358.3		
16	3	21,487.9	26,769.1	28,625.9	20	30,996.5	27,556.4	26,793.9		
17	4	30,238.9	26,882.9	26,346.6	21	32,264.8		28,328.4		
18	5	30,443.9	26,729.7	26,958.5						
19	6	26,123.8	27,985.0	26,627.8						
20	7	23,832.5	29,689.9	28,819.8						
21	8	34,018.8	30,243.3	30,823.7						
22	9	36,230.1	31,810.0	30,611.3						
23	10	31,004.7	33,213.7	32,851.8						
24	11	27,544.4	34,314.1	34,147.1						
25	12	39,658.6	35,257.2	35,045.9						
26	13	39,462.8	34,648.2	35,884.3						
27	14	30,172.0	32,321.6	34,243.3						
28	15	24,349.6	30,334.1	30,774.4						
29	16	31,420.4	27,933.2	29,012.5						
30	17	30,129.8	26,453.9	26,336.7						
31	18	24,080.3	25,795.9	25,470.1						
32	19	21,187.9	26,395.3	25,358.3						
33	20	30,996.5	27,556.4	26,793.9						



3-PERIOD SMA RESULTS

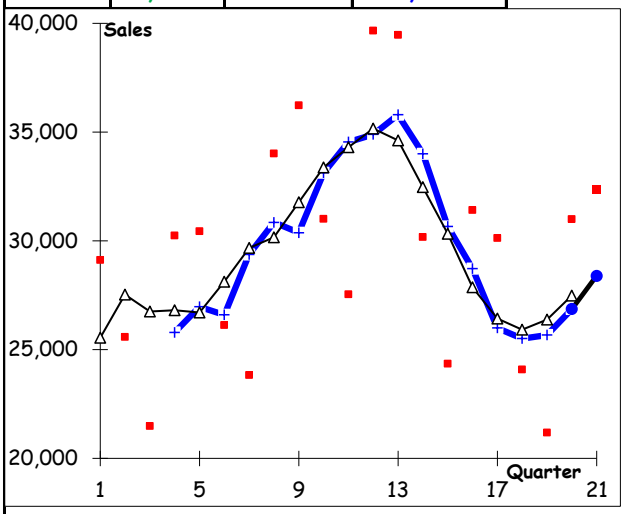
Next the template is Moved and Copied, and the Solver changed to include also changing cell **C11**, the value of P_2%, for a 3-period SMA. P_2 is the percent of the difference between the 3rd and 2nd last data points to be added to the last data point. The Solver is executed, and results are shown in **Figure 7**. The true r^2 values are all improved by an over 10% increase in values and nearly identical in cells **D11:F11**. Since these values are all so close, it makes no sense to run models minimizing MAD or MAPE. This is not always the case, but since minimizing RMSE is the most stable of the error measures to minimize, RMSE should always be minimized initially. See Appendix for explanation of switching from W_i weights to P_i proportions for SMA3

and SMA3. The forecast for actual sales is again found in cells **C5** and **F17** as a result of multiplying cell **H17** by Index$_1$ of 114.00%.

Another thing to note is that the two SMA models have almost identical indices for seasonality, even when initializing both templates with 100% indices, and using the original or successive starting points for P_1. This gives us confidence that we have been able to de-seasonalize the data to be used with the upcoming **REGRESSION** template. This is a very robust solution regardless of the starting Solver "Changing Variable Cells".

Figure 7. Results for 3-period SMA3

	A	B	C	D	E	F	G	H	I	J
1	Quarterly Seasonal 3-period Solver Moving Average (SMA3) Model									
2		Foodservice Co. Sales ($k)								
3		Minimize	Reseasonalized	Error Measures						
4		R.M.S.E.	Forecast	R.M.S.E.	M.A.D.	M.A.P.E.	df	Seasonality doesn't exist		
5	3-Per Solver Model		32,367.0	840.19	708.97	2.39%	17	Quarter	Index	
6	N Total Average		33,484.5	3,194.14	2,898.03	9.76%	19	1	114.01%	
7	A Cum Average		33,484.5	3,355.51	2,801.76	9.18%	19	2	92.91%	
8	I 3-Per Smooth		30,320.8	2,404.53	2,046.37	6.87%	17	3	80.31%	
9	V 2-Per Smooth		30,707.1	1,854.06	1,531.48	5.14%	18	4	112.78%	
10	E Next = Last		31,335.0	1,363.46	1,169.07	3.95%	19	Δ = 33.70%	400.0%	
11		88.11%	-14.05%	62.0%	63.2%	63.2%	r^2		188.1%	W_{-1}
12	n = 20	P_1	P_2	$F_4 = W_{-3}*Y_1 + W_{-2}*Y_2 + W_{-1}*Y_3$					-102.2%	W_{-2}
13	Quarter	Sales	Adjusted	SA Forecast	Quarter	Sales	Adjusted	SA Forecast	14.1%	W_{-3}
14	1	29,120.6	25,542.7		18	24,080.3	25,917.9	25,510.9		
15	2	25,581.9	27,534.1		19	21,187.9	26,383.5	25,669.8		
16	3	21,487.9	26,757.2		20	30,996.5	27,485.1	26,865.5		
17	4	30,238.9	26,813.3	25792.7	21	32,367.0		28,390.3		
18	5	30,443.9	26,703.5	26972.0						
19	6	26,123.8	28,117.4	26598.8						
20	7	23,832.5	29,676.7	29378.6						
21	8	34,018.8	30,165.1	30851.9						
22	9	36,230.1	31,778.8	30376.3						
23	10	31,004.7	33,370.8	33132.0						
24	11	27,544.4	34,298.8	34546.7						
25	12	39,658.6	35,166.0	34892.8						
26	13	39,462.8	34,614.2	35799.7						
27	14	30,172.0	32,474.4	34006.2						
28	15	24,349.6	30,320.6	30666.5						
29	16	31,420.4	27,861.0	28723.5						
30	17	30,129.8	26,428.0	25996.5						
31	18	24,080.3	25,917.9	25510.9						
32	19	21,187.9	26,383.5	25669.8						
33	20	30,996.5	27,485.1	26865.5						



The forecasted value in cell **F17** for Quarter 21 shows a revenue forecast of $32,367k, an increase by just about 100k from the 2-period SMA forecast, which is a small percentage

increase. However, the 3-period SMA does show stronger and consistently better error measures for RMSE, MAD and MAPE. Since r^2 values for all three error measures are virtually identical, there is no need to optimize MAD or MAPE. Note that W_{-2} is negative, which again allows the SMA 3-period model to forecast more effectively than 2-period or 3-period smoothing or even 1-period (Next=Last), which is the best naïve model. In fact, SMAs are usually an improvement on Next=Last.

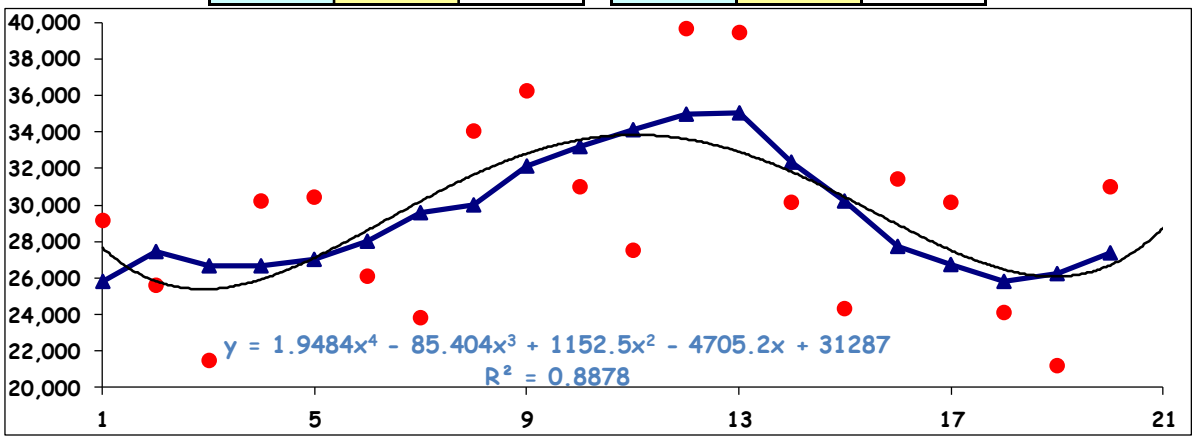
Another option for using these templates in **Figure 6** and **Figure 7**, is to copy the indices in **Figure 2** into the Seasonal Template and not allow cells **I6:I9** to be used in the Solver, but just run the Solver with allowing only cell **B11** or **B11:C11** to change.

LINEAR REGRESSION

Since the pattern of the seasonally adjusted data looks like a "W", the template **LINREG5** will fit a quartic or 4th order polynomial with five parameters (a, b, c, d and e). We first need to get the data set from **SEASONGRAPHS** used in **Figure 2** and shown now in **Figure 8**. The adjusted data is just the original sales for each quarter divided by the respective quarterly index.

Figure 8. Raw Data (Red) and Adjusted Data (Blue) in Column F of SEASONGRAPHS

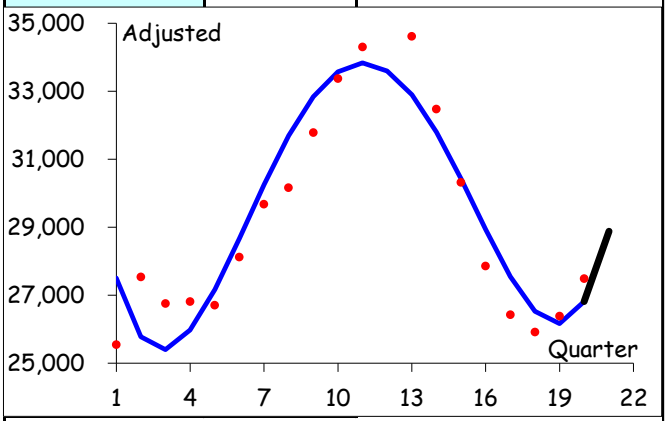
X	Y	Y'	X	Y	Y'
Quarter	Sales($k)	Adjusted	Quarter	Sales($k)	Adjusted
1	29,121	25,843	11	27,544	34,144
2	25,582	27,414	12	39,659	34,995
3	21,488	26,636	13	39,463	35,021
4	30,239	26,683	14	30,172	32,333
5	30,444	27,017	15	24,350	30,184
6	26,124	27,995	16	31,420	27,725
7	23,832	29,543	17	30,130	26,738
8	34,019	30,018	18	24,080	25,805
9	36,230	32,152	19	21,188	26,264
10	31,005	33,225	20	30,996	27,351



For this paper the Solver was run for each fit starting with linear and then progressing to the quartic which gives a slightly smaller RMSE than either Excel Trendline in Figures 8 and 11. The results for the Solver Regression are shown in **Figure 9**, but the biased and unbiased r^2 values in cells **B7:D7, G6** are only computed comparing with Y-bar (Total Average) and instead of Next=Last**.** We could also use the Trendline graph option in Excel to determine the coefficients (see also **Figure 11** later) and then typing in the five parameter values to be able to determine the values of RMSE, MAD and MAPE. However, this gives a larger value of RMSE 1,247.20 versus 1,245.47 shown in cell **B5** in **Figure 9.**

Figure 9. Quartic Regression Solution

	A	B	C	D	E	F	G
1	Regression Model F = a + bX + cX² + dX³ + eX⁴						
2	Foodservice Co. Sales ($k)						
3	Minimize	Regression Forecast Measures			Degrees of		
4	R.M.S.E.	R.M.S.E.	M.A.D.	M.A.P.E.	Freedom		R.M.S.E.
5	Model	1,245.47	1,236.82	4.26%	15	Y-Bar (y)	Biased r²
6	Y-Bar (y)	3,194.14	2,898.03	9.76%	19	29,370.44	88.00%
7	Unbiased r²	84.80%	81.79%	80.91%	n = 20		N=L RMSE
8	F =	a +	bX +	cX² +	dX³	eX⁴	1,363.46
9	Parameters	31,021.404	-4,564.3998	1,131.98458	-84.287638	1.928683	True r²
10	X	Y	F	X	F		8.65%
11	Quarter	Adjusted	Model	Quarter	Model		
12	1	25,542.7	27,506.6	20	26815.46	Re-seasonalized	
13	2	27,534.1	25,777.1	21	28878.64	32,923.7	
14	3	26,757.2	25,396.5				
15	4	26,813.3	25,974.9				
16	5	26,703.5	27,168.5				
17	6	28,117.4	28,679.9				
18	7	29,676.7	30,258.0				
19	8	30,165.1	31,697.8				
20	9	31,778.8	32,841.0				
21	10	33,370.8	33,575.1				
22	11	34,298.8	33,834.1				
23	12	35,166.0	33,598.5				
24	13	34,614.2	32,894.8				
25	14	32,474.4	31,795.8				



These inflated R^2 values are the result of comparing model error with the squared error of the average of the data, as shown in cell **B6** of 31,194.14, and for cell **B7** the r^2 is calculated as

$$\text{Unbiased } R^2 = 1-(1{,}245.47/3{,}194.14)^2 = 84.80\% \quad (1)$$

and is very misleading.

But the actual best naïve model has a squared error of 1,363.46 (**D10** in **Figure 6** and **Figure 7**, entered manually in cell **G8** in **Figure 8**). The true r^2 value for what looks like a great fit is just 8.65% (**G10**), which is calculated as

True $r^2 = 1-(1{,}245.47/1{,}363.46)^2 = 8.65\%$. (2)

Because we are dealing with time series data, we are not constrained to calculating r^2 against a flat line, but we can compare the model error with the best of the error measures of the five naïve models found in **Figure 6** and **Figure 7**. This Next=Last RMSE error from **Figure 6** & **7** (cell **D10**) is shown in cell **G8** and the r^2 result shown in cell **G10**. This reveals the hidden, dirty secret of R^2 and regression when time series data is used.

CONCLUSION

Given the results of the error measures of the SMA2, SMA3 and the Regression Quartic, the two SMA forecasts (32,264.8, 32,367.0) can be trusted and the Regression forecast of 32,923.7 not trusted (almost a half-million dollars more). However, the difference in all three forecasts for Quarter 21 is less than 2%, but the true r^2 values measures of these three models should make a huge difference to any analyst. Thus, the SMAs' forecasts fit the data much more closely than Regression for this data set.

Figure 10. Indices for all 3 Models

Comparative Indices			
	a)	b)	c)
Qtr 1	112.68%	113.90%	114.01%
Qtr 2	93.32%	93.35%	92.91%
Qtr 3	80.67%	80.27%	80.31%
Qtr 4	113.33%	112.48%	112.78%

Figure 10 shows the quarterly indices for all 3 models. Index set a) is the quick and dirty indices from Seasonality Graphs and b) is from the results of the SMA2 and c) from SMA3. The fact that there is negligible difference between each quarter all three columns should confirm there is enough seasonality to take it into account. However, the graph in **Figure 2** does show that by the 5th year of data perhaps the seasonality is lessening.

No one best forecast model

However, there may be times when Regression curves yield a better fit than SMA's for time series data. Also it is not always true the SMA3 will be better than SMA2 since there is a loss of one degree of freedom. Each data set will have its own quirks, and it is recommended that all three templates be employed for whatever data sets you have until you are confident that one template is consistently outperforming the others. As time marches on, you may want to add the latest quarter and delete the original starting quarter to keep 20 data points as you update your analysis. A good rule of forecasting is to manually adjust data and forecasts considering special circumstances, but always annotate what has been adjusted and why.

Sample size

Although 5 years of quarterly data (n=20) was used for this paper, each of the templates could use the last 4 years (n=16) or even just 15 or 18 last quarters or any number of quarters but should really have at least 4 years of data. In any case, the data should reflect the same general ups and downs over that time period. The only thing to be adjusted is the graph axis. The idea is to use enough valid data to get a sense of the seasonality.

Files for this paper
If you desire the Excel templates, PDF file, and/or PP slides for this paper, please contact the author, who will email the file(s).

APPENDIX

Transitioning from weighted averages to proportion of increase/decrease for Solver Moving Average is accomplished for SMA2 and SMA3 as shown below.

2-Period Solver Moving Average (SMA). Hesse (1999-2020)

Before the inclusion of the Solver and gradient search engine by Frontline Systems into spreadsheets (first Lotus, then Excel and Quattro Pro in the late 1980s), we could only guess at what weights (W_{-1} and W_{-2}) for the last 2 periods would reduce the error. To demonstrate this technique, we will use two periods to get an average or track a trend and the Solver will determine two parameters for a 2-period Solver Moving Average (SMA) and we will lose two degrees of freedom (**k=2**) out of the **n** data points. If we have weights for the last two periods of W_{-2} (two periods ago,) and W_{-1} (last period), we can forecast this period with the following formula for period 3 with the weights unrestricted. This means the could be negative or ABS(W_i) > 100%.

$$F_3 = (W_{-1}*Y_2 + W_{-2}*Y_1)/(W_{-2} + W_{-1}) \quad (3)$$

where Y stands for the data and F the forecast. If we require $W_{-2}+W_{-1} = 100\%$, we don't need to divide by the sum. We can also use the spreadsheet Solver to find the optimal weights to minimize any error measure we choose (RMSE, MAD, or MAPE).

$$F_3 = (W_{-1}*Y_2 + W_{-2}*Y_1) \text{ for all forecasts from 2, 3, ..., n.} \quad (4)$$

- Use Solver to determine weights. Now break the rules: make weights unrestricted (could be negative or > 100%) and add a constraint that weights must add up to 100%.
- Algebraically substitute last weight in terms of the others and change problem from 2 or 3 variables with a constraint to just 1 or 2 variables unconstrained.

$$\text{Since } W_{-1} + W_{-2} = 1 \rightarrow W_{-1} = 1 - W_{-2} \quad (5)$$

$$\text{Substituting into equation (3) gives } F_3 = (1- W_{-2})*Y_2 + W_{-2}*Y_1 \quad (6)$$

Recombining gives

$$F_3 = Y_2 - W_{-2}*(Y_2-Y_1) = Y_2 - W_{-2}*\Delta_1 \quad (7)$$

with

$$\Delta_1 = Y_2-Y_1 \quad (8)$$

Thus

$$F_3 = Y_2 + P_1*\Delta_1 \quad (9)$$

and

$$P_1 = -W_{-2} \quad (10)$$

where Δ_1 is the difference between the last two data points and P_1 is the percent of that difference (positive or negative) added to the last data point. This reduces the equation to just one unknown value, P_1, and no constraint.

3-PERIOD SOLVER MOVING AVERAGE (SMA3) Hesse (1999-2020)

A 3-period SMA might allow for more stability by using the last three periods of data to forecast a fourth. However, it may not minimize more error than SMA2. The original algebraic model would look as follows with unrestricted weights for the last 3 data points to predict the next data value.:

$$F_4 = W_{-3}*Y_1 + W_{-2}*Y_2 + W_{-1}*Y_3 \quad (11)$$

and

$$W_{-3} + W_{-2} + W_{-1} = 1 \quad (12)$$

Again the weights should sum up to 100%. This can be rewritten in much simpler terms. Solving for W_{-1} in [10] and substituting into [9] and adding zero as ($W_{-3}*Y_2 – W_{-3}Y_2$) and combining terms gives:

$$F_4 = Y_3 – (W_{-2} + W_{-3})*(Y_3 – Y_2) –W_{-3}*(Y_2 – Y_1) \quad (13)$$

or

$$F_4 = Y_3 +P_1*\Delta_1 + P_2* \Delta_2 \quad (14)$$

where Δ_2 is now the difference between the 2nd and 3rd last data points.

We have now simplified the mathematical problem from 3 variables and one constraint to just two variables (**P_1, P_2)** and no constraints. The second P_2 value allows the forecast to be modified a bit from the initial trend, and as we will see, this forecast model can fit a parabola, where SMA2 perfectly fits a linear model. Hopefully SMA3 will anticipate curves up or down in the data. Because this is a quadratic fit, it allows the model to predict upturns and downturns.

REFERENCES

Google Image **Figure 1** retrieved from https://www.google.com/search?q=feeding+California+firefighters&rlz=1C1CHBD_enUS773US773&sxsrf=ALeKk00mpP0z6cUWb2r3mIeYz5IgEXInQQ:1599573472268&tbm=isch&source=iu&ictx=1&fir=aq-JN0t-ZXvMnM%252Cc0-XEwi4BMrpLM%252C_&vet=1&usg=AI4_-kQFwI5WEt8gBRTgw9ku_7OfFmtSbw&sa=X&ved=2ahUKEwiM3ezE29nrAhXw01kKHV9vAmsQ9QF6BAgOEDk#imgrc=aq-JN0t-ZXvMnM

Hesse, R, (2018). *A New r^2 for Time Series Data*, Proceedings of Decision Sciences Institute National Conference 2018, Chicago, Il.

Hesse, R, (1997). *Managerial Spreadsheet Modeling and Analysis*, Richard D. Irwin, 1997 (now McGraw-Hill). LUMAT is an electronic updated version of this book with Statistics and several other topics added.

Hesse, R, (1999-2020). *LUMAT: Learning to Use Managerial Analysis Templates*, Azel/Blue Pond Publishing, © 1999-2020, Electronic Text LUMAT 2: Forecasting (LUMAT 204 pp 11, 22).

DECISION SCIENCES INSTITUTE
A Resilient Global Supply Network Design Model

Kanchan Das
East Carolina University, College of Engineering and Technology.
Greenville, NC 27858, dask@ecu.edu, Tel: 252 737 1905.

ABSTRACT

We propose a design model for creating a resilient supply network applicable for local and global markets. It considers the ineffectiveness of existing supply network that contributed to current economic, food and hospital supply crisis due to COVID 19 Pandemic situation. The research includes resilience creation approach for the known previously experienced as well as unknown and unforeseen disruption situations. This research includes a resilience management-based approach considering such methods in the literature to building abilities of supply networks for absorption, adaptation and recovery for containing disruptions that have severe negative effect on supply chain operations and society.

KEYWORDS: Supply network, Resilience creation, Design model; Resilience management approach, Absorption, Adaptation, Recovery

INTRODUCTION

COVID 19 created economic crisis and disrupted total social ecosystems that we have been facing every day. It also challenged supply chain (SC) resilience creation approaches of the SC literature (Ivanov & Dolgui, 2020, Ivanov et al,2020). Supply system disruptions for hospital supply to day-to-day useable items establishes that our approach should examine the supply network and their intertwined structure, not individual SCs and the set of suppliers dealt by them. Since supply networks are intertwined and complex (Ivanov and Dolgui, 2020; Braziotis et al, 2013), examination of supply networks will provide insights about the options for containing such disruptions by identifying better performers or still operating suppliers. Selection of such suppliers may provide options for supply system restoration and continuity. Since supply network disruptions experienced by SCs in this COVID 19 Pandemic are unprecedented, risk management-based approach does not seem to address containment options effectively (Linkov et al, 2017). This is because such approach is probability based. While assumptions and scenario-based steps may provide some options to proceed with risk management-based approach, this research follows resilience management approach considering absorptions, adaptation and recovery strategy considering unforeseen and uncertain disruptions and disruption impacts.

Extent of COVID 19 pandemic disruptions is enormous everyway that may not be contained until regional and global coordination is planned (Kimura et al, 2020). This is such, because SC networks are always inter-country, inter regional and global in nature. For a robust approach such regional and inter regional cooperation may provide containment options by finding less disrupted and not disrupted organizations in some safer locations. Such regional cooperation also supports early detection and initiation steps to flatten the effect of any disease outbreaks pandemic curve which will also ease out demand side panic for over stocking. Such early mitigation policies are also supported by study or World Bank (2020). It is also important to note that such regional and inter-regional policy coordination and the stability of global value chain will be valuable in the post-pandemic recovery for the businesses in the regions.

Such coordination has been broken down and may not be obtained soon. But since this research is not time bound, we design our supply network applicable to global and interregional supply network, when needed and applicable.

SC resilience and sustainability may refer to same state of a system (Marchese et al, 2018; Fiksel et al., 2014) when we apply these terms for normal business operating conditions and persistence of the system over time. SC intermediaries may be used to facilitate overcoming failure of supply system sustainability and resilience (Cole and Aitken, 2020, Vedel and Ellegaard , 2013).). Such intermediaries are often a part of global supplier network. By SC intermediaries Cole and Aitken, (2020) considered distinct organizational form that occupies a position in global sourcing, may be a supply contractor and may sometimes become a supplier organization for supplying finished goods. In this Pandemic situation markets have been operating only in a limited scale for online transactions, and most of the production organizations have not been operating due to non-availability of required employees. As such traditional supply management systems are also in non-operative conditions.

In this Pandemic disruption situation such options of involving SC Intermediaries may be considered, especially, when all other options are closed or not seemed to be working.

In the case where supply management system is still operating with limited capacity (for example for e-markets, medical supply chains, pharmaceutical industries) by facing the pandemic situation crisis, we need to understand the factors and facilitating conditions for making them stronger in future. Based on the findings and recommendations in Matapoulos et al, (2019) perceived organizational justice in buyer supplier relationships for risk sharing, and development of stronger relationships and commitment contributes in financial and operational performances in the crisis situation. Development of such risk sharing, and development of stronger relationships may be considered to face future Pandemic situations. In addition to empirical research findings, the approach proved successful with the outcomes Matopoulos et al's study obtained from their analysis of grocery retailing network during the Greek financial crisis, which further exemplifies the usefulness of the approach.

In this research we plan to design a resilient supply network in place of traditional supply system for an individual SC for containing unforeseen and unprecedented business and SC disruptions taking resilient management-based approach. We shall include above two(considering SC intermediaries, and perceived organizational justice in buyer supplier relationships for risk sharing, and development of stronger relationships and commitment) as the potential resilience improvement options in our designs as much as possible.

In addition, information visibility has been given importance in the recent study of Li et al. (2017) and past literature (Basole and Bellamy, 2014; Craighead et al, 2007). All these researches emphasized importance of information visibility in terms of access to disruption information for failure mitigation taking risk management-based approaches. From the resilience management perspectives such visibility could provide adaptation and quick recovery options. In addition, such visibility could remove demand shock and out of stock panic by the consumers by communicating through social media and taking advantages of already established IBM's IoT based higher level technology (Sanders, 2014).

Based on above discussions and information this research introduces a noble supply network design model that will be able to handle a set of SCs considering them as customers to supply networks. Since it is supply network, it will include flexibilities, decentralizations and inventory creation at strategic locations in the perspectives of the supply networks (applicable for intertwined global and local situations. The research is presented in the following structures: Section 2 includes relevant literatures and section three includes Research Methodology in terms of Mathematical Modeling based approach and section 4 includes a numerical example and Section 5 discusses and concludes.

LITERATURE REVIEW

Literature on SC resilience is quite rich. But only a limited number of researches are available that cover SC disruptions and relevant resilience considerations considering disruptions and changes created by COVID 19 situations. Since number of articles are limited, we shall study literature under the streams of 1. supply management disruptions and resilience; and 2. Supply network management disruptions and resilience.

Supply Management Disruptions and resilience

Taking the advantage of globalization SCs go for multitier supplier for cost and quality related advantages. Bimkis et al. (2019) followed a modeling-based approach where they plan production in their multitier SC management system after observing realized output by suppliers and by that they optimize cost. Planning of production based on confirmed supply items availability is equivalent to spot buying. Such approach may not be considered suitable to build resilience in COVID19 like Pandemic situations.

Implications of COVID 19 disruptions on Food supply chains (Hobbs, 2020) have not been much different from other products considering initial demand side peaks and shocks out of panics. Such demand shocks with the time became lower demand for prepared and packaged food when people were consuming home-made food only. Supply side shocks in food supply chain continued mainly because of labor shortages and transportation disruptions. Hobbs (2020) discussed on Canadian situation including the issue of traffic thickening at the USA-Canadian Border during this COVID 19 pandemic. Hobbs (2020) also studied longer lasting effects on the food supply chains including on-line grocery delivery sector and whether that sector will influence local food supply chains. On the resilience creation the author provided insights on collaborative supply chain relationships with long-term partnerships with suppliers that help reduce transaction costs, share risks, could provide access to complementary resources and expertise, and enhance productivity. The author considered buyer supplier relationship following Matopoulos et al, (2019)'s approach for risk sharing, and development of stronger relationships and commitment within buyer–supplier relationships.

Rezapour et al (2017) studied an automobile supply chain disruption case study for the supply system disruptions. Based on their detailed analysis they planned three policies that include: keeping emergency stock at the retailer's level. Such stock keeping is in effect reserving back-up capacity at the supplier's level, and multiple sourcing for containment of disruptions. Such steps seem to have some effectiveness even in the COVID 19 disruptions, when one does not run production and not need supply by suppliers at lower tiers. But for taking advantage of stock at the retailer's level and making the disruption containment steps to work, transportation arrangement should be organized.

Purchasing and supply management (PSM) plays a key role for supply management performance and overall supply chain performance, as reinforced in a recent study by Pereira et al.(2020). PSM has been responsible for pursuing proactive and reactive measures for ensuring procurement management performance, thus contributing to resilience. For such measures PSM had the authority over resources in addition to external and internal relations. From this perspective they had the ability to take reactive measure to procure from supply chain intermediaries as proposed by (Cole and Aitken,2020) to tackle the traditional supply management failure cases considered in our research. Keeping backup supplier improves SC resilience under uncertain demand and supply disruption potential. Chakraborty et al. (2020) studied effect of keeping backup supplier under uncertain demand. Their finding using a game theoretic and price strategy-based approach showed improvement of supply management resilience through the backup supplier. But with COVID 19 situations such backup supplier did

not work when suppliers' organizations stopped their operations. In some cases, operations of the supply organizations failed because of transportation failures. So, options were multiple layers of backups with alternative transportations. The study of Tsai (2016) considered optimum layers of backups and recommended a dynamic approach for including such layers of backups. Question may come, how to decide number of layers. Since the study is based on dynamic sourcing, multiple scenarios based, probabilistic approach will support such backups. In this case probabilities and scenarios may be developed dynamically, considering characteristics of disruptive situations to decide optimum backup layers.

Supply Network Disruptions and Resilience

Select research considered supply networks as the intertwined type or interconnected within the network. These organizations are conceptually considered to take buyer suppler roles (Ivanov et al. 2020; Ivanov & Dolgui, 2020). Since they supply or interchange various goods (products, production inputs) and sometimes take the role of service provider for some services (mobility services), these interconnected SCs on their own can support each other to attain required resiliency. The authors considered developing resilience mechanisms to interchange supports (type of collaborative organizations) to each other based on visibility for requirements and capabilities of each other. Such resilience mechanisms seem to have the potential to support organizations in COVID 19 type pandemic situations by sharing and interchanging resources, capacities and capabilities. The study in Ivanov & Dolgui (2020) has not provided more details beyond conceptualization.

Chibani et al., (2018) considered a set of supply chains and planned e-procurement using a model-based approach for dynamic optimization of cost to handle uncertainty in procurement quantity and cost. The authors used genetic algorithm for solving the model. In each generation of solution, the algorithm changed the supplier mix. The overall problem and the solution for building resilience seem to work in the conceptual level only.

Supply networks face random disruptions from natural calamity (earthquakes, tsunami) and targeted disruptions from military interventions/war, strike, employee noncooperation. For building resilience to such disruptions Shi et al. (2019) proposed an undirected graph theoretic model for transactions of products between suppliers (node) and enterprises(node) and from enterprises to retailers. They used a computer simulation to solve the model when such transactions are disrupted from random disruption events. The authors claimed their results in terms of LCC (number of subnetworks where each pair of nodes can be interconnected) better than the similar solutions in the literature. The result seems interesting but not much suitable to provide managerial insights for resilience when the approach is of conceptual in nature.

Azad and Hassini (2019) proposed recovery strategy for a supply network disruption situation. Their strategy considered dynamic pricing strategy, inventory keeping, transshipment and multiple sourcing for recovery from supply network disruptions. Each of these strategies proven to work very well in normal business situation but does not seem suitable for COVID 19 like situations.

Based on the above literature review; within the limited number of researches that addressed supply network disruptions to plan resiliency that had the potential to address disruptions caused by COVID 19 like Pandemic situations, resilience development approaches are still at the conceptual level. Individualized supply management resilience planning for containing disruptions are well planned but not suitable to handle disruptions that may arise from COVID 19 like pandemic situations. Only exception here is the Food SC study (Hobbs,2020) which is well suited for addressing Pandemic situations. In our proposed model we considered supply network that may be formed by suppliers, supply contractor (SC intermediaries, as mentioned

by Cole and Aitken, 2020). As such it can only be partially shut down which will be in operation quickly with some better situations. This is such, because SC intermediaries operate virtually in computer-based systems, sometimes as a broker (Li and Choi, 2009) to establish relations between a supplier and buyer through computer-based information exchange. Such intermediaries may only need some rented warehouses to use as collection centers / distribution centers for creating planned stock of items as needed by the customers and distribute them to customers by taking support from 3PL logistics providers, which have been working in some form in this pandemic situations.

METHODOLOGY

Considering various parameters and number of entities to be included for containing disruptions, this research follows a mathematical modeling-based approach for designing a resilient supply network. We have developed a bi-objective mixed integer programming model. This section includes *Notations* for defining model equations, Problem Statement, and Formulation of Model.

Notations

C: set of customer organizations (used as customers in the paper) $c \in C$
G: set of 3PL services provider (3PL SP) company for providing transportation services mainly; $g \in G$.
I: set of supply items $i \in I$. Such items include components, accessories, small product, like mask, and big product, like ventilator, generator.
L: Set of Locations $l \in L$; set of locations for suppliers $l \in L^{s;}$ set of locations for distribution nodes $l \in L^{n}$; set of locations for customers $l \in L^{c}$; we consider L^{s}, L^{n}, $L^{c;}$ are partitions of *L*.
M: set of markets $m \in M$ for supply contactor (SCON) to supply product *i*
N: set of nodes $n \in N$; set of distribution node (DSN used in the body of the paper) $n \in N^{k}$, *here* N^{k} is the set of *DCs* (Warehouses, collection center) maintained by SCONs; set of distribution centers(warehouses) that are identified to be in comparatively safer location; $n \in N^{SF}$; set of customers or customer nodes $n \in N^{c}$ that order items to SCON *v*. Customer nodes N^{c} are sometimes mentioned as destination node when warehouses/collection centers cannot operate due to some disruptions or some abnormal situation as COVID 19.
S: set of actual suppliers (ACS) $s \in S$ representing local or foreign Pharmaceutical or any other manufacturing companies, or a grower in certain occasions.
V: set of supply contractor (SCON) $v \in V$ (right term for such SCON is the SC intermediaries or agents that procure supply item $i \in I$ from actual local or foreign suppliers *s* and supply to their customer c.

Parameters

AVY_{isng}: average per day inventory for item *i supplied by supplier s to distribution node (DSN)* $n \in N^{k}$ *by taking transportation support from 3PL SP: g.*
αu_{sl}= *is a 0/1* parameter; =1, if supplier s is from location *l*, 0 otherwise.
CS_{isn}: cost of supplying item *i* to DSN $n \in N^{k}$, by suppliers *s.*
$CAP_{is:}$ capacity of supplying item *i*.by supplier s.
FSC_{is}: fixed cost for ordering *i* to supplier s.
FTR_{sng}: Fixed cost for selecting 3PL SP *g* for transporting items from supplier *s* to DSN $n \in N^{k}$.
$PR_{r:}$ probability of scenario *r.* Let us consider *n* scenario with probability value 1/n for each scenario.
SLT_{sn}: average supply lead time of supplier *s* to supply input items to DSN *n* based *on* scenario-based analysis for scenario *r to be included in the constraint* (16) through $AVSL_{snr}$ (average

supply lead time for supplying item from supplier *s* to DSN n at a scenario r) based on Google Map based disatnce and assumed supplier location. Scenarios for average transportation time may be defined in the following way: scenario 1: normal operating condition with average traffic density; scenario 2:transportation time in COVID 19 situation with rare traffic but the driver cannot take coffee, food, and often closed rest area due to state emergency situation; scenario 3:emergency lifted, limited services to drivers including food, coffee and water available ; scenario 4: traffic condition and services to drivers are Okay, but traffic movement is sometimes disturbed due to , processions/some form of natural calamity.
SW_{nl}: is a 0 / 1 parameter decided by SCON for selecting location *l* (position) for DSN *n* considering position criteria (nearness to market *m*, nearness to supplier *s* availability/nearness to transportation hub, and safe locations to avoid risk proneness for natural calamity and disease outbreak.
$TRC_{isng:}$ cost of transporting item *i* from supplier *s* to DSN *n* through 3PL SP *g*.

Decision variables

aux_{sn} : an auxiliary 0/1 variable that acts to mathematically balancing constraints (24) and (25).
bu_{isl} =1, if supply order for item *i* is assigned to supplier *s* of location $l \in L^{s:}$; 0 otherwise
dy_{incmg}: distribution of supply item *i* from DSN *n* to customer *c* of market *m* through 3PL SP contractor *g* to fulfill market requirement
ey_{isng}: effective amount of item *i* supply to DSN *n* as computed in constraint (17) by considering normal supply item stock received from supplier s (y_{isng}) and safety stock as computed by considering supply lead time ($SLT_{sn)}$ *and* average per day inventory consumption (AVY_{isng}).

$ntr_{isng:}$ 1, if for transporting item *i* from supplier *s* to DSN *n* 3PL SP (transportation) *g* is selected; *0* otherwise.
$u_{is:}$ 1, if order for supplying item *i* is assigned to supplier *s*; 0 otherwise.
w_{in}: 1, if item *i* is allocated to DSN *n* and is open to receive load for the item; 0 otherwise.
x_{iv}: requirements of supply item *i* by the SCON *v.*
$xz_{vsi:}$ actual supplier *s* is assigned supply order for item *i* to be delivered to distribution node by the SCON *v*
y_{isng}: item *i* from supplier *s* be to be transported to a DSN *n* through 3PL SP *g*.
$z_{vicm:}$ SCON *v* is assigned a procurement orders for supplying item *i* by a customer $c \in C$ from market *m.*
αI_{ncm}: 1, if DSN *n* is allocated to supply customer *c* in market m, 0 otherwise.
β_{isn} =1, if supplier s is assigned to supply comparatively non-risk prone DSN $n \in N^{SF}$, 0 otherwise.
τ_{isn} : 1, if supplier *s* is assigned to supply DSN *n*, 0 otherwise.

Problem Statement

We assume a supply network formed by a set of supply contractor (SCON) *V* from which a set of items(products) *i* are procured by a set of customer (customer organization) *C*. The supply contactors (SCONs) V ($v \in V$) are globally located with different virtual and website addresses in different countries and may be reached by any company in the USA or from any other country. SCON *v* uses set of different transportation modes through transportation contractor (3PL SP (service provider) *g* for supplying item (product) *i* to customer *c*. Let z_{vicm} be the item $i \in I$ procured by the customer *c* of market *m* from SCON v. The SCON *v* organizes supply of items *i* needed by customer *c* of market *m* through a set of actual supplier *S*. The SCON *v* assigns $xz_{vsic:}$ supply order to actual supplier (ACS) *s* for delivery of item *i to* a set of distribution

/collection nodes (DSN) N^k ($n \in N^k$). The DSNs or distribution center or collections nodes are combination of warehouses and cold storages used as distribution nodes (DSNs, DCs) and for stocking items by the SCON. The SCONs maintain the stock and keep them ready to be opened when needed for complying with delivery schedule of customers. Such stocks also provide distribution flexibility and resilience to transportation and other failures on the inbound supply of items from suppliers to DSNs and ensures distribution of items to customer when some supply base become disconnected due to some disruptions or COVID 19 like problems. We assume actual suppliers are from a set of locations L^s ($l \in L^s$). Location for each of these supplies are known based on their enlistment with the SCON *v.* SCONs maintain supply portfolio with all details for suppliers and supply locations. Based on the locations of ACSs and customers, SCON assign dedicated ACSs to some valuable and big customers in some occasions when DSNs are disrupted. SCONs plan supplier assignment to supply to DSNs ensuring supplier flexibility (more than one supplier per supply item); supplier location flexibility (assigning suppliers from more than one location for the crucial items of a customer based on information from customer); assigning at least one supplier for a DSN (distribution node, warehouse, distribution center) located in safe or non-risk prone location for each crucial item.

The overall objective of the model is to make the supply networks resilient to disruptions and optimization of overall supply cost to customers. Based on the experience SCON (v) about supply situation, SCON (v) work with suppliers (ACS) to organize transportation of the supply items *I* through 3PL SP (transportation contractor), with which SCON has the contractual arrangement). SCON informs transportation cost per unit of supply items through 3PL SP to suppliers to support them in transporting items to DSN. Such support through 3PL SP (service provider) will eliminate not getting transports/trucks or any other mode of transportation by *S* (ACS) during any abnormal situation (calamity, disease outbreak). Let y_{isng} : input $i \in I$ to be transported to assigned/selected DSN *n* through the 3PL SP *g*. Objective of the Model is to maximize *Resilienc*y and minimize **Total Cost** for supply network operation as defined in Objective functions 1 and 2 to be followed.

Formulation of the Model

Constraint (1) estimates overall requirement of a supply item for a SCON's customers from all markets.

$$x_{vi} = \sum_{c \in C} \sum_{m \in M} z_{vicm} \qquad \forall v, i \qquad (1)$$

Equation (2) balances supply order received by a SCON from the customers of different markets with procurement order placed by the SCON.

$$\sum_{s \in S} xz_{vis} = \sum_{m \in M} \sum_{c \in C} z_{vicm} \qquad \forall v, i \qquad (2)$$

Constraint (3) limits assignment of product /item to a supplier based on its capacity for the item. Constraint (4) assigns more than one supplier for supplying a product/item. Constraint (5) verified and validated 0/1 variable bu_{isl} for the location of the supplier for assigning order to obtain location flexibility. Based on constraint (6), each input item is assigned to suppliers from more than one location. Equation (7) balances quantity of items transported by the suppliers (ACSs) to DSNs to the quantity ordered to them. Constraint (8) ensures transportation of an

item to DSNs that is open, and it also limits transported quantity according to the capacity of the DSN for accommodating the quantity. Equation (9) computes average per day inventory as supplied by a supplier considering y_{isng} to be the supply quantity for one month with 20 operating days.

$$\sum_{v\in V} xz_{vis} \le u_{is} CAP_{is} \qquad \forall i,s \qquad (3)$$

$$\sum_{s\in S} u_{is} \ge 1 \qquad \forall i \qquad (4)$$

$$au_{sl} bu_{isl} \le u_{is} \qquad \forall i \qquad (5)$$

$$\sum_{s\in S}\sum_{l\in L} bu_{isl} \ge 1 \qquad \forall i \qquad (6)$$

$$\sum_{n\in N}\sum_{g\in G} y_{isng} = \sum_{v\in V} xz_{vis} \qquad \forall i,s \qquad (7)$$

$$\sum_{g\in G}\sum_{s\in S} y_{isng} \le w_{in} CAP_{in} \qquad \forall i,n \qquad (8)$$

$$AVY_{isng} = \sum_{s\in S} y_{isng} / (20 * \sum s) \qquad \forall i,n,g \qquad (9)$$

Constraint (10) ensures allocation of more than one DSN (distribution node: distribution centers, warehouses), for each supply item *i*. Constraint (11) plans to allocate DSNs based on their pre-decided locations/positions. Considering non-risk proneness as defined in the notations. Constraint (12) balances supply item to DSN with the distributed product from the DSN to Customers in the market. Constraint (13) balances distribution of product to customer at the markets with the order quantity committed by SCON.

Constraint (14) allocates DSN (distribution centers, DCs, warehouses) to a customer in a market. Constraint (15) allocates more than one DSN (DCs, warehouses) to supply a customer of a market. Equation (16) estimated supply lead time based on Average lead times taking scenario-based analysis. Since there is no supply lead time data to estimate standard deviation for a situation like COVID 19. Equation (17) computes effective supply quantity to be sent to a DSN by including safety stock as computed in the equation and original supply quantity y_{isng}., We plan to maintain safety stock for the entire estimated supply lead time as computed in the equation (17). We assume that it will be a onetime safety stock creation. Constraint (18) assigns 3PL logistics service provider (for transportation services) for sending effective supply quantity to DSN.

$$\sum_{n\in N} w_{in} \ge 1 \qquad \forall i, n \in N^k \qquad (10)$$

$$w_{in} \le sw_{nl} \qquad \forall i,n,l \in L^n \qquad (11)$$

$$\sum_{s\in S} y_{isng} = \sum_{c\in C}\sum_{m\in M} dy_{incmg} \qquad \forall i, n\in N^k, g \qquad (12)$$

$$\sum_{n\in N}\sum_{g\in G} dy_{incmg} = zv_{icm} \qquad \forall i, c, m \qquad (13)$$

$$\sum_{i\in I}\sum_{g\in G} dy_{incmg} \le al_{ncm} M \qquad \forall n, c, m \qquad (14)$$

$$\sum_{n\in N} al_{ncm} \ge 1 \qquad \forall c, m \qquad (15)$$

$$SLT_{sn} = \sum_{r\in R} AVL_{snr} PR_r \qquad \forall n, c, m \qquad (16)$$

$$ey_{isng} = y_{isng} + \sum_{l\in L} SLT_{sn} AVY_{isng} \qquad \forall i, s, n, g \qquad (17)$$

$$\sum_{i\in I} ey_{isng} \le ntr_{gsn} M \qquad \forall s, n, g \qquad (18)$$

Constraint (19) allocates DSN to a supplier for ensuring its replenishment. Constraint (20) assigns more than one supplier for replenishing a DSN. Such assignment improves supply resiliency to a DSN through increase in supply flexibility. Constraint (21) ensures assignment of supplier to a DSN is effective when the DSN is in operation. Constraint (22) ensures assignment of at least one supplier to a DSN which is located in a safer zone. Constraint (23) ensures that safer location DSN is a part of DSNs operated by SCON. Constraints (24) and (25) work in a combined way and decides extra quantity of supply items(Inventory) to be procured from partner suppliers to compensate the shortages between the quantity of items ordered by SCONS to suppliers to fulfill market requirements and the quantity of items supplied by the supplier to DSNs. Extra inventory to be procured will be triggered by the 0/1 auxiliary variable aux_{sn}. These two equations in effect created resiliency improvement provisions for obtaining the extra inventory from partner suppliers if needed. The extra inventory or improvement options has the basis in constraint (4) that included supplier flexibility to have more than one supplier for each item.

$$\sum_{g\in G} y_{isng} \le \tau_{isn} M \qquad \forall s, n, g \qquad (19)$$

$$\sum_{s\in S} \tau_{isn} \ge 1 \qquad \forall i, n \qquad (20)$$

$$\tau_{isn} \le w_{in} \qquad \forall i, s, n \qquad (21)$$

$$\sum_{s\in S} \beta_{isn} \ge 1 \qquad \forall n \in N^{SF} \qquad (22)$$

$$\beta_{isn} \le \tau_{isn} \qquad \forall i, s, n \in N^k; n \in N^{SF} \qquad (23)$$

$$\sum_{i\in I}\sum_{g\in G}(y_{isng}+IA_{isng})-\sum_{i\in I}\sum_{v\in V}\sum_{c\in C}\sum_{m\in M} z_{vicm}\le M(1-aux_{sn}) \qquad \forall s,n \tag{24}$$

$$IA_{isng}\le aux_{sn}M \qquad \forall s,n \tag{25}$$

Objective Function 1: maximizes resilience (*RS*) as defined in equation (26). RS is defined as a measure of coefficient of performance in equation (27) as the ratio of quantity of items procured and sent to DSNs by the suppliers to the quantity of items SCON (supply contractor) committed to supply to customers by accepting orders. We added extra quantity IA_{isng} based improvement options in the case desired resilience performance is not achieved. Extra quantity-based improvement option will be triggered by the constraints (24) and (25) in a combined way.

Maximize ***Resiliency*** (*RS*) (26)

$$RS=\frac{\sum_{i\in I}\sum_{s\in S}\sum_{n\in N}\sum_{g\in G}(y_{isng}+IA_{isng})}{\sum_{v\in V}\sum_{i\in I}\sum_{c\in C}\sum_{m\in M} z_{vicm}} \tag{27}$$

Objective function 2 defined in equation (28) minimizes total supply network operation cost *TC. TC* is defined in equation (29) in terms of its components: cost of supply network operations *COP*, cost of inventory maintained in DSNs *INVC*. COP is defined in equation (30 in terms of its components: supply management cost *CSUP*, cost for transporting supply items to DSNs *CSTR* and costs for distributing product from DSN to customers CSDR.

Objective Function 2: Minimize cost (*TC*) (28)

$$TC=COP+INVC \tag{29}$$

$$COP=CSUP+CSTR+CSDR \tag{30}$$

Supply management cost *CSUP* is defined in equation (31) by considering cost of supply items and fixed cost for assigning orders to suppliers and cost of including inventory for improvement options when it will be triggered by constraints (24) and (25). Equation (32) computes *CSTR* transportation cost from supplier location to DSN according to contractual rate of per product for such transportation and fixed cost for selecting 3PL SP for transportation. Equation (33) computes distribution cost *CSDR* for distributing product from DSN to customer in each market plus fixed cost for allocating DSN to supply customer in the markets.

$$CSUP=\sum_{i\in I}\sum_{s\in S}\sum_{n\in N}CS_{isn}\sum_{g\in G} ey_{isng}+\sum_{i\in I}\sum_{s\in S} FSC_{is}u_{is}+\sum_{i\in I}\sum_{s\in S}\sum_{n\in N}CS_{isn}\sum_{g\in G} IA_{isng}CS_{isn} \tag{31}$$

$$CSTR=\sum_{i\in I}(\sum_{s\in S}\sum_{n\in N}\sum_{g\in G} ey_{isng}TRC_{isng})+FTR_{sng}ntr_{sng} \tag{32}$$

$$CSDR=\sum_{g\in G}\sum_{i\in I}(\sum_{n\in N}\sum_{c\in C}\sum_{m\in M} dy_{incmg}CDR_{incm}+FDR_{ncm}al_{ncm}) \tag{33}$$

Constraint (34) imposes integrality.

$$u_{is} \in \{0,1\}, \forall i,s; bu_{isl} \in \{0,1\} \forall i,s,l; w_{in} \in \{i,n\}, \forall i,n; sw_{np} \in \{0,1\}, \forall n,p; al_{ncm} \in \{0,1\}, \forall n,c,m$$
$$aux_{in} \in \{0,1\} \forall i,n; ntr_{sng} \in \{0,1\} \forall s,n,g; \tau_{isn} \in \{0,1\} \forall s,n,g; \beta_{isn} \forall i,s,n \quad (36)$$

NUMERICAL EXAMPLE

We assume a set of 5 supply contractors (SCONs) operate in 6 markets where they have customer organizations with demand for 12 products of their product portfolio based on their previous business in the markets. Table 1 presents typical product demand in Market 1.

	Table 1: Typical customer demand for products								
	Demand for products from the 8 Customer in Market 1								
Product	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	Total
1	54	45	61	63	56	59	56	58	452
2	66	60	65	52	49	58	62	54	466
3	60	59	57	59	47	65	53	58	458
4	60	66	52	67	58	51	45	64	463
5	60	68	56	57	69	65	58	56	489
6	66	59	64	63	51	63	57	72	495
7	61	54	67	58	50	56	49	52	447
8	72	48	53	61	64	60	51	54	463
9	42	54	67	54	51	51	59	49	427
10	54	48	61	59	59	59	51	59	450
11	54	56	50	66	59	54	59	65	463
12	62	52	54	54	51	63	55	67	458

The SCONs are networked with several suppliers (ACSs) in USA, other countries and continents through their counterparts (SCONs) operating in those countries and continents. Considering shut down of air travel, export and import provisions in March-April-May of 2020 due to COVID 19 pandemic situations, the SCONs connected their 8 suppliers in USA for the orders they received for their 12 products. The SCONs identified 6 suitable DSNs (Distribution centers/collection centers at an optimum location considering supply time from suppliers and distribution distances from DSNs to 8 customers in 6 markets that ordered products out of their product portfolio. Within the 6 DSNs one of them is in a comparatively safe location where effect of COVID 19 is lower over the period (April-May-June-July).The DSNs will be used for receiving items to be supplied by suppliers and distributing products from there to customers. The SCONs organized opening them for operation. Table 1 presents customer Orders (demand) for 12 products from 8 customers of market1 placed to 5 supply contractors. To take an example for illustrations, based on Table 1, Customer 1 of market 1 ordered 54 units of product 1 and 66 units of product 2.

We solved the proposed model for the example numerical problem using commercial solver LINGO 18.0 in an standard PC Dell Latitude 5590 with Processor Intel Core (TM) i7 8650U CPU @1.90 GHz 2.11 GHz 32-bit operating system with Windows 10 Enterprise, Installed RAM 16 GB. The model solution for the example problem involved 21, 546 total variables, 1249 integers and 5930 constraints.
In this section we present limited input information. We shall mainly present model outputs. Table 2 presents typical model decision for assigning supply orders for products (1 to 6) to suppliers. Based on Table 2 each product is assigned to 2 to 3 suppliers. For an example product 1 assigned to suppliers 3,4 and 5 whereas product 2 is assigned to suppliers 2 and 6 as may be observed in Table 2.

Table 2: Assignment of product supply orders to Suppliers

	Suppliers						
Product	1	2	3	4	5	6	7
1			Y	Y	Y		
2		Y				Y	
3			Y		Y		
4			Y		Y		
5	Y						Y
6					y	y	

Legend: Y: Product at the left hand columns is assigned to suppliers shown in the top Row of the Table

Table 3 presents model decision on allocation of DSNs (distribution nodes) to products. For example, DSN 1 is allocated to Products (3'4,6,8,9, and 10). I we look product wise each product is allocated three DSNs. For an example Product 1 allocated to DSNs 3,4 and 6. Similar are allocations for Products 2,5,7, and 11. Each product is allocated to 3 DSNs as may be observed in Table 3. DSN 3 is in the comparatively safer location.

Table 3: Allocation of DSNs to Products

	Products allocated to distribution node (DSNs) by constraint 8 using (w_{in})											
DSNs	1	2	3	4	5	6	7	8	9	10	11	12
1			Y	Y		Y		Y	Y	Y		Y
2												
3	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
4	Y	Y			Y		Y				Y	
5												
6	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Legend Y: depicts allocation of DSN row for the relevant Y to the column of Y which is product

Table 4 presents typical model decisions on allocations DSNs 1 and 2 to Markets and products. For an example, DSN 1 is allocated to distribute products to customer 2 in Market 6, customer 6 in market 3 and customer 7 in market 4.

	Table4: Typical allocation of DSNs to markets and customers							
	Assigned DSNs to markets and customers by constraint (14) using 0/1 variable (al_{ncm})							
DSN	1	2	3	4	5	6	7	8
1		M6				M3		
1							M4	
2		M4						
2			M4					
2						M4		
2						M6		

We solved the model for the following scenarios to provide clear insights to Managers who may be interested to plan Resiliency creation for their unique business. The scenarios are also important for general readers. The scenarios are:

1.Normal Business operating condition (100% coefficient of performance, COE)); 2. Due to COVID19 or similar situation from a disease outbreak, big natural calamity (earth quake, tsunami) businesses run in partially shut down condition (say at 65% COE); 3: some transportation related problems due local or political disturbance in the country (80% COE); 4. Some suppliers are unable to supply due to their internal problems (90% COE). The table also presents taking resiliency improvement steps based on set desired level of COE Resilience performance in % as defined by the proposed model.

Table 5: Scenario analysis Table			
Scenario	Resilience metric: Objective 1 value	Total cost (TC) for the Objective 1 value ($ M)	Objective 2, minimize TC $M
1	1	0.56	0.301
2	0.65	0.426	0.2
3	0.8	0.464	0.205
4	0.9	0.496	0.222

CONCLUSIONS

The proposed research introduced a new and innovative alternative approach for fulfilling business requirement in a difficult business environment like COVID 19 Pandemic situations, by involving SC intermediaries (like supply contractors, supply agents, supply agency networks). The research includes resiliency creation options to address business requirements through Resilience management-based approach for containing unforeseen and unanticipated business

disruption. The research also includes various practical scenario-based analysis to provide clear insights to SC managers for improving resilience and cost performances of their unique businesses. The research has the scope of extending it by including classical supplier pools and SC intermediaries in a combined way. The research has the potential to go for future extension by considering more customers, suppliers and DSNs to make suitable for global SC disruptions situations.

References

Azad, N. & Hassini, E.(2019). Recovery strategies from major supply disruptions in single and multiple sourcing networks*, European Journal of Operational Research*, 275,481-501

Basole, R.C., Bellamy, M.A. (2014). Supply network structure, visibility, and risk diffusion: a computational approach. *Decision Sciences*, 45 (4), 1–49.

Bimpikis, K., Candogan, O., & Ehsani, S. (2019). Supply disruptions and optimal network structures. *Management Science*, 65(12), 5504-5517.

Chakraborty, T., Chauhan, S.S., Ouhimmou. M. (2020). Mitigating supply disruptions with a backup supplier under uncertain demand: competition vs. cooperation. *International Journal of Production Research*, 58 (12), 3618-3649.

Cole, R., Aitken, J. (2020). The role of intermediaries in establishing a sustainable supply chain. *Journal of Purchasing & Supply Management*, In press.

Craighead, C.W., Blackhurst, J., Rungtusanatham, M.J., & Handfield, R.B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences*, 38, 131–156.

Chibani, A., X. Delorme, Dolgui,A. & Pierreval,H. (2018). Dynamic Optimization for Highly Agile Supply Chains in e-Procurement Context. *International Journal of Production Research* 56 (17): 5904–5929.

Fiksel, J., Goodman, I., Hecht, A., (2014). Resilience: navigating toward a sustainable future. *Solutions Journal*. 5(5), 38–42.

Hobbs, J., E. (2020). Food supply chains during the COVID-19 pandemic. *Canadian Journal of Agricultural Economics*, 1-6. DOI: DOI: 10.1111/cjag.12237.

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research*, 58(10), 2904-2810.

Ivanov, D., Dolgui, A., Blackhurst, J., Choi, T.M. (2020). Special Issue: Viability of Supply Networks and Ecosystems: Lessons Learned From COVID-19 Outbreak, *International Journal of production Research,* editorial.

Kimura, F., Thangavelu,S.M., Dionisius,N., & Findlay, C. (2020). Pandemic (COVID-19) policy, regional cooperation and the emerging global production network. *Asian Economic Journal* 34(1), 3–27.

Linkov, I., Trump, B.D., & Fox-Lent, C. (2017). Resilience in the face of danger. *ISE Magazine*, 49(3), 32-37.

Li, Y., Zobel, C.W.,Russell, R.S. (2017). Value of supply disruption information and information accuracy. *Journal of Purchasing & Supply Management*, 23, 191-201.

Li, M., Choi, T.Y. (2009). Triads in services outsourcing: bridge, bridge decay and bridge transfer. *Journal of Supply Chain Management,* 45 (3), 27–3.

Matopoulos, A., Didonet, S., Tsanasidis, V., & Fearne, A. (2019). The role of perceived justice in buyer–supplier relationships in times of economic crisis, *Journal of Purchasing and Supply Management*, 25(4), 100554.

Merchese, D., Renolds, E., Bates, M.E., Morgan, H., Clark, S.S.,& Linkov, L.(2018). Resilience and sustainability: Similarities and differences in environmental management applications, *Science of the Total Environment* 613–614 (February) 1275–1283.

Pereira, C.R.,da Silva, A. L., Tate, W.L., and Christopher, M.(2020). Purchasing and supply management (PSM) contribution to supply side resilience. *International Journal of Production Economics,* 228 (October), 107740.

Rezapour, S.,Farahani, R.Z., & Pourakbar, M.(2017). Resilient supply chain network design under competition: A case study. *European Journal of Operational Research*, 259 (3),1017-1035.

Sanders, N.R. (2014). *Big data driven supply chain management: a framework for implementing Analytics and turning information into intelligence*. 1st ed., Pearson Education, Inc.

Shi, X.-Q.,Long, W., Li, X.-Y., Deng, D.-S., Wei, Y.-L. Liu,H.G. (2019). Research on supply network resilience considering random and targeted disruptions simultaneously, *International Journal of Production Research*, Ahead-of-print, 1-19 https://doi.org/10.1080/00207543.2019.1685697

Tsai, W.C. (2016). A dynamic sourcing strategy considering supply disruption risks, *International Journal of Production Research*, 54(7),2170-2184.

Vedel, M., Ellegaard, C. (2013). Supply risk management functions of sourcing intermediaries: an investigation of the clothing industry. *Supply Chain Management: An International Journal*, 18(5), 509-522.

World Bank (2020). World Bank East Asia and Pacific Economic update, April 2020: East Asia and the Pacific in the time of COVID-19. Washington, DC: *World Bank*. [Online; cited April 2020]. Available from URL: https://www.worldbank.org/en/region/eap/publication/east-asia-pacific-economic-update

DECISION SCIENCES INSTITUTE
Considerations for Artificial Intelligence in Cybersecurity

Deanna House
University of Nebraska at Omaha
Email: deannahouse@unomaha.edu

Davin Norman
University of Tampa
Email: davin.norman@spartans.ut.edu

ABSTRACT

The use of artificial intelligence in areas of cybersecurity is necessary for advanced protective measures in the field. AI can take on more tedious functions to free up human workers for less repetitive tasks. Current AI technology is set up to do what its expert workers program/develop. However, future AI and the potential harms that could occur in times of superintelligence need to be taken into consideration now.

KEYWORDS: Cybersecurity, Artificial Intelligence, Ethics, Superintelligence, Collective Intelligence

INTRODUCTION

Artificial intelligence, or AI, provides promise for the future of technology and computing. It allows for opportunities for humans and machines to work together to effectively thwart cyberattacks. Cybersecurity events that put organizations at risk are becoming more difficult to detect and can be persistent in nature (Craig, 2018). While the use of humans for many tasks is the most efficient and effective, there are specific tasks in cybersecurity that are more appropriate for machines. The use of AI in cybersecurity is still in the early stages of development with very specific tasks such as those discussed by Masobuka, Grobler, & Watson (2018). These tasks are focused around detection and prevention of malware attacks.

In order to create artificial intelligence technology, expert knowledge is gathered from experts and implanted into computer software (Mason, 1986). The push to develop AI technology quickly put its development at risk of ethical challenges (Jennings, 2019). AI being used maliciously or altered with malware can put a great strain on society (Wu, Huang, & Gong, 2020). Within cybersecurity, many technical challenges have been solved, but there is a need for merging human factors with the computational aspects (Guitton, 2019).

“Companies and institutions should feel an obligation to think about more than how to design and execute – they should engage in socially responsible processes from the very start” (Schwab (2018), p. 37). Design should also be responsible and include appropriate culture (Clarke, 2019). AI, like humanity, is evolving and adapting to perform more difficult tasks. In order to reach this level of man and machine working together to push organizations to their full potential, we must consider ethical AI tools; particularly in the field of cybersecurity. It is

imperative that cybersecurity AI move forward at a quick pace to keep up with (and preferably stay ahead of) malicious attacks.

LITERATURE REVIEW

Ethics

Ethics determines what is right and what is wrong and why. However, ethics can vary greatly from an individual, cultural, and societal level. According to Bishop (2013) “ethics concerns the moral behavior of individuals based on an established and expressed standard of the group, which is in and of itself a collection of individual values” pg 636. The interesting and also challenging because of the many perspectives and nuances that can exist related to ethics within organizations and in professional groups. Typically, the use of a code of ethics can help establish guidelines and recommendations for ethical behavior (Bishop, 2013).

Ethics is typically broken into the categories normative, meta, and applied ethics. Each category has several subcategories. Some of these subcategories are more relevant to technology such as AI and cybersecurity (as mentioned by Carle, 2003). For example, under normative ethics, utilitarianism takes the position of increasing the good and decreasing the bad in order to make life better (Carle, 2003). Utilitarianism also has a focus on consequences in that the choice should be whatever produces the best consequences (Nathanson, n.d.).

Virtue ethics focuses on character and virtue. The idea of a virtuous person is internal and developed from natural tendencies (Athanassoulis, n.d.). The idea of virtue ethics is not to come up with universal principles that can be applied anywhere.

The greatest good, utilitarianism, may be the best for superintelligent systems (Jennings, 2019). As mentioned by Wu et al., (2020) it is important to build ethical principles in to the software specification process in design and implementation of AI. Areas such as decision making (Denning & Denning, 2020), transparency (Wu et al., 2020), bias (Denning & Denning, 2020;

Artificial Intelligence

AI has definitions that can vary depending on the audience. One method for determining the intelligence of a machine is to conduct a Turing Test (Turing, 1950). While this method will help one determine that a machine is no different than man in its ability to answer questions and is thus intelligent, there are potential shortcomings specific to AI. Specifically, as mentioned by (French 1990) the lack of world experiences related to physical, cognitive, and subcognitive function that are part of our human nature. There is still much to be explored related to AI, as it is not currently intelligent in all areas on its own. For now, humans are required to input whatever general intelligence and skills the machines are lacking (Malone, 2018).

There are three areas of focus related to AI, narrow, general, and superintelligence. Narrow AI focuses on a specific task (narrow area) in which it outperforms a human (Techopedia, 2020). General AI is considered such when AI reaches the same level of intelligence as humanity. General AI will possess “learning, memory, reasoning and abstraction, problem solving algorithms, divergent thinking, convergent thinking, emotional intelligence, and the ability to perform each of these at reasonable speeds” (Ackermann, 2018). We are expected to reach this point by year 2045 or later and then what’s known as Superintelligence shortly

after (between 10-15 years). Superintelligence is when AI's intelligence surpasses that of humans. Because superintelligence can potentially cure cancer while at the same time cause humanity's extinction, this could be either the best or worst thing for mankind depending on what framework is used to create the AI.

There are concerns about the development of technology and the possibility of duality related to AI. In this context, the potential for anything that is developed for good being used for bad (de Ágreda, 2020). For example, AI that is designed to protect an organization's network could be used to attack a network instead. AI that is designed for the personalization of a marketing campaign could instead be used to spear phish someone.

Cybersecurity and AI

Analysts must work alongside AI and provide the tool with a large, labeled and organized, dataset to get the job done and eliminate false positives. People are a key element of a cybersecurity program that is successful (Craig, 2018). In order to address challenges faced in the field, multidisciplinary human behavior research must be conducted along with the integration of knowledge from other fields (Guitton, 2019). It is also important to establish a formal monitoring process for AIs to remain safe and friendly (Jennings, 2019). In addition, implementing ethics by design to address areas of bias related to gender, socio-economic status, educational backgrounds, race or sexual orientation should be a priority (Scroxton, 2019).

By design, AI is becoming ubiquitous to the user and is sometimes even invisible (de Ágreda, 2020) which puts an even greater emphasis on human involvement in the design and in post-implementations. This, coupled with the uncertainty of AI in dynamic environments, can potentially create issues in maintaining a trustworthy and safe system (Wu et al., 2020).

Moving Forward – Collective Intelligence and Ethics

AI of the future will evolve to be in its unique form specific to its tasks (Malone, 2018). This would mean that in the area of cybersecurity, robots will not be physically present protecting our networks and files but instead programs moving around performing tasks. While logs can track what is happening on systems and networks, it is feasible think that a superintelligent AI could learn obfuscation techniques to remain undetected. Obfuscation is used in malware to hide from logs and anti-malware programs. These techniques change frequently and creatively so that it is difficult to detect (Osborne, 2018).

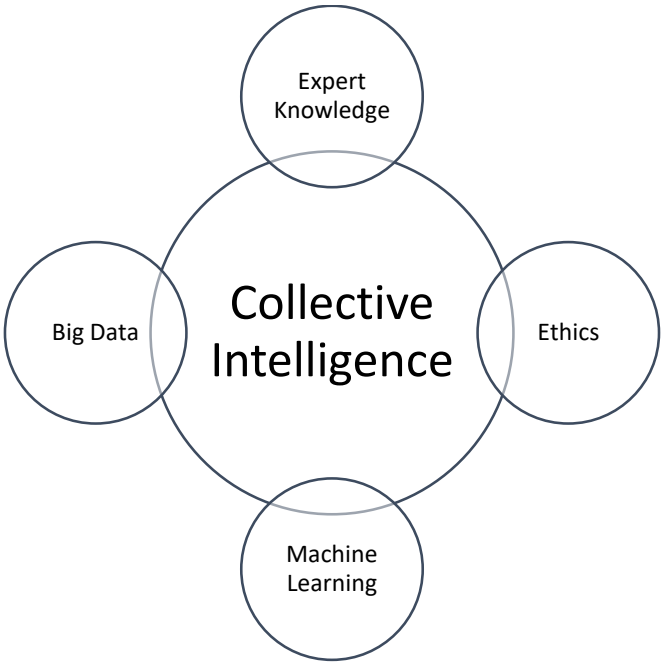


Figure 1: Expanded Collective Intelligence

"Efforts will be needed to ensure that decisions made by machines are programmed in a secure manner, resistant to being subverted or exploited through cyberattacks" Schwab, 2018, pg 129. However, engineers typically do not want to take on the responsibility of making social AI decisions (Jennings, 2019). One area of research that may be a solution to these challenges is that of collective intelligence. The authors propose the expanded collective intelligence model shown in Figure 1. Collective intelligence can be described as human intelligence combined (Malone, 2018). Collective intelligence can also include AI in collaboration with humans take learning to a higher level (Gavriushenko, Kaikova, & Terziyan, 2020). AI transparency is critical to building trust and confidence among the public related to its decision-processes (Wu et al., 2020). While this research is still relatively unexplored, it is important for researchers to take steps to explore what future AI interactions will look like.

References

Ackermann, N. (2018, December 13). Artificial Intelligence Framework: A Visual Introduction to Machine learning and AI. Retrieved June 10, 2020 from https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87.

Athanassoulis, N. (n.d.). Virtue Ethics. *Internet Encyclopedia of Philosophy.* Retrieved June 10, 2020 from https://www.iep.utm.edu/virtue/.

Bishop, W.H. (2013). The Role of Ethics in 21st Century Organizations. *Journal of Business Ethics,* 118(3), 635 – 637.

Clarke, R. (2019). Principles and Business Processes for Responsible AI. *Computer Law & Security,* 35, 410 – 422.

Craig, J. (2018). Cybersecurity Research – Essential to a Successful Digital Future. *Engineering,* 4, 9 – 10.

de Ágreda, Á.G. (2020). Ethics of Autonomous Weapons Systems and its Applicability to Any AI Systems. *Telecommunications Policy,* 44, 1-15.

Denning, P.J. & Denning, D.E. (2020). The Profession of IT Deliemmas of Artificial Intelligence. *Communications of the ACM,* 63(3), 22 – 24.
French, R.M. (1990). Subcognition and the Limits of the Turing Test. *Mind,* 99(393), 53 – 65.

Gavriushenko, M., Kaikova, O., & Terziyan, V. (2020). Bridging Human and Machine Learning for the Needs of Collective Intelligence Development. *Procedia Manufacturing,* 42, 302 – 306.

Guitton, M.J. (2019). Facing Cyberthreats: Answering the New Security Challenges of the Digital Age. *Computers in Human Behavior,* 95, 175 – 176.

Jennings, C. (2019). *Artificial Intelligence: Rise of the Lightspeed Learners.* Lanham, MD: Rowman & Littlefield.

Malone, T. W. (2018). *Superminds: The Surprising Power of People and Computers Thinking Together.* New York, NY: Little, Brown and Company.

Masombuka, M., Grobler, M., & Watson, B. (2018). Towards an Artificial Intelligence Framework to Actively Defend Cyberspace. *European Conference on Cyber Warfare and Security. Retrieved from http://search.proquest.com.esearch.ut.edu/docview/2077000322?accountid=14762.*

Mason, R.O. (1986). Four Ethical Issues of the Information Age. *MIS Quarterly,* 10(1), 5 – 12.

Nathanson, S. n.d. Act and Rule Utilitarianism. *Internet Encyclopedia of Philosophy.* Retrieved June 10, 2020 from https://www.iep.utm.edu/util-a-r/.

Osborne, C. (2018). “A Question of Security: What is Obfuscation and How Does it Work?” Retrieved June 10, 2020 from https://www.zdnet.com/article/a-question-of-security-what-is-obfuscation-and-how-does-it-work/.

Schwab, K. (2018). *Shaping the Future of the Fourth Industrial Revolution: A Guide to Building a Better World.* New York, NY: Crown Publishing Group.

Scroxton, A. (2019). AI Needs to Be Given Ethics by Design. *Computer Weekly,* 19-25 November 2019.

Techopedia. (2020). Narrow Artificial Intelligence (AI). Retrieved June 10, 2020 from https://www.techopedia.com/definition/32874/narrow-artificial-intelligence-narrow-ai.

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind,* 59(236), 433 – 460.

Wu, W., Huang, T., & Gong, K. (2020). Ethical Principles and Governance Technology Development of AI in China. *Engineering,* 6, 302 – 309.

DECISION SCIENCES INSTITUTE

An Empirical Analysis of Physical Access on Emergency Department Utilization under Insurance Coverage Expansion

Eric Xu, Anant Mishra
Operations and Management Science Department, Carlson School of Management, University of Minnesota, 321 19th Avenue South, Minneapolis, MN 55455, e-mails: xuxx1064@umn.edu, mish0049@umn.edu

Kevin Linderman
Supply Chain & Information Systems Department, Smeal College of Business, Pennsylvania State University, Business Building, University Park, PA 16802, email: kevinlinderman@psu.edu

ABSTRACT

In 2010, the Patient Protection and Affordable Care Act (ACA) set out to improve access to healthcare by expanding insurance coverage. This expansion in healthcare coverage, Medicaid in particular, was expected to reduce one of the major cost drivers in American healthcare, emergency department visits. However, since the passage of the ACA, emergency department visits are rising, most notably in Medicaid expansion states. We find that the characteristics of the nearest primary care clinic, hours of operation and the distance to said clinic, are strong indicators of whether an individual will choose to use the emergency department.

Keywords: Healthcare Operations, Access to Care, Physical Access, Affordable Care Act

DECISION SCIENCES INSTITUTE

Unequal but fair: incorporating distributive justice in operational allocation models

David Rea
University of Cincinnati
Email: readj@mail.uc.edu

Craig Froehle
University of Cincinnati
Email: craig.froehle@uc.edu

Suzanne Masterson
University of Cincinnati
Email: suzanne.masterson@uc.edu

Brian Stettler
University of Cincinnati
Email: brianastettler@gmail.com

Greg Fermann
University of Cincinnati
Email: fermangj@ucmail.uc.edu

Arthur Pancioli
University of Cincinnati
Email: pancioam@ucmail.uc.edu

ABSTRACT

Fairness is a natural concern when model-based decisions affect human beings. For individuals to perceive decisions as fair, outcomes must reflect two competing concepts: equity and equality. Using a multi-objective, non-linear optimization model, we derive a Pareto frontier of fair solutions that reward employee contributions. The model is implemented in a real-world service environment to allocate employee time across multiple locations. The amount of employee time allocated to locations they did not prefer was substantially reduced. Pre- and post-implementation surveys indicate statistically significant improvements in employee perceptions of fairness, transparency, and overall satisfaction with the work-time allocation process.

KEYWORDS: Field Case Studies, Nonlinear Programming, Ethical Decision Making, Workforce Scheduling, and Individual Preferences

DECISION SCIENCES INSTITUTE

Does leader disability status improve performance outcomes for workers with a disability? an empirical study in the apparel industry

Dustin Cole
Michigan State University
Email: coledust@msu.edu

Sriram Narayanan
Michigan State University
Email: narayanan@broad.msu.edu

Shawnee Vickery
Michigan State University
Email: vickery@broad.msu.edu

ABSTRACT

This research examines the impact of disability status similarity between line leaders and workers on productivity and quality performance in a garment manufacturing environment in which people with disabilities comprise most of the workforce using micro data gathered over multiple years from a pair of garment manufacturing facilities. The results show that the presence of a line leader with disability beneficially impacts team productivity as both the number and ratio of employees with disability increases through mitigation of the productivity decline seen as an increasing number or ratio of workers has a disability in the workgroup.

KEYWORDS: Workforce Diversity, Social Enterprise, Productivity, Manufacturing, Textiles

DECISION SCIENCES INSTITUTE
Contributing Sustainability Factors to Organizational Success in Healthcare

Susan M. L. Zee
Department of Marketing and Supply Chain Management
Southeastern Louisiana University
Hammond, Louisiana 70402
susan.zee@selu.edu

Lillian Y. Fok
Department of Decision Sciences
Western Washington University
Bellingham, Washington 98225
fokl@wwu.edu

Yun-Chen Morgan
Department of Management and Business Administration
Southeastern Louisiana University
Hammond, Louisiana 70402
ymorgan@selu.edu

ABSTRACT

This research examines the relationships between healthcare organizations' performance and 1) green sustainability orientations, 2) organizational culture, 3) Total Quality Management (TQM), and 4) impact of green sustainability initiatives. Surveys completed by 69 full-time employees from the healthcare industry were evaluated. Results showed that healthcare organizations performance are affected by the total number of green initiatives implemented in an organization, the use of traditional quality management programs, and employees' perceptions of positive impact from their organizations' commitment to practice social sustainability.

KEYWORDS: Healthcare; Green orientations; Sustainability, Quality Management; Organizational Culture.

INTRODUCTION

Due to our current pandemic—hospitals, medical clinics, healthcare facilities and beyond—have been thrusted into the spotlight. The state of wellness of hospital systems, current healthcare employees, and the general well-being of all organizations in this industry is something none of us can afford to ignore at this time. In this research, we examine a set of sustainability factors which contribute to healthcare organizations' performance success. In particular, we look at the relationships between organizational success and sustainability orientations, organizational culture, Total Quality Management (TQM), and the impact of sustainability movement. These elements all have great potential to affect the general well-being and performance level of healthcare organizations.

LITERATURE REVIEW

Environmental sustainability initiatives can offer significant financial benefits for healthcare organizations, and these benefits can help hospitals prosper in the long run (Wagner, 2017). According to Kaplan, Sadler, Little, Franz, and Orris (2012), the U.S. healthcare industry could save roughly $15 billion by adopting more sustainable practices. Hospitals can save millions of dollars by being energy-efficient conscious, through waste reduction efforts, and environmentally responsible purchasing. To be environmentally and socially sustainable, healthcare organizations need to evaluate their building infrastructure, organizational practices, and related systems, in order to implement measures in line with their budgets. They need to have reliable and resilient engineering systems to ensure safety for all patients (Biason & Dahl, 2016).

Total Quality Management and Sustainability

TQM has its roots in the integrative approach in customer satisfaction and a company's overall success (Chin, Pun, & Hua, 2001). Applying TQM to the healthcare environment means providing the best possible care through continuously improving service to stakeholders of which exceed the needs and expectations of them (Grossu & Kalkis, 2019). Stakeholders in the health care environment could be the patients, staff, physicians, and the community.

Sustainable development is necessary for companies to address environmental deterioration. Zairi and Liburd (2001) defined sustainability development as the ability of an organization to adapt to change in the business environment in order to deploy the best contemporary methods to achieve and further maintain superior performance. Their take on sustainability development implies competitiveness and an organization's competitiveness is partly dependent on their quality management practices. Sustainability is a new tool in company planning (Beatley & Manning, 1998) and a fundamentally important concept which should influence all policy developments within a firm (Loffler, 1998).

Organizational Culture

Employees of the same organization often share similar values and attitudes, and in turn, behave similarly. Therefore, organizational culture often is an invaluable tool to implement company strategies. When sustainability movement inevitably becomes a part of organizational strategy, the culture that binds members of the organization together also impact the level of success on green sustainability initiatives. Healthcare organizations incorporating these initiatives within part of their organizational culture could reach a higher level of sustainability and profitability (Acharya, Vadher, & Acharya, 2014).

RESEARCH MODEL

In this research, we extend the examination of green movement issues to consider organizational culture, employee perceptions of organizational commitment to the green movement and the relationships which may exist between personal and organizational commitment and the QM Maturity of the organization. Additionally, we examine the factors that may be affecting the organizational performance in healthcare organizations.

Research Question: Healthcare organizations will have different levels of organizational green orientation, organizational culture, organizational performance, and impacts of the green movement.

METHODOLOGY

Subjects in the sample were approximately 69 full-time employees from different healthcare organizations in the South. They were roughly 30.4 % male and 69.6% female with an average age of 42.61 years. These subjects had an average 19.8years of working experience with 8.58 years in management positions. Roughly 52.2% of the subjects are employed in a company that has more than 500 employees, 5.8% of the subjects worked in a company that has 251 to 500 employees, 14.5% of the subjects worked in a company that has 51 to 250 employees, and 27.5% of the subjects worked in a company that has less than 50 employees. Subjects responded to a survey which asked about their perceptions and experiences about the green movement, quality management, and organizational culture in their own firms. In this study, we will concentrate on the relationships among perceptions of support for the organizational green movement, organizational culture, the impact of green movement, QM maturity and organizational performance.

DISCUSSIONS AND CONCLUSIONS

Introducing environmental sustainability measures is often believed to lead to increased operating costs. In this study, we found that it may just be the opposite. With little capital investments, hospitals and other healthcare facilities could realize significant operating savings and better performance. Our results showed that healthcare organizations performance is significantly affected by the total number of green initiatives implemented in an organization, the use of traditional quality management programs, and employees' perceptions of positive impact from their organizations' commitment to practice social sustainability. We believe that healthcare facilities are in a prime position now to make sustainability a part of their mission and long-term strategic plan. This will ensure the organizations' longevity and increased profitability.

REFERENCES

Reference is available via email to the author.

DECISION SCIENCES INSTITUTE
Healthcare Supply Chain Management Strategies against Medical Supply Shortages

Xiaojun (Gene) Shan
University of Houston – Clear Lake
Email: shan@uhcl.edu

ABSTRACT

Supply chain management is rapidly evolving, constantly incorporating advanced technologies. However, there seems a gap between supply chain management in general and healthcare supply chain management in particular. There is an urgent need to bridge this gap in facing the inefficiency in healthcare supply chains, which manifest in a number of ways. A significant one is the shortages in medical supplies. The aim of the paper is to survey the current literature on the causes and mitigation strategies of different types of medical supply shortage and propose future strategies of healthcare supply chain management in combating it.

KEYWORDS: Healthcare supply chain management, Medical supply shortage

INTRODUCTION

Supply shortage is a common problem faced by many different industries on a global scale. It has occurred in a variety of industries, including auto (Hauschild, 2017), health care (E.g., vaccines in Center for Disease Control and Prevention, 2001, medical isotope in Gould, 2009 and Shea & Morgan, 2010, generic injectable agents in Gatesman & Smith, 2011 and Chabner, 2011, EpiPen in Antico, 2018, IV saline bags in Weber, 2018), and mobile phones (Norrman & Jansson, 2004). Besides causing businesses to lose revenues, it could place people's safety and health at danger if the shortage happens to medical supplies (Jaret, 2014). The situation is much worse for developing countries (e.g., Venezuela according to National Public Radio, 2018, and many others according to Gray & Manasse, 2012).

The requirements of a typical supply chain are quite different from those of a supply chain in health care. In particular, the goal of the design of a traditional supply chain is usually to maximize efficiency and minimize costs by reducing inventory and avoiding redundancies whereas in health care making necessary supplies available to patients are prioritized over cost reduction. Inventory management, risk management and production efficiency provide theoretical foundation in preventing medical supply shortage. Specifically, safety stocks are designed to deal with uncertain demand to achieve a certain level of services (in health care, the service is treating patients with medicines, which is supported with medical equipment and supplies). Buffers can be strategically embedded into the supply chain to possibly account for unexpected disruptions. It is imperative to pay attention to senior citizens (Garcia & Bumblauskas, 2017).

This paper aims to understand the causes and possible remedies to medical supply shortage by surveying a rich body of literature. In particular, this is a conceptual paper integrating the broad literature with a relatively unique perspective. Whereas there exist plenty of research studies focusing on particular types of medical products, few comprehensive reviews exist. Given that some policy regulations as mitigation strategies are suggested, which are likely to be broad affecting a range of medical products, it might be useful to compare the root causes and mitigation strategies for different medical products.

The rest of the paper is organized as follows: the next section provides the literature review as the main body of the paper. In particular, we first give a broad overview of research on

healthcare supply chain, then discuss different types of medical products and likely causes of their shortages, current and proposed additional mitigation strategies. Finally, we provide conclusions.

LITERATURE REVIEW

Health Care Supply Chain Research

There is a significant amount of research that has been devoted to the healthcare supply chain. Pishvaee et al. (2014) propose an accelerated Benders decomposition algorithm to solve a sustainable supply chain design problem, which was formulated as a multi-objective probabilistic programming model. Nagurney et al. (2012) also develop a network optimization model to design a complex supply chain for blood. Medical supply shortage was often associated with disaster management (Mete & Zabinsky, 2010). Another topic of interest is pediatric vaccine. Jacobson et al. (2006) used simulation to evaluate different stockpiling levels of pediatric vaccines to cope with potential vaccine shortages. There are some concerns that raw material shortage could impact healthcare services (e.g., Clark & Kline in 1981 discussed the impact of oil shortage on medical life support).

Interestingly, whereas there are significant coverages in the media, not as much supply chain management research in medical supply shortage has been conducted. In a recent review by Volland et al. (2017), they categorize the current supply chain management literature into four groups: (1) supply and procurement, (2) inventory management, (3) distribution and scheduling, and (4) holistic supply chain management. Uthayakumar & Priyan (2013) explore a pharmaceutical supply chain between a pharmaceutical company and a hospital, with a focus on inventory management. Gunpinar & Centeno (2015) develop stochastic integer programing models between a hospital and a blood center, whose solution also reduces blood waste and shortages. However, based on media reports, the main cause of medical supply shortage does not seem at the level of inventory but the manufacturing disruption. Also, they did not consider operations during disasters, which could have a severe impact on the whole nation or even across national boundaries.

To reduce supply chain costs, lean philosophy and six sigma are being adopted by healthcare systems at an increasing speed (Henrique & Godinho, 2020). It is in conflict with risk management in health care supply chain, which aim to make the supply chain resilient to external shocks.

Types of Medical Products

Different types of products are likely to have different underlying causes of shortage and thus we first consider product types. Motivated by commercial applications, Fisher (1997) proposes that two types of products require two different types of supply chains. One is functional and the other is innovative. There could be several causes for supply shortage, mainly due to uncertain demand, and disruption along the supply chain from production to transportation. Fisher (1997) claims that heavy discounts could cause greater variations in demand. However, there might be other types of products demanding new supply chains as in health care. Health care products are typically not subject to discounts, with the exception of over-the-counter medicines. Variation in demand could be due to natural and man-made disasters such as hurricanes, epidemics/pandemics and terrorist attacks. Production or transportation disruption could be due to quality issues and/or natural and man-made disasters. Privett & Gonsalvez (2014) list shortage avoidance as the 6th global health pharmaceutical supply chain challenges.

Whereas several medical supplies could be in shortage simultaneously, their root causes could be fundamentally different and related to the product type. The following is a list of medical products vulnerable to shortages.

Medicines and Vaccines

Their shortages have been observed in a range of countries with various income levels (Iyengar et al., 2016). Canada experienced a supply shortage in medical cannabis (Buckner, 2018). In 2012, there was a shortage of generic injectable medications in Canada (Kaposy, 2013). Zheng et al. (2008) analyze the case where China had a drug shortage. A total of 21 countries experience different levels of medicine shortage, ranging from Afghanistan to Zimbabwe (Gray & Manasee, 2012). Its consequences could have significant clinical implications (Mazer-Amirshahi et al., 2015).

Medical Equipment

Expensive medical equipment such as ventilators and less expensive medical products such as IV saline bags, Personal Protective Equipment (PPE), and generic injectable agents are some examples. A recent shortage of IV saline bags is resulting from the disruption to its manufacturers in Puerto Rico due to Hurricane Maria.

Blood

Blood is a unique healthcare product since the supply is mainly determined by availability of donors and their willingness to donate.

Causes of Shortage

There seems three main causes of medical supply shortage: 1) manufacturing disruptions, 2) inefficiency in logistics, and 3) substandard inventory management. During Hurricanes Harvey, some shelters have extra medical supplies whereas others are in need of them, suggesting some gaps in the logistics process. It appears that better visibility and tracking of medical supplies during a disaster could alleviate this problem. Privett (2015) explores state-of-the-art technologies in enhancing information visibility in humanitarian operations. Bolineni (2016) interviews three supply chain managers in Indian pharmaceutical industries to learn about the supply chain management strategies employed in Indian pharmaceutical supply chain. In the following, causes of different types of medical supplies are discussed. Table 1 summarizes the types of medical supplies that experience shortage, and their causes with references.

Table 1: Types of Medical Supply Shortage, Causes and References

TYPE OF MEDICAL SUPPLY	CAUSE	REFERENCE
Medicine and Vaccines	1) Disruption to manufacturing 2) Disruption to distribution 3) Misalignment of incentive 4) Government disqualification of drug manufacturers 5) Government sanction	1) Hollak et al. (2009); Steinbrook (2009); FDA (2011) 2) Wood et al. (2006); Harris (2014). 3) Bradford (2018) 4) Minion et al. (2010) 5) Kirkpatrick (1996)

Medical Equipment	1) Lack of incentive to stockpile (Ventilators) 2) Disruption to global supply chain (PPEs) 3) Misalignment of incentive (PPEs)	Mehrotra et al. (2020)
Blood	1) Low donation rate 2) Seasonality	

Medicines and Vaccines

Its shortage is frequently caused by disruption to their manufacturing and supply chain. According to the U.S. Food and Drug Administration (FDA) (2011), a total of 43% of 127 drug shortages were due to manufacturing issues. A large number of medicines consumed in the U.S. originate from other countries. For example, India produces 40% of over-the-counter and generic prescription drugs used in the U.S. although there are safety concerns by the U.S. FDA (Harris 2014). A disruption to global supply chain due to natural disasters and others would lead to shortage. There could be a limited number of manufacturers for some other medicines if the profit margin is relatively low. Bradford (2018) suggests that the government or the Medicare and Medicaid system is a main reason behind medical supply shortages besides natural disasters in the U.S.. The government might be responsible for low profit margin for producing generic medicines, which discourages medicine manufacturers to produce them. Other causes of medical supply shortage include government regulations disqualifying some manufacturers (e.g., Streptomycin, a medicine to treat TB, Minion et al., 2010) and government sanctions (e.g., the role of U.S. in medicine shortage in Cuba, Kirkpatrick, 1996).
There is a legitimate reason to examine the problem of medical supply shortage from a global perspective. The coincidence of significant reduction in heroin use in Canada with Australian heroin shortage suggests a global supply chain, which could affect more than one country at the same time (Wood et al., 2006). For some medicines such as radiopharmaceuticals, there are a limited number of facilities in the world manufacturing the product and any disruption to the manufacturing process could lead to potential shortage (Steinbrook, 2009). Another example is imiglucerase, which is used to treat Gaucher disease and solely produced by Genzyme Cooperation (MA, U.S.). A production disruption leads to a global shortage of imiglucerase, putting patients at risk (Hollak et al. 2009).

Medical Equipment

Many healthcare systems observes shortage in ventilators and PPEs such as facial masks, alcohol-based hand rub, gown, goggles and gloves during COVID-19. The reason behind the shortage in ventilators is that health systems with limited resources have little incentive to invest significantly on expensive ventilators, which might not be used unless a large number of patients are infected and hospitalized in a pandemic within a short time period. Considering the low probability of such extreme events, there is no surprise to see a shortage in ventilators. Take COVID-19 as an example, although the first case was reported in Wuhan, China on December 31, 2019 and other countries than China might have more time to prepare, there are countries with a limited number of COVID-19 cases as of June 9, 2020, who might not experience a shortage of ventilators. Therefore, it might not be cost-effective to stockpile ventilators at individual health system level. Besides, it is difficult to ramp up production capabilities to the level that meet the rapidly increasing needs. U.S. President Donald Trump has invoked the Defense Production Act to order General Motors to manufacture ventilators. However, this conversion still takes time. By efficiently allocating limited supplies of ventilators, it

is possible to satisfy the uncertain needs according to stochastic modeling with resource sharing (Mehrotra et al., 2020). There is a need to have adequate medical staff to operate the ventilators. From health systems' perspective, one lesson learned from COVID-19 would be to cross-train medical staff. In response to personnel shortage, a large number of volunteers and retired medical personnel has come to the frontline to fight COVID-19. Given the furlough and layoffs of healthcare workers due to cancelled elective surgeries during the lockdown, it might be beneficial to cross-train the medical staff so that they could be deployed to other positions. In contrast, PPE is relatively inexpensive but still experiences significant shortage. While PPE shortage is also caused by a mismatch in the supply and demand, its supply shortage is mainly due to disruption to the global supply chain besides surging demand. There are a limited number of manufacturers for facial masks in developed countries due to low profit margin. As a result, many manufacturing activities have been outsourced to overseas countries such as Asian countries.

Blood

The root causes is the low donation rate, which is exaggerated by natural and/or man-made disasters. Also, blood shortage seems seasonal and happens more frequently during holidays. Disasters could increase the demand and decrease the supply at the same time. In countries with voluntary blood donation programs, a blood shortage occurs frequently.

Current Mitigation Strategies

In general, there are a number of ways to mitigate supply shortage, including pricing, rationing, queueing, and interruption. Woo (1994) develops a common framework to compare these strategies in managing water supply shortage. Chopra and Sodhi (2004) provide a comprehensive review of causes of supply-chain disruption and potential remedies. A number of the strategies might not be directly applied to healthcare supply chain. For example, one way to curtail the demand is to raise the price. However, there might be some ethical issues and it also gives medicine manufacturers a legitimate reason to raise prices in the case of medicine shortage. Therefore, there is little incentive for them to increase production capabilities unless there are some government regulations.
Some other mitigation strategies could include improved classification of products, suppliers, markets, customers, enhanced agility and flexibility, well-developed key performance indexes, and coordination across the healthcare supply chain (Ebel et al., 2013). In particular, the healthcare logistics management could apply ABC inventory classification to their portfolio of products (Jacobs and Chase, 2017). For A items, which are critical or lifesaving medicine/devices, they might build redundancy into the supply chain so that any single event is extremely unlikely to cause wide-spread shortages, which could last for a prolonged period. Products labelled as B and C probably have viable alternatives or substitutes and could be delayed in their employment. Simpson (2008) proposes that producing more might not be the best solution to the global shortage of antisnake venom and should be used together with waste reduction. This logic might apply to other medical supplies as well. Suppliers could contribute by increasing data transparency and thus the efficiency of the healthcare supply chains. Singleston et al. (2013) develop a framework to plan for drug supply shortage. U.S. FDA is mandating drug manufacturers to report anticipated drug shortage and publishes the list of drugs in shortage on its website and its mobile app. In addition, U.S. FDA extends the use date of some of those products since at least some drugs can be used beyond their expiration dates (Lyon et al., 2006). Ventola (2011) provides a list of causes and mitigation strategies to tackle the drug shortage in the U.S. They propose that the main cause is lack of advanced notice of shortage. However, U.S. FDA has implemented such mandates and has not solved the drug shortage

problem. This suggests that there might exist other causes. The mitigation strategies during a global shortage could differ. Emergency production could temporarily increase production outputs of some critical medical supplies. In the U.S., a number of manufacturers are altering their production lines to produce ventilators and PPEs. To meet domestic demands, U.S. President Donald Trump has blocked facial mask exports by 3M and order it to increase manufacturing outputs with the Defense Production Act.

Innovation and improvising also play important roles in combating medical supply shortage. Convissar et al. (2020) propose to use some widely available medical supplies to make N95 facial masks. Some modifications to ventilators have been proposed so that one machine can be used to treat multiple patients simultaneously (Kumar and Kumar, 2020).

Using alternatives or substitutes seem a useful strategy when one particular type of health care supplies is in shortage and there are safe alternatives. Lyra et al. (2011) suggest using an alternative production method for molybdenum-99 (important for medical diagnosis). According to IMS Institute for Healthcare Informatics (2011), two-thirds of the supply shortage have only three or fewer suppliers. Which they did not report but might be interesting is how likely they are affected by the same event such as a hurricane. For example, most of the manufacturing facilities of IV saline bags are located in Puerto Rico, and were all affected by Hurricane Maria leading to a nationwide shortage of IV saline bags for a prolonged period in the U.S.. Fortunately, the healthcare providers were innovative and used some substitutes for IV saline bags such as Gatorade (a sports drink) to minimize the impact on patient care. Similarly, during a propofol shortage, dexmedetomidine was a viable substitute (Thoma et al., 2014). However, what if the medical supplies in discussion do not have safe alternatives? Therefore, it is imperative for healthcare supply chain professionals to proactively develop other mitigation strategies in advance.

A more efficiency use and optimal inventory control are the strategies to mitigate blood shortage. There has been significant research efforts attempting to find a suitable synthesized substitute to blood. Unfortunately, they carry a significant negative effect even though some have been used in practice (Solomon, 2012). Economic incentives have been used to encourage voluntary blood donation. Alternatively, making donors aware of blood shortage is effective in the short term but not in the long term. A family replacement program, where patients in need could recruit their family and friends to donate, is marginally effective in the short run but discourage first-time donors (Sun et al., 2016). Some healthcare supply chain researchers have focused on optimal inventory control to increase the efficiency in blood usage (Wang and Ma, 2015). To cope with blood shortage, practitioners have been innovative in finding solutions to this unavoidable problem (McCarthy, 2007). Some of the mitigation strategies mentioned in McCarthy (2007) root from supply chain management strategies such as having more than one supplier of blood, and rationing and prioritizing patients, who are in need of blood.

Potential Mitigation Strategies

There are several methods to further improve the healthcare supply chain. First, health system management could have a better understanding and thus forecasting of the demand. For example, the management from individual health systems might be able to estimate how many PPEs are required to care for an average patient based on historical data and report such data to local health authorities. Public health departments for each county and state could compare these data from different health systems accounting for the difference in the number and severity of patients and make recommendations about PPE usage. Local governments could work with federal governments, non-profit organizations and private companies to strategically allocate limited PPE resources. Whereas healthcare workers in the front line of fighting contagious diseases should be provided with adequate and sufficient PPEs, vulnerable

populations would also need adequate protections to be prevented from burdening the healthcare systems.

Second, local governments could assess different industries in terms of vulnerability to contagious diseases. In addition, health system management might be prepared for unexpected demand increases due to epidemics/pandemics such as the need for dialysis of some COVID-19 patients. Social distancing might be easier to implement in some production processes than others. Whereas manufacturers would be mainly responsible for workers' safety, local governments could be aware of the potential risk and the care need. Whereas the demand for some medicines might be difficult to control but relatively easy to forecast, the demand for care needs could be relatively easier to control. For example, strictly enforcing social distancing would allow local governments to prevent potential virus outbreaks.

Third, one potential mitigation strategy is for the government to subsidize the manufacturing of some critical medicines. The government subsidized manufacturers in a number of categories including auto (Rothman, 2017), electronics and others (DVIRC, 2012). As a result, there would be more manufacturers in the market, reducing the probability of manufacturing disruption (simultaneous disruption to all manufacturing processes are much more unlikely than disruption to a limited number of manufacturers located physically close to each other).

Supply chains frequently affect other management disciplines such as information technology (IT), finance, sales and marketing. Healthcare supply chains are no exception. As one example, significant advances in supply chain management usually come together with advances in IT. Radio-frequency identification (RFID) in supply chains might not be widely implemented if the IT infrastructures were not developed. Some future mitigation strategies against medical supply shortage might be related to IT solutions.

Finally, the general public also plays a critical role in the combat against medical supply shortage. For instance, right after September 11, 2001 terrorist attacks, a large number of people volunteered in donating blood due to strong patriotism and awareness of a probable significant blood need. By coping with the supply side issue, this behavior has prevented any blood shortages but resulted in significant wastes, which suggests that a better coordination mechanism is needed. In contrast, by complying with stay-home orders and social distancing policies, the public could also assist with alleviating the demand by preventing oneself from being sick, possibly hospitalized and burdening the health system.

CONCLUSION

Medical supply shortage poses a significant threat to public safety and might be alleviated with some improvements in healthcare supply chain. In general, there are three distinct types of medical supplies vulnerable to shortage: 1) medicines and vaccines, 2) medical equipment such as ventilators and Personal Protection Equipment (PPE), and 3) blood. There seems to be three main root causes: 1) manufacturing disruption, 2) logistics disruption and 3) surging demands. The mitigation strategies would involve all stakeholders including the government, healthcare providers, manufacturers, suppliers, non-profit organizations and the general public. The government could create central mechanisms to stockpile strategically critical medical supplies such as ventilators and PPEs, subsidize manufacturers of critical medical supplies, and ration and prioritize medical supplies according to needs and impacts on the public health. Healthcare providers could increase the efficiency in using medical supplies and strategically allocate the limited resources. Manufacturers could diligently report expected disruption to the government and invest in flexible manufacturing and quality programs so that it becomes easier to increase production capacity if needed and reduce defects. Suppliers could work with manufacturers and provide transparent information to the authority or the public about the medical supplies. In addition, healthcare supply chain personnel could invest in the information technology in healthcare supply chains to provide better tracking and verify medical supplies to ensure timely

delivery and minimize counterfeit products. The non-profit organization could facilitate the production and distribution of medical supplies by providing finance resources to increase the manufacturing capacity and strengthen supply chains and aid vulnerable groups (e.g., the elderly) to reduce demand for health care. Finally, the general public could comply with governmental orders or recommendations to avoid becoming burdens on the health care system

REFERENCES

Antico, P. (2018). Mylan confirms EpiPen "supply disruption" (i.e. shortage) and more. Retrieved from AllergyEats.com https://www.allergyeats.com/mylan-confirms-epipen-supply-disruption-i-e-shortage-and-more/, December 2018.

Abutaleb, Y. (2018). U.S. healthcare spending to climb 5.3 percent in 2018: agency, Health News, February 14, 2018. Retrieved from https://www.reuters.com/article/us-usa-healthcare-spending-idUSKCN1FY2ZD, December 2018.

Bolineni, P. (2016). *The Indian pharmaceutical industry's supply chain management strategies*, PhD dissertation, Walden University.

Bradford, J. C. (2018). Medical equipment shortages compromise patient care, Me&My Doctor, Texas Medical Association, October 12. Retrieved from http://www.meandmydoctor.com/2018/10/medical-equipment-shortages-compromise.html, December 2018.

Buckner, D. (2018). Medical cannabis users anxious over supply shortages. Retrieved from https://www.cbc.ca/news/business/medical-pot-shortages-1.4887318, December 2018.

Burns, L. (2001). *The healthcare value chain*. Jossey-Bass, New York, NY.

Center for Disease Control and Prevention (2001). Update on the supply of tetanus and diphtheria toxoids and of diphtheria and tetanus toxoids and acellular pertussis vaccine. Morb Mortal 50(10):189–190

Chabner, B. A. (2011). Drug shortages–a critical challenge for the generic-drug market. *New England Journal of Medicine*, 365: 2147-2049 doi: 10.1056/NEJMp1112633 pmid: 22040167.

Chopra, S. & Sodhi ,M. S. (Fall 2004). Managing risk to avoid supply-chain breakdown, *MIT Sloan Management Review*, 53-61.

Clark, G. B. & Kline, B. (1981). Impact of oil shortage on plastic medical supplies, *Public Health Reports*, 96(2), 111.

Clark, D. (2020) Trump invokes Defense Production Act to force GM to make ventilators for coronavirus fight. NBCNews, Retrieved from https://www.nbcnews.com/politics/donald-trump/trump-invokes-defense-production-act-force-gm-make-ventilators-coronavirus-n1170746, June 2020.

Convissar, D., Berra, L., Chang, M. G., & Bittner, E. A. (2020). Personal Protective Equipment N95 Facemask Shortage Quick Fix: The Modified Airway From VEntilatoR Circuit *(MAVerIC). Cureus, 12(5)*, e7914.

DVIRC, (2012). U.S. manufacturing benefiting from government subsidies, May 11, Retrieved from http://www.dvirc.org/u-s-manufacturing-benefiting-from-government-subsidies/, December 2018.

Ebel, T., Larsen, E. and Shah, K., (2013). Strengthening health care's supply chain; A five-step plan, McKinsey&Company, Retrieved from https://www.mckinsey.com/industries/healthcare-

systems-and-services/our-insights/strengthening-health-cares-supply-chain-a-five-step-plan, December 2018.

Fink, A. (1998). *Conducting research literature review: From paper to the Inter*net, Sage, Thousand Oaks.

Garcia, M. & Bumblauskas, D. (Nov. 2017) Creating a culture of continuous improvement in senior services, Decision Science Institute (DSI) Annual Conference, Washington, DC.

Gatesman, M. L., & Smith, T. J. (2011). The shortage of essential chemotherapy drugs in the United States. *New England Journal of Medicine*, 365: 1653-1655 doi: 10.1056/NEJMp1109772 pmid: 22040130.

Gould, P. (2009). Medical isotope shortage reaches crisis level; robust solutions sought urgently to shore up fragile supply chain. *Nature*, *460*(7253), 312-313. Retrieved from https://link.gale.com/apps/doc/A204482868/HRCA?u=txshracd2589&sid=HRCA&xid=63743cfb, December 2018.

Gray, A. & Manasse, H., (2012). Shortages of medicines: A complex global challenge. Bulletin *of the World Health Organization*, Retrieved from https://www.who.int/bulletin/volumes/90/3/11-101303/en/, December 2018.

Gunpinar, S. & Centeno, G., (2015). Stochastic integer programming models for reducing wastages and shortages of blood products at hospitals. *Computers & Operations Research*, 54, 129-141.

Harris, G. (2014). Medicines made in India set off safety worries, the New York Time, February 14, Retrieved from https://www.nytimes.com/2014/02/15/world/asia/medicines-made-in-india-set-off-safety-worries.html, December 2018.

Hauschild, M. (2017). Supply shortage sees BMW production plummet. Retrieved from https://www.themanufacturer.com/articles/supply-shortage-sees-bmw-production-plummet/, December 2018.

Henrique, D. B., & Godinho Filho, M. (2020). A systematic literature review of empirical research in Lean and Six Sigma in healthcare. *Total Quality Management & Business Excellence*, 31(3-4), 429-449.

Hollak, C. E., vom Dahl, S., Aerts, J. M., Belmatoug, N., Bembi, B., Cohen, Y., Collin-Histed, T., Deegan, P., van Dussen, L., Giraldo, P., & Mengel, E., (2010). Force majeure: therapeutic measures in response to restricted supply of imiglucerase (Cerezyme) for patients with Gaucher disease. *Blood Cells, Molecules, and Diseases*, *44*(1), pp.41-47.

IMS Institute for Healthcare Informatics, (2011). *Drug shortages: a closer look at products, suppliers and volume volatility.* Danbury CT. Retrieved from http://www.imshealth.com/vgn-exttemplating/v/index.jsp?vgnextoid=4737087078483310VgnVCM100000ed152ca2RCRD, December 2018.

Jacobson, S. H., Sewell, E. C., & Proano, R. A., (2006). An analysis of the pediatric vaccine supply shortage problem. *Health Care Management Science*, 9(4), 371-389.

Jaret, P. (2014). Dangerous drug shortages, October, Retrieved from https://www.aarp.org/health/drugs-supplements/info-2014/prescription-medication-shortage.html, December 2018.

Kaposy, C. (2014). Drugs, money, and power: the Canadian drug shortage. *Journal of Bioethical Inquiry*, 11(1), 85-89.

Kirkpatrick, A. F. (1996). Role of the USA in shortage of food and medicine in Cuba. *The Lancet*, 348(9040),1489-1491.

Kohler, J. (2001). Prices jump at gasoline pumps in the area, St. Louis Post-Dispatch, Aug. 29, sec. A, 1.L

Kumar, P., & Kumar, M. (2020). Management of potential ventilator shortage in India in view of on-going COVID-19 pandemic. *Indian Journal of Anaesthesia*, 64(14), 151.

LaPointe, J. (2017). 78% of hospital staff still face manual supply chain management, RevcycleIntelligence, February 16, 2017. Retrieved from https://revcycleintelligence.com/news/78-of-hospital-staff-still-face-manual-supply-chain-management, December 2018.

Lee, H. L., & Whang, S., (March/April 2001). Demand chain excellence: a tale of two retailers, *Supply Chain Management Review*, 5/3: 40-46. L

Lyon, R. C., Taylor, J. S., Porter, D. A., Prasanna, H. R., & Hussain, A. S. (2006). Stability profiles of drug products extended beyond labeled expiration dates. *Journal of Pharmaceutical Sciences*, 95(7), pp.1549-1560.

Lyra, M., Charalambatou, R., Roussou, E., Fytros, S., & Baka, I. (2011). Alternative production methods to face global molybdenum-99 supply shortage, *Hellenic Journal of Nuclear Medicine*, 14(1): 49-55.

Iyengar, S., Hedman, L., Forte, G., & Hill, S. (2016). Medicine shortages: a commentary on causes and mitigation strategies. *BMC Medicine*, 14(1), 124.

McCarthy, L. J. (2007). How do I manage a blood shortage in a transfusion service? *Transfusion*, *47*(5), 760-762.

Marshall, A. (2017). What hospitals waste. Propublic.org Medicine shortages: a commentary on causes and mitigation strategies. Retrieved from https://www.propublica.org/article/what-hospitals-waste, December 2018.

Mazer-Amirshahi, M., Hawley, K. L., Zocchi, M., Fox, E., Pines, J. M., & Nelson, L. S. (2015). Drug shortages: implications for medical toxicology. *Clinical Toxicology*, 53(6), 519-524.

Mehrotra, S., Rahimian, H., Barah, M., Luo, F., & Schantz, K. (2020). A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. *Naval Research Logistics*, in press.

Mete, H. O., & Zabinsky, Z. B., (2010). Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics,* 126(1), 76-84.

Minion, J., Leung, E., Menzies, D., & Pai, M. (2010). Microscopic-observation drug susceptibility and thin layer agar assays for the detection of drug resistant tuberculosis: a systematic review and meta-analysis, *The Lancet Infectious Diseases*, 10(10), 688-698.

Nagurney, A., Masoumi, A. H., & Yu, M. (2012). Supply chain network operations management of a blood banking system, *Computational Management Science*, 9(2), 205-231.

Norrman, A., & Jansson, U. (2004). Ericsson's proactive supply chain risk management approach after a serious sub-supplier accident, *International Journal of Physical Distribution & Logistics Management* 34(5), 434-456.

NPR. (2018). Venezuela's health care system ready to collapse amid economic crisis, February 1, Retrieved from https://www.npr.org/2018/02/01/582469305/venezuelas-health-care-system-ready-to-collapse-amid-economic-crisis, December 2018.

Privett, N. (2016). Information Visibility in Humanitarian Operations: Current State-of-the-Art. In: Zobel C., Altay N., Haselkorn M. (eds) *Advances in Managing Humanitarian Operations. International Series in Operations Research & Management Science*. Springer.

Rothman, J. (2017). Should we subsidize manufacturing? August 14, Retrieved from https://www.newyorker.com/news/news-desk/should-we-subsidize-manufacturing, December 2018.

Sanborn, B. J. (2018). How hospitals are grappling with medical supply shortages caused by Hurricane Maria. *Healthcare Finance.* Retrieved from https://www.healthcarefinancenews.com/news/how-hospitals-are-grappling-medical-supply-shortages-caused-hurricane-maria, December 2018.

Shea, D. A. and Morgan, D., (September 2010). *The helium-3 shortage: Supply, demand, and options for Congress*. Congressional Research Service, Library of Congress.

Simpson, I. D., (2008). The "worldwide shortage" of antisnake venom: is the only right answer "produce more" or is it also "use it smarter?" *Wilderness & Environmental Medicine*, 19(2), 99-107.

Singleton, R., Chubbs, K., Flynn, J., Kaposy, C., Peckham, G., Penney, J. and Pullman, D., (April 2013). From framework to the frontline: designing a structure and process for drug supply shortage planning. In Healthcare Management Forum (Vol. 26, No. 1, pp. 41-45). Sage CA: Los Angeles, CA: SAGE Publications.

Solomon, G. D. (2012). Segmentation and communications to solve the blood shortage: An exploration of the problem with recommendations. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 23(2), 415-433.

Steinbrook, R. (2009). Drug shortages and public health. New England Journal of Medicine, 365(361), 1525-1527 doi: 10.1056/NEJMp0906922 pmid: 19828529.

Sun, T., Lu, S. F., & Jin, G. Z. (2016). Solving shortage in a priceless market: Insights from blood donation. *Journal of Health Economics*, 48, 149-165.

Thoma, B. N., Li, J., McDaniel, C. M., Wordell, C. J., Cavarocchi, N. and Pizzi, L. T., (2014). Clinical and economic impact of substituting dexmedetomidine for propofol due to a US drug shortage: examination of coronary artery bypass graft patients at an urban medical centre. *Pharmacoeconomics*, 32(2), 149-157.

Volland, J., Fügener, A., Schoenfelder, J., Brunner, J. O., (2017). Material logistics in hospitals: A literature review, *Omega*, 69, 82-101.

Webster, L. (2018). Hurricane Maria's effect on the healthcare industry is threatening lives accress the U.S., *Huffpost*, September 20, Retrieved from https://www.huffingtonpost.com/entry/iv-bag-drug-shortage-puerto-rico-hurricane-maria_us_5ba1ca16e4b046313fc07a8b, December 2018.

United States Food and Drug Administration, (2011). *A review of FDA's approach to medical product shortages.* Silver Spring. Retrieved from http://www.fda.gov/downloads/aboutfda/reportsmanualsforms/reports/ucm277755.pdf, December 2018.

Ventola, C. L., (2011). The drug shortage crisis in the United States: causes, impact, and management strategies. Pharmacy and Therapeutics, 36(11), 740-757.

Wang, K. M., & Ma, Z. J. (2015). Age-based policy for blood transshipment during blood shortage. *Transportation Research Part E: Logistics and Transportation Revie*w, 80, 166-183.

Woo, C-K. (1994). Managing water supply shortage interruption vs. pricing, *Journal of Public Economics*, 54: 145-160.

Wood, E., Stoltz, J.-A., Li, K., Montaner, J. S. G., & Kerr, T. (2006). Changes in Canadian heroin supply coinciding with the Australian heroin shortage, *Addiction*, 101(5), 689-695.

Zheng, W., Yao, C., & Zheng, W. (2008). Analysis of the phenomenon of the shortage of the zero-profit drug supply in the CHS facility, *Chinese Health Economics*, 7.

DECISION SCIENCES INSTITUTE
Leaderboard or Digital Badges?
Improving Engagement in Online Discussions

Brian S. Bovee
Dakota State University
Email: Brian.bovee@trojans.dsu.edu

Thomas Jernejcic
Dakota State University
Email: Brian.bovee@trojans.dsu.edu

Dr. Omar El-Gayar
Dakota State University
Email: omar.el-gayar@dsu.edu

ABSTRACT

Engaging students in online discussions is difficult. Gamification increases engagement in non-game settings such as online classrooms by adding game design elements. Since game elements such as leaderboards and digital badges are typically used together, there remains a lack of awareness on the effectiveness of each element. Thus, the aim in the present study is to compare leaderboards and badges for improving engagement in online discussions. Overall, digital badges were more effective in increasing behavioral engagement when compared to a leaderboard. The research provides insights into the efficacy of gamification and alternatives to address low student engagement in online discussions.

KEYWORDS: Gamification, Engagement, Online discussion, Leaderboard, Badges

INTRODUCTION

Online discussions are critical to online education as a means for student collaboration, enhancing learning and supporting intrinsic motivation to succeed. Engagement in online discussions, however, is difficult to maintain (Cheung & Hew, 2005; Hara, Bonk, & Angeli, 2000; Hewitt, 2005). Gamification, defined as the use of game design elements for non-game applications (Deterding et al, 2011), has been proposed as a method that can promote learning and engagement (Kapp, 2012, Hamari et al, 2014). Thus, while the research supporting gamification and learning is prevalent, studies examining the gamification of online discussions are sparse (Ding et al, 2017; Ding et al, 2018; Ding, 2019).

Moreover, there remains a lack of awareness on the effectiveness of each particular form of gamification as the majority of research in gamification has employed multiple gamification methods while using a single measurement of engagement (Looyestn et al, 2017). Accordingly, the aim in the present study is to address this gap by examining two game design elements for possible differences in their effect on two dimensions of engagement with an online discussion. Consistent with Fredricks et al. (2004) established model for engagement, the goals of this study are to separately evaluate two game elements, leaderboards and digital badges, for their potential in evaluating behavioral and cognitive engagement. Toward this goal, each design

element was evaluated in an online, undergraduate information systems course in which all students were exposed to a single gamification element each week, either the leaderboard or digital badges. The independent variable was the use of the different game design elements: badges vs. leaderboard. The dependent variable is two forms of engagement: Behavioral engagement, (measured through posts and replies in online discussions) and cognitive engagement (measured through student grades in the discussion forums).

The current study contributes to the extant literature in at least three ways. First, both leaderboards and digital badges are compared as game elements for their respective ability to promote user engagement in online discussions. Secondly, from a theoretical perspective, the research provides insights into the role of gamification of online discussions in inducing active engagement and increased cognitive performance. Finally, from a practical perspective, the research provides insights into gamification alternatives to address low student engagement in asynchronous online discussions reflected in both the quality and quantity of student posts.

The paper is structured as follows: in the following section we give a brief overview of the related work and we conclude this section with our hypotheses. Next, we describe the methodology followed by the results obtained and a discussion of key findings and insights. Finally, conclusions are presented and opportunities for future research are highlighted.

LITERATURE REVIEW AND HYPOTHESES

Gamification

Gamification involves the use of game design elements for non-game applications (Deterding et al, 2011). While “serious” games are designed for a purpose other than pure entertainment, gamification involves the application of game-like elements such as leaderboards, digital badges, and point systems to increase engagement and tap the learner’s normal drive for achievement (Deterding et al, 2011; Cronk, 2012). Findings in the current literature regarding the use of gamification to promote learning are mixed. The majority of the research demonstrates the ability of gamification to promote intrinsic motivation to learn (Leong & Luo, 2011; Ding, Kim & Orey, 2017) and encourage participation (Barata, Gama, Jorge, Gonçalves, & Fonseca, 2013). However, a few studies have found gamification to result in lower intrinsic motivation and satisfaction compared to the non-gamified groups (Hanus & Fox, 2015). Moreover, despite the success of gamification, research has identified key barriers that prevent teachers from adopting gamification in the classroom such as available resources, instructor/student readiness, game flexibility, and usefulness (Baek, 2008). Instructors must carefully consider each gamification method to ensure there is adequate time and effort to implement a successful gamification pedagogy.

Typical methods of gamification include features such as leaderboards, digital badges, points, rewards, and stories (Looyestyn et al, 2017). Regardless of the methodology used, the goal of all forms of gamification is to have a positive effect on the user’s motivation to engage with the system (Deterding et. al., 2011). However, the specific gamification feature used to increase motivation has not been conclusive. Most studies to date have examined combinations of gamification features and, therefore, further research is needed detailing the benefits of each form of gamification (Looyestyn et al, 2017). Leaderboards and digital badges are currently among the most popular elements of gamification (Looyestyn et al, 2017) and were, thus, selected as the focus for this study.

Leaderboards

Origins of leaderboards date back, as with badges, to online gaming in which a user's performance would be placed in relationship to other players who played the game. Within education, leaderboards rank players according to their relative success in achieving a task related to the learning objectives (Butler, 2013). It is in this context that leaderboards are said to instill a sense of competition into the learning environment in which the user's performance in completing the task is placed in relation to the performance of others (Butler, 2013; Garcia, 2013). Preliminary evidence has shown that this social comparison amongst peers, combined with the tangible nature of leaderboards as compared to digital badges, makes the leaderboard a particularly effective game element (Loooyestn et al, 2017).

However, like digital badges, research on the ability for leaderboards to instill engagement with the information system has been mixed. The majority of the research supports the notion that competition caused by leaderboards creates social pressure to increase the user's engagement and, therefore, learning (Butler, 2013. On the other hand, there is also research showing the presence of a leaderboard caused low scoring users to actually be de-motivated to engage with the information system (Glover, 2013; Hanus & Fox. 2015). Werbach and Hunter (2015) expanded on these findings demonstrating that leaderboards can be effective motivators for increased engagement with the software if there are only a few points left to the next level.

Digital badges

Digital badges provide a way to represent an accomplishment that is visual, available online and contains information related to the meaning of an activity completed by the user (Gibson et al, 2015). The origins of badges can be traced back to the medieval ages in which knights wore badges as a way to display political allegiance and/or achievements. More recently, trophies and badges such as those worn by the Boy Scouts Association became a modern-day representation of achievement (Gibson et al, 2015). With the advent of the Internet, digital badges allowed users to designate achievements in online gaming and encouraged a positive reputation in systems such as those used by Ebay and AirBnb.

In order to achieve a digital badge the user can engage in a variety of activities such as engaging in online tasks, reaching performance milestones, and/or completing an assessment. Users have displayed badges on social websites as a means to increase credibility and reputation in regards to the subject matter. Within education, badges have traditionally been used to promote three activities: a) to promote engagement with the information system in a manner that promotes learning, b) to act as digital "breadcrumbs" documenting the location of the user within the entire storyline of completing the course, and c) to document specific learning achievements. First, digital badges are designed to promote the user to engage with the software in order to complete activities that will help improve learning goals (Gibson et al, 2015; Stefaniak & Carey 2019). Second, digital badges act as "digital breadcrumbs" documenting the location of the user's journey through the entire "storyline" of the information system. Within the online gaming world, digital badges "unlock" levels to previously restricted content. Similarly, within education, digital badges act as a credential signifying the current status of the user in regards to progression towards an overall educational goal (Stefaniak & Carey 2019). Examples include passing a single module in an entire course or acquiring a particular skill in which many others may still need to be completed. Third, badges provide a way to communicate evidence of learning to other students, employers, social networks, or any other relevant stakeholder (Gibson et. al., 2015).

However, the research on digital badges has not been entirely positive. There is growing evidence to support the notion that the extrinsic reward provided by badges tend to wear off after a short novelty period (Van Kleek et al, 2013; Nakajima & Lehdonvirta, 2013; Hanus & Fox, 2015). A popular example of this is the app 'Four-square' which experienced a large reduction in engagement six to twelve months after its initial implementation ("Foursqure", n.d.). In addition, digital badges have been viewed as are more arbitrary than other game elements and, therefore, can lack meaning resulting in users less motivated to engage with the information system (Xu & Buhalis, 2013).

Engagement in online discussions

Asynchronous online discussions represent a critical aspect of the online learning process. Low student engagement, in the form of low quantity and quality of posts, has represented a significant challenge to overcome for instructors (Hara, Bonk, & Angeli, 2000; Hewitt, 2005). While engagement has been defined in many ways, Fredericks et. al (2004) widely accepted (da Rocha et al, 2016; Finn & Zimmer, 2012) model of engagement focuses on examining the three elements of how students feel, behave, and think. Accordingly, in the present study, engagement in online discussions was evaluated based on behavioral (active participation) and cognitive engagement (Fredricks et al, 2004).

Active participation in online discussions involves behavioral aspects of composing original posts, reading or viewing other posts, and responding to other student posts. Game design elements such as leaderboards or digital badges provide external motivation to engage with discussions via constructive competition toward a goal (Lo & Hew, 2018; Ding, 2019). As the user engages with the game, the motivation to engage can shift from extrinsic to intrinsic (Ryan et al, 1991; Lepper, 1988; Deci, Eghrari, Patrick, & Leone, 1994). Self Determination Theory (Deci & Ryan, 2008) and Flow Theory (Csikszentmihalyi, 1998) describe this as a process in which one identifies with an activity's value and integrates it into their sense of self. In applying flow theory to gamification of asynchronous discussions, students are more likely to be motivated to engage with the discussion by clear goals (Locke & Latham, 2006), challenging content, and appropriate feedback. When expectations are not set or vague, students struggle with both the amount and type of content in posts (Dennen, 2007).

In order for learning to occur, active participation must be accompanied by cognitive thinking in the discussion. Active participation, alone, will not result in the desired outcome of learning. Common strategies for developing cognitive thinking in discussions include instructor scaffolding (Zhu, 2006), clear guidelines (Ng, Cheung & Hew, 2009), and critical thinking questions (Garrison, 2005). Game elements can provide the intrinsic motivation for students to critique, construct, and comprehend knowledge in the context of online discussions (Ding, 2019). For example, in recent a study using digital badges, more students in the gamified group reported the online discussions forced them to think harder than students in the non-gamified group (Ding, 2019).

Based on the aforementioned detailed literature review, there remains a lack of awareness on the effectiveness of leaderboards as compared to digital badges due to the fact the majority of research in gamification has employed multiple gamification methods while using a single measurement of engagement (Looyestn et al, 2017). Accordingly, the aim of the present study

is to address this gap by comparing two game design elements, leaderboards and digital badges, for possible differences in their effect on two dimensions of engagement with an asynchronous online discussion board.

Following are the two hypotheses that guided the study:

H1: Leaderboards and digital badges will demonstrate no significant differences in boosting behavioral engagement in asynchronous online discussions.

H2: Leaderboards and digital badges will demonstrate no significant differences in boosting cognitive engagement in asynchronous online discussions.

METHODOLOGY

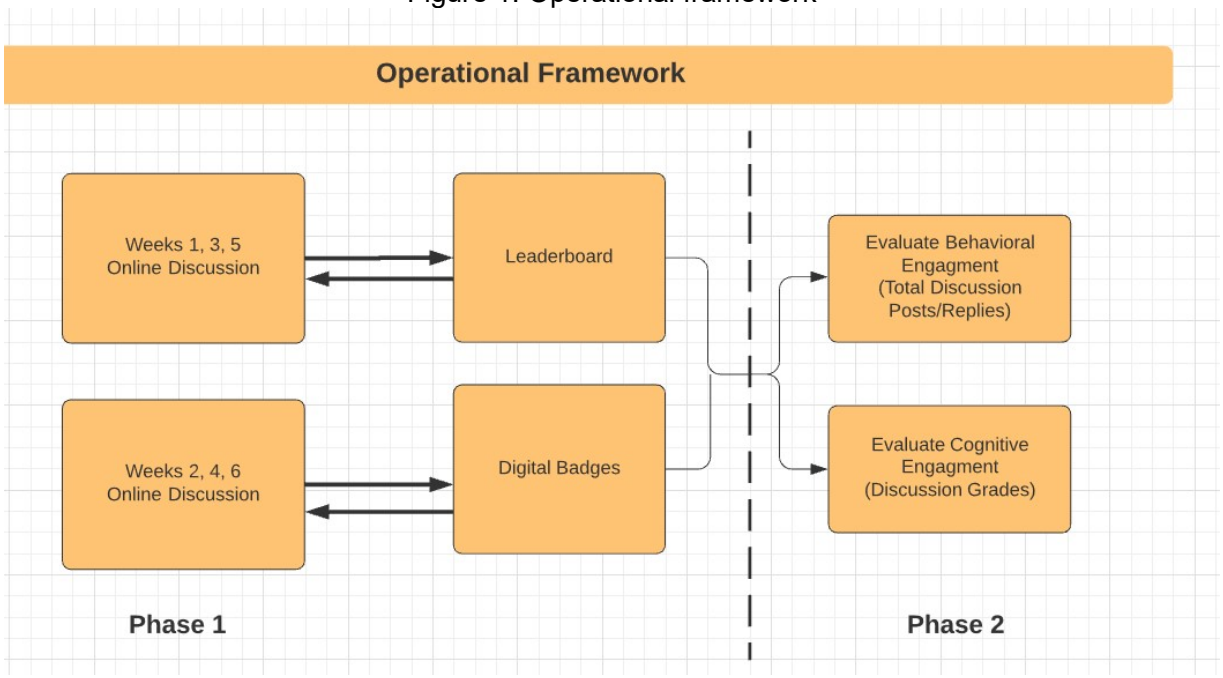
Subjects

Twenty-nine undergraduate students taking an online information systems course at a large regional university were chosen as participants. Students in the course were required to record a minimum of one video discussion post and one video reply using their webcam or cell phone.

Experimental setup

Figure 1 shows the operational framework for this study which used a two phased approach. First, an online discussion board was gamified using both a leaderboard and digital badges. Second, the leaderboard and digital badges were each individually assessed for behavioral and cognitive engagement.

Figure 1: Operational framework



Two game elements, a leaderboard and digital badges, were added to an online discussion board in alternating weeks. In order to examine each game element separately for engagement, only one game element was used each week. During weeks 1, 3, and 5 subjects used the leaderboard. During weeks 2, 4, and 6 subjects used digital badges.

Subjects were informed at the start of the course and reminded weekly of two goals:
1) Achieve a place on the leaderboard game element by placing in the top 20% of total posts/replies on weeks 1, 3 and 5, or
2) Unlock a total of three digital badges during weeks 2, 4, and 6.

This design ensured all students were exposed to each game element. In addition, multiple game elements were not used together on any given week and could, therefore, be analyzed separately. During weeks 1, 3, and 5 only the leaderboard was used as the game element in the discussion board. Subjects were encouraged in these weeks to "lead the discussion" in terms of total posts/replies. Students in the top 20% of the class were then recognized each week on the leaderboard. See Figure 2. The remaining students, and their associated ranking, were not shown in order to reduce elements of destructive competition (Fülöp, 2009) associated with social comparison.

Figure 2: Leaderboard of total posts and replies

During weeks 2, 4, and 6 only badges were used by subjects as the game element in the discussion board. The goal during these weeks was to attempt to unlock one of three possible digital badges: the "Weekly Top 3" badge for being among the top three students for the particular week in terms of total posts/replies, the "Personal Best" badge for creating more posts and/or replies in a given week than in any other week prior to that point, and the "Early Bird" badge for being the first individual to post on a topic. See Figure 3.

Figure 3: Digital badges for online discussion

Badge	Description
	Earn the "Weekly Top 3" badge by being in the top 3 students for total posts/replies in weeks 2, 4, or 6.
	Earn the "Personal Best" badge by creating more posts and/or replies in a given week than in any other week before. Example: Week 1: You created 2 posts and 3 replies for a total of 5 posts/replies. Week 2: You created 2 posts and 4 replies for a total of 6 post/replies. You would have then exceeded your personal best for total posts/replies and your digital badge for week 2 would be unlocked giving you the extra credit points for that week.
	Earn the "Early Bird" badge by being the first individual to post on a topic in a given week! To qualify, your post must be made Sunday after midnight for the given week.

Each week high performing students were recognized in the class via reports that displayed results from the prior week's discussion. Students could access the reports at any time to see if they appeared on the leaderboard of students with the most total posts/replies and/or to review progress on which badges were achieved.

Statistical analysis

The independent variable was the use of the different game design elements: leaderboard vs. digital badges. The dependent variables included two forms of engagement: Behavioral engagement, (measured through post and replies in online discussions) and cognitive engagement (measured through student grades in the discussion forums). Engagement in online discussions was evaluated based on behavioral (active participation) and cognitive engagement (Fredricks et al, 2004).

Analysis consisted of statistical discovery of mean differences between the weeks in which the leaderboard game element was the focus versus the weeks in which the badges game element was the focus. Subjects in which the leaderboard was the focus are identified as group 1 and subjects in which digital badges were the focus are identified as group 2. Variables observed and analyzed include total posts and replies in the online discussion. 2-sample t tests permit the comparison of two means to validate whether there is significant statistical evidence that two means are equivalent (Anderson et al, 2015). Following the analysis of each of the three separate dependent variables, specifically, mean differences specific to student posts, student replies, and total student activity, we sought to assess the statistical significance of the difference between groups based on the vector representing the collective means of the two dependent variables, student posts and student replies. This was performed using Hotelling's t-squared statistic t^2, which is a specialized form of the MANOVA technique and an extension of the univariate t test (Hair et al, 1995). Hotelling's t^2 is appropriate for cases where there are

multiple dependent variables and one independent variable consisting of two categories –in our case, leaderboard-based discussion boards vs. digital badges-based discussion boards. Significance levels (alpha) of 0.05 and 0.10 were utilized for investigation of the data collected.

RESULTS

Behavioral Engagement

Table 1 depicts the statistical measures of central tendency and dispersion for each of the behavioral engagement variables for leaderboard group (1) and for digital badges group (2) . The same twenty-nine students participated in both discussion board groups, alternating between weeks, with odd weeks 1, 3, and 5 dedicated to leaderboards and even weeks 2, 4, and 6 dedicated to digital badges. Regarding posts, we observe a mean of approximately 1.1826 versus 1.1736 posts per student with standard deviations of 0.6153 and 0.4412, respectively (sample size n = 29). Minimum and median are the same. Maximums vary slightly from 5.0 to 4.0. Consequently, we conclude a slight decrease between groups when focusing on student posts for leaderboards. Regarding replies, we observe a mean of 1.1043 versus 1.4380 replies per student with standard deviations of 0.7988 and 1.3590, respectively. Minimum and median are the same. Maximums vary significantly from 5.0 to 11.0. Consequently, we conclude an increase regarding replies for digital badges. Based on these descriptive measures, we observe an overall decrease in student engagement for leaderboards regarding total student activity (mean of approximately 2.2870 versus 2.6120); however, we observe a slight increase for student posts. We observe an overall increase for digital badges regarding student replies.

Table 1: Statistical Measures of Central Tendency and Dispersion for Behavioral Engagement

VARIABLE	GROUP	MEAN	STDEV	MIN	MEDIAN	MAX
Student Posts	1	1.1826	0.6153	1	1	5
	2	1.1736	0.4412	1	1	4
Student Replies	1	1.1043	0.7988	0	1	5
	2	1.4380	1.3590	0	1	11
Total Student Activity	1	2.2870	1.1300	1	2	8
	2	2.6120	1.4110	1	2	12

The results of the 2-sample t tests are presented in Table 2. The means for the three variables for each group are shown along with the resulting p-values and confidence intervals for the significance levels investigated. We observe no significant statistical evidence that the means are different for student posts at either significance level (0.05, 0.10). Regarding student replies, the results reveal significant statistical evidence that the means are different at both significance levels. Regarding total student activity, we observe no significant statistical evidence at the 0.05 significance level; however, we observe significant statistical evidence that the means are different at the 0.10 significance level.

Table 2: 2-Sample t Test Results						
	GROUP MEAN				EQUAL ?	
VARIABLE	1	2	DIFF	P-VALUE	0.05 ALPHA	0.10 ALPHA
Student Posts	1.1826	1.1736	0.0090	0.897	Yes	Yes
Student Replies	1.1043	1.4380	-0.3337	0.022	No	No
Total Student Activity	2.2870	2.6116	-0.3246	0.052	Yes	No

Results for the Hotelling's t^2 resulted in a test statistic of 0.02222, F statistic of 2.648 with 2 degrees of freedom, and a p-value of 0.073. From a collective vantage point at the 0.05 significance level, we observe that the vector means are the same and conclude there is no overall significant difference regarding the effect of the two groups on the two dependent variables total posts and total replies. However, at the 0.10 significance level, we observe a significant difference in student engagement between leaderboard-based discussion boards and digital badges-based discussion boards.

The three statistical activities, descriptive statistics, 2-sample t tests, and Hotelling's t^2, facilitated a means of triangulating towards a sound set of conclusions when comparing leaderboard-based discussion boards and digital badges-based discussion boards. We observe no statistical difference regarding student posts, a significant statistical difference between student replies, and a significant statistical difference regarding overall student activity when considering the 0.10 significance level. Consequently, we conclude no change in engagement regarding posts, an increase in engagement regarding replies for the digital badges group, and an overall increase in engagement for digital badges at the 0.10 significance level. The latter does not support H1, that leaderboards and digital badges will demonstrate no significant differences in boosting behavioral engagement in asynchronous online discussions.

Cognitive Engagement

Discussion grades were retrieved from the university's Learning Management System. Analysis consisted of statistical discovery of mean differences between the weeks 1, 3, and 5 using a leaderboard (group 1) and the weeks 2, 4, and 6 using digital badges (group 2). The average grades were analyzed using a 2-sample t test at the 0.05 and 0.10 significance levels. In the three weeks using a leaderboard, students received an average of 6.930 points with a standard deviation of 2.287. In the three weeks using digital badges, students received an average of 7.08 points with a standard deviation of 2.66. Based on this data, we observe that the means for the discussion grades were not statistically different at both the 0.05 and 0.10 significance levels with a p-value of 0.920. We support the second hypothesis (H2), that leaderboards and digital badges will demonstrate no significant differences in boosting cognitive engagement in asynchronous online discussions.

DISCUSSION

The aim in the present study was to compare two game design elements, leaderboards and digital badges, for possible differences in their effect on two dimensions of engagement with an asynchronous online discussion board. The findings of the present study suggest digital badges are a particularly effective way to increase behavioral engagement. One possible reason for this finding is students who were not able to achieve the goal of being in the top 20% of posts to appear on the leaderboard chose to focus on unlocking digital badges. This behavior is in line with previous research showing the negative impact of engagement for students not appearing near the top of the leaderboard (Glover, 2013; Hanus & Fox. 2015). For these students, digital

badges seem to offer a less socially competitive option in which there is more control over the outcome. For example, in order to achieve the “personal best” digital badge, students simply had to reply more than their own highest total of replies in any previous week. However, the positive effect on increasing discussion replies could not be replicated for the initial discussion post using either game element. The mostly likely explanation the initial discussion posts were not significantly different is because students typically post only one initial post and use multiple replies to improve their performance.

While the goal of our study was to isolate and examine individual game elements for improving engagement, an important byproduct of our findings seems to suggest improvements in behavioral engagement may be best accomplished using multiple game elements. High performing students, for example, may focus efforts more on a leaderboard game element. Digital badges, in contrast, may offer a viable option for students that are unable to appear near the top of a leaderboard.

In regard to cognitive engagement, the present study was not able to determine any significant differences in using a leaderboard or digital badges as a game element to improve learning outcomes. While previous studies have shown improved cognitive engagement using both leaderboards (Butler, 2012) and digital badges (Gibson et al, 2015; Stefaniak & Carey 2019), no significant differences could be identified using either game element based on evaluating discussion grades. One possible reason for the similarity in cognitive performance between the two game elements stems from the requirement that only one post and one reply were used to grade each student’s work. Therefore, all replies beyond the one required post and reply were not measured in comparing cognitive engagement.

CONCLUSION

Despite the research supporting the use of gamification in improving student engagement, studies examining the gamification of online discussions are sparse (Ding et al, 2017; Ding et al, 2018; Ding, 2019). Moreover, these studies typically use multiple gamification methods (Looyestn et al, 2017) resulting in a lack of awareness on the effectiveness of each game element. The aim in the present study was to address this gap by examining two popular game elements, leaderboards and digital badges, for possible differences in their effect on behavioral and cognitive engagement. Each game element was evaluated in an online, undergraduate information systems course in which all students were exposed to a single gamification element each week; either the leaderboard or digital badges. The independent variable was the use of the different game design elements: badges vs. leaderboard. The dependent variables were two forms of engagement: Behavioral engagement, (measured through posts and replies in online discussions) and cognitive engagement (measured through student grades in the discussion forums).

Our findings suggest digital badges are a particularly effective way to increase behavioral engagement. This behavior is in line with previous research showing the negative impact of engagement for students who do not appear near the top of the leaderboard (Glover, 2013; Hanus & Fox. 2015) and also speaks to the importance of using multiple game elements to stimulate motivation to engage for all types of learners. In regard to cognitive engagement, the present study was not able to determine any significant differences in using a leaderboard or digital badges.

The present study makes several important contributions to the information systems literature: First, we fulfill a significant need in gamification research for more studies which evaluate individual game elements for improving engagement. Second, we provide a valuable framework to address low student engagement in online discussions via the gamification of an online discussion board. Finally, from a practical perspective, our study shows that the implementation of digital badges as a game element is particularly effective in promoting increased collaboration and engagement with online discussions.

The present study has limitations, which offer significant starting points for future research. First, we only examined the short-term effects of leaderboards and digital badges. The short eight-week course was sub-divided into two three-week segments focused on each game element. Future studies should focus on repeating this experiment using a longer duration. Second, in regards to cognitive engagement, the present study examined only the first post and reply for evidence of cognitive performance. Future studies examining the impact of game elements in cognitive outcomes for online discussions should focus evaluation on all discussion posts and replies. Finally, the present study focused on two of three aspects of engagement: behavioral, cognitive and affective (Fredricks et al, 2004). Future studies should use established research instruments such as the technology acceptance model (Davis, 1989) and/or IMMS (Huang et al, 2006) to evaluate affective engagement differences in game elements.

REFERENCES

Baek, Y. K. (2008). What Hinders Teachers in Using Computer and Video Games in the Classroom? Exploring Factors Inhibiting the Uptake of Computer and Video Games. CyberPsychology & Behavior, 11(6), 665–671. https://doi.org/10.1089/cpb.2008.0127

Barata, G., Gama, S., Fonseca, M. J., & Gonçalves, D. (2013). Improving student creativity with gamification and virtual worlds. Proceedings of the First International Conference on Gameful Design, Research, and Applications, 95–98. https://doi.org/10.1145/2583008.2583023

Butler, C. (2013). The Effect of Leaderboard Ranking on Players' Perception of Gaming Fun. In A. A. Ozok & P. Zaphiris (Eds.), Online Communities and Social Computing (pp. 129–136). Springer. https://doi.org/10.1007/978-3-642-39371-6_15

Cheung, W. S., & Hew, K. F. (2005). How can we facilitate students' in-depth thinking and interaction in an asynchronous online discussion environment? A case study. 2005 Annual Proceedings-Orlando: Volume, 113.

Cronk, M. (2012). Using Gamification to Increase Student Engagement and Participation in Class Discussion. 311–315. https://www.learntechlib.org/primary/p/40762/

Csikszentmihalyi, M. (1998). Finding Flow: The Psychology of Engagement With Everyday Life (p. 144).

da Rocha Seixas, L., Gomes, A. S., & de Melo Filho, I. J. (2016). Effectiveness of gamification in the engagement of students. Computers in Human Behavior, 58, 48–63. https://doi.org/10.1016/j.chb.2015.11.021

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319–340. JSTOR. https://doi.org/10.2307/249008

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating Internalization: The Self-Determination Theory Perspective. Journal of Personality, 62(1), 119–142. https://doi.org/10.1111/j.1467-6494.1994.tb00797.x

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology/Psychologie Canadienne, 49(3), 182–185. https://doi.org/10.1037/a0012801

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, 9–15. https://doi.org/10.1145/2181037.2181040

Ding, L. (2019). Applying gamifications to asynchronous online discussions: A mixed methods study. Computers in Human Behavior, 91, 1–11. https://doi.org/10.1016/j.chb.2018.09.022

Ding, L., Er, E., & Orey, M. (2018). An exploratory study of student engagement in gamified online discussions. Computers & Education, 120, 213–226. https://doi.org/10.1016/j.compedu.2018.02.007

Ding, L., Kim, C., & Orey, M. (2017). Studies of student engagement in gamified online discussions. Computers & Education, 115, 126–142. https://doi.org/10.1016/j.compedu.2017.06.016

Finn, J. D., & Zimmer, K. S. (2012). Student Engagement: What Is It? Why Does It Matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), Handbook of Research on Student Engagement (pp. 97–131). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_5

Foursquare May Not Be Toast Yet, But It's Browning At The Edges—Business Insider. (n.d.). Retrieved March 11, 2020, from https://www.businessinsider.com/foursquare-may-not-be-toast-yet-but-its-browning-at-the-edges-2012-4

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Fülöp, M. (2009). Happy and Unhappy Competitors: What Makes the Difference? Psihologijske Teme, 18(2), 345–367.

Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The Psychology of Competition: A Social Comparison Perspective. Perspectives on Psychological Science, 8(6), 634–650. https://doi.org/10.1177/1745691613504114

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating Cognitive Presence in Online Learning: Interaction Is Not Enough. American Journal of Distance Education, 19(3), 133–148. https://doi.org/10.1207/s15389286ajde1903_2

Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., & Knight, E. (2015). Digital badges in education. Education and Information Technologies, 20(2), 403–410. https://doi.org/10.1007/s10639-013-9291-7

Glover, I. (2013). Play As You Learn: Gamification as a Technique for Motivating Learners. 1999–2008. https://www.learntechlib.org/primary/p/112246/

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1984). Multivariate data analysis with readings, 1995. Tulsa, OK: Petroleum Publishing.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. 2014 47th Hawaii International Conference on System Sciences, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Computers & Education, 80, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. Instructional Science, 28(2), 115–152. https://doi.org/10.1023/A:1003764722829

Hewitt, J. (2005). Toward an Understanding of How Threads Die in Asynchronous Computer Conferences. Journal of the Learning Sciences, 14(4), 567–589. https://doi.org/10.1207/s15327809jls1404_4

Huang, W., Huang, W., Diefes-Dux, H., & Imbrie, P. K. (2006). A preliminary validation of Attention, Relevance, Confidence and Satisfaction model-based Instructional Material Motivational Survey in a computer-based tutorial setting. British Journal of Educational Technology, 37(2), 243–259. https://doi.org/10.1111/j.1467-8535.2005.00582.x

Kapp, K. (2012). The gamification of learning and instruction: Game-based methods and strategies for training and education. San Francisco, CA: Pfeiffer.

Leong, B., & Luo, Y. (2011). Application of game mechanics to improve student engagement. International Conference on Teaching and Learning in Higher Education.

Lepper, M. R. (1988). Motivational Considerations in the Study of Instruction. Cognition and Instruction, 5(4), 289–309. https://doi.org/10.1207/s1532690xci0504_3

Lo, C. K., & Hew, K. F. (2018). A comparison of flipped learning with gamification, traditional learning, and online independent study: The effects on students' mathematics achievement and cognitive engagement. Interactive Learning Environments, 1–18. https://doi.org/10.1080/10494820.2018.1541910

Locke, E. A., & Latham, G. P. (2006). New Directions in Goal-Setting Theory. Current Directions in Psychological Science, 15(5), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x
Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. PLoS ONE, 12(3). https://doi.org/10.1371/journal.pone.0173403

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. PLoS ONE, 12(3). https://doi.org/10.1371/journal.pone.0173403

Nakajima, T., & Lehdonvirta, V. (2013). Designing motivation using persuasive ambient mirrors. Personal and Ubiquitous Computing, 17(1), 107–126. https://doi.org/10.1007/s00779-011-0469-y

Ng, C. S. L., Cheung, W. S., & Hew, K. F. (2009). Sustaining Asynchronous Online Discussions: Contributing Factors and Peer Facilitation Techniques: Journal of Educational Computing Research. https://doi.org/10.2190/EC.41.4.e

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. Motivation and Emotion, 15(3), 185–205. https://doi.org/10.1007/BF00995170

Stefaniak, J., & Carey, K. (2019). Instilling purpose and value in the implementation of digital badges in higher education. International Journal of Educational Technology in Higher Education, 16(1), 1–21. https://doi.org/10.1186/s41239-019-0175-9

Van Kleek, M., Smith, D. A., Hall, W., & Shadbolt, N. (2013). “the crowd keeps me in shape”: Social psychology and the present and future of health social machines. Proceedings of the 22nd International Conference on World Wide Web, 927–932. https://doi.org/10.1145/2487788.2488082

Werbach, K., & Hunter, D. (2015). The Gamification Toolkit: Dynamics, Mechanics, and Components for the Win. Wharton School Press.

Xu, F., Weber, J., & Buhalis, D. (2013). Gamification in Tourism. In Z. Xiang & I. Tussyadiah (Eds.), Information and Communication Technologies in Tourism 2014 (pp. 525–537). Springer International Publishing. https://doi.org/10.1007/978-3-319-03973-2_38

Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. Instructional Science, 34(6), 451. https://doi.org/10.1007/s11251-006-0004-0

DECISION SCIENCES INSTITUTE

Does Extrinsic Motivation to Participate in Class Discussions Encourages Intrinsic Motivation for Learning? A Case Study in a Virtual Engineering Classroom

Ruholla Jafari-Marandi
School of Business,
University of Redlands
Email: roy_jafari@redlands.edu

ABSTRACT

The influential role of extrinsic or intrinsic motivation on student learning is undeniable. However, the complex and under-researched relationship between intrinsic and extrinsic motivation and their interaction influence on student learning and performance make designing an optimum class experience very difficult. This paper poses the question of whether students' intrinsic motivation for learning is increased if they are encouraged extrinsically, specifically by grade, to participate in class discussion. The paper also showcases a real-classroom experiment that offers valuable insights for designing empirical research to answer the posed question.

KEYWORDS: Intrinsic Motivation, Extrinsic Motivation, Student Learning and Performance

INTRODUCTION

It is not a myth that class engagement predicts students' success and learning (Furrer & Skinner, 2003). If this is a matter of encouragement for undergraduate level students, it should be regarded as a necessity for graduate-level classes (Kyei-Blankson, 2009). This paper presents an innovative evaluation assessment that proved to enhance students' engagement and showed promise to improve students learning.

LITERATURE REVIEW

The relevant studies are introduced in three subchapters: motivation, innovative course evaluation, and active learning & class engagement. The last paragraph in this section place this paper in the presented literature.

Motivation

It is difficult to argue with the vital role of motivation in pedagogical development (Wise, 2004). The intrinsic vs extrinsic classification of motivation has proven extremely useful in understanding students' attitudes and designing college courses (Sansone & Harackiewicz, 2000). While intrinsic motivation correlates highly with perceived usefulness, and higher levels of satisfaction (McEvoy, 2011; Dunn, 2014), extrinsic motivation jointly with intrinsic motivation predict student success (Lee, McInerney, Liem, & Ortiga, 2010; Cerasoli, Nicklin, & Ford, 2014). There is evidence that even graduate students are highly grade-motivated, and the motivation of many graduate students is largely extrinsic and performance-based rather than intrinsic and mastery-based (Hancock, 2007; Hegarty, 2010). Moreover, there are other indirect factors that students are

grade-motivated overly enough to even commit academic misconduct (Rettinger, Jordan, & Peschiera, 2004).

The overwhelming evidence that mixing intrinsic and extrinsic motivation on the effectiveness of enhancing performance (Matsumoto & Sanders, 1988; Guay, Vallerand, & Blanchard, 2000; Becker, McElvany, & Kortenbruck, 2010; Rogstadius, Kostakos, Kittur, Smus, Laredo, & Vukovic, 2011) should motivate educators to design assessment methods that engage both types of motivations. However, despite many efforts to demystify the interaction between intrinsic and extrinsic motivation, the state of literature remains at either subjective investigation of the interaction (Covington & Müeller, 2001) or ignoring of the interaction in the empirical studies (Areepattamannil, Freeman, & Klinger, 2011; Chen, 2014). In short, this article presents a new graduate assessment method that combines the intrinsic and extrinsic motivation to contribute to the class discussion which is known to enhance students' learning and performance. This new assessment method could help design research experiments that empirically demystify the influence of the interaction on student learning.

Innovative learning assessment methods

As students are shown to be grade motivated, designing a course assessment method can be one of the most important steps in mixing intrinsic and extrinsic motivation (Carless, 2015). For instance, in contemporary higher education, standard-based assessment methods are adopted to create a more transparent, and mastery-based approach toward classroom design (Sadler, 2005; Ajjawi, Bearman, & Boud, 2019). As educators now we better understand the detrimental effects of one-dimensional assessment on students learning and appreciate the value of qualitative assessment (Suskie, 2018). As such, the need for testing innovative assessment methods to engage both intrinsic and extrinsic motivations and actively engage students is highly palpable.

Active Learning and Class Engagement

Evidence is abundant about the benefits of active learning and class engagement toward improving students learning. When the lectures are interactive, students are not just passively observing information pass through their conscious mind, but rather student's mind goes through simultaneous and energetic activities that lead to a subjective ducking of the material as own experience (Van Dijk, Van Der Berg, & Van Keulen, 2001). Such engagement has shown to improve both students' perceived usefulness (Daouk, Bahous, & Bacha, 2016), and student performance (Freeman, Eddy, McDonough, Smith, Okoroafor, Jordt, & Wenderoth, 2014).

Study placement

This paper presents a new course assessment method that is designed to extrinsically motivate class engagement and active learning in graduate-level courses. The name of the assessment method is the Contribution Point System.

THE NEW ASSESSMENT METHOD – CONTRIBUTION POINT SYSTEM

The contribution point system solely evaluates students' performance based on the amount, depth, and value of their contribution to the class. The assessment method is designed for graduate-level courses, as there is an expectation for graduate students to contribute to the class. Graduate

students need to reach an acceptable level of contributions to pass the course. The contributions throughout a course may be recognized by two possible badges: Excellent Contribution (EC), and Good Contribution (GC). Students need to get at least 5 ECs to pass the course. Every two GCs are worth one EC.

The instructor may introduce different ways a graduate student may contribute to the class: 1) contribute to the class conversation during the lecture, class meetings or video calls, 2) post or comment a very thorough analysis in one of the discussion forums, 3) leading a group conversation, 4) present a relevant and novel topic, 5) lead a discussion. Moreover, the instructor may encourage students to be creative and find other ways of contributing to the class.

There are a couple of caveats that need attention. First, this assessment method is not suitable for a typical lecture-based course and rather is ideal for courses that are discussion-based and have a more open structure. For instance, flipped courses where the students learn the material before coming to class and work on a case study or a problem in small groups or as a class during class time. Second, this assessment method must be mixed with a more formative assessment where students are encouraged to study for the course learning objectives. For instance, the letter grades can be assigned from a combination of achieved contribution points and a final exam. Figure 1 portrays an example of how the two can be synthesized.

		B	5 ECs & 75 on Final	C-	5 ECs
A	5 ECs & 90 on Final	B-	5 ECs & 70 on Final	D+	4 ECs
A-	5 ECs & 85 on Final	C+	5 ECs & 65 on Final	D	3 ECs
B+	5 ECs & 80 on Final	C	5 ECs & 60 on Final	D-	2 ECs

Figure 1 Letter grading policy example

EXPERIMENT

The presented assessment method was implemented in a graduate-level course at the Industrial and Manufacturing (IME) department at California Polytechnique State University in San Luis Obispo (Cal Poly – Obispo) during the 2020 spring quarter. The quarters in Cal Poly – Obispo last for ten weeks added with a final exam week. The course name is IME 520 - Advanced Information Systems for Operations and it carries 4 seminar units which means the class meets four hours weekly and 40 hours in total. As a rule, Cal Poly students are advised to study an hour per hour of class seminar units. That means the students taking IME 520 are advised to attend the classes for four hours and self-study for another four hours, in total eight hours a week and 80 hours during the quarter.

In total 32 students enrolled for IME 520 offered during the 2020 spring quarter, and as this quarter coincided with the worldwide COVID-19 pandemic, the course was delivered 100% virtually. To take advantage of the opportunities of the new mode of content delivery and avoid its shortcoming a few measures were taken.

First, the class was flipped, that is, the 4 hours weekly synchronous class meeting was dedicated to class discussions about case studies, and the lectures and other materials were delivered asynchronously. This was done to defeat the monotony of watching lectures online, and students mentally checking out due to lack of much engagement. From personal experiences and also

empirical research (Guo, Kim, & Rubin, 2014), we know that it is a fact that long video lectures are not useful learning tools. One technique to beat that monotony during an asynchronous delivering lecture is to implement relevant group discussions in between the lecture. Zoom breakout rooms provide an excellent technological solution for that purpose (Mercer, Pisutova, & Rogers, 2018). However, to meaningfully and productively engage students for this course, the discussion should be about real case studies, otherwise, the concepts presented during the lectures are very abstract. As engaging students with real case studies has a very high lead time and also needs students forethought, the technique would not have been effective for this course. Therefore, the courses were flipped to effectively rectify the problem of student's high tendency to mentally check out during long online video exposures.

Second, instead of instructor's putting the abstract concepts into perspective using examples and explanations during lectures, as the material was being delivered online anyways, many video case studies were used to prime students with practical examples of the material, and prepare them to be exposed to the abstract concepts. Third, the course was designed to allow contributions for students who are less social and prefer to deliberate their thoughts before sharing. This was done by allowing students to answer case study questions and chapter questions in discussion forums. Lastly, as proctoring online exams are difficult and it has become a contentious issue (James, 2016), the final exam for this course was decided to be a design-based take-home final.

Week 3 Anonymous Feedback

To empower students by giving them a voice and also having a chance to adjust the design of the class due to its many new aspects, during the third week of instruction students were encouraged to participate in an anonymous survey. Figure 2 displays the questions that students were asked and Figure 3 shows the analysis of 31 student unique submissions. The participation rate in this feedback was 97% given that the size of the class was 32 students.

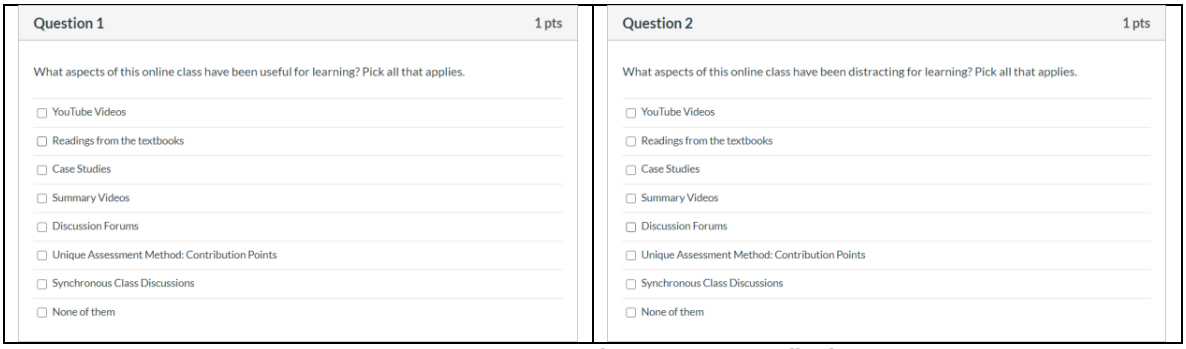


Figure 2 Questions in Week 3 Anonymous Feedback

As Figure 3 a) reveals the majority of the students find most aspects of the designed classroom useful. The only rather inconclusive opinion of the student body was the new assessment method. Even though more students found the new assessment distracting than useful, only about 40% of the students found it distracting and more than 60% of the students either found it useful or non-distractive. That is an expected and favorable outcome for new methods of course assessment. However, to alleviate the students' concern regarding the assessment method two more options were presented to the students at week 4. The default option was the one presented in the syllabus. The first alternative is for students to take two finals: one designed-based and

take-home final and one conceptual final from the textbook. The final ground would be calculated by an equally-weighted average of the two finals with standard grading scales. The first alternative was created for students who found the contribution point system distracting, however, the second alternative was designed for students who found the pace of the course too slow. Students during the anonymous feedback were also asked about the pace of the class and 19%, 72%, and 9% were the response rate respectively for too fast, just right, and too slow. The second alternative was for students to improve their grades from simple passing to higher letter grades by doing a class project with defined deliverables. For this alternative, the interested students would have to book an appointment with the instructor and together decide about the relevant class project. Despite the nearly 40% disagreement of students with the default assessment method, no students chose to do either first or second alternative and everyone persisted with the default option.

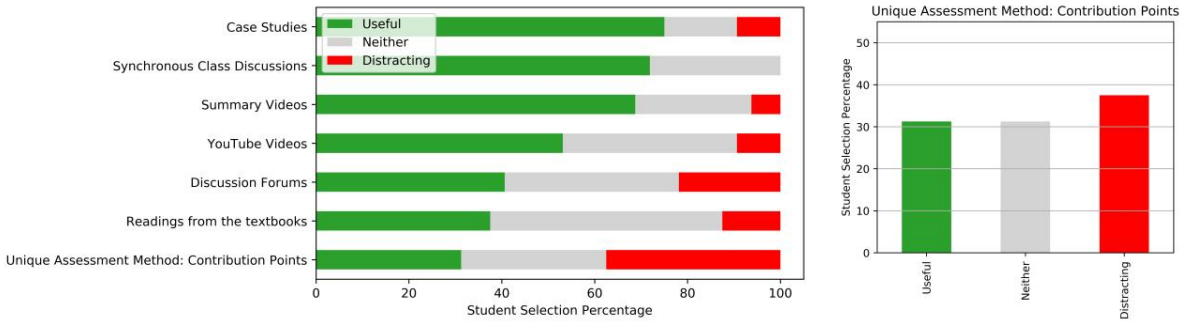
Week 10 Anonymous Feedback

At the end of the course, another anonymous feedback was requested from the students. Apart from the two questions in Week 3 anonymous feedback (Figure 2), the following questions were also asked. ADQ stands for Agree or Disagree Questions.

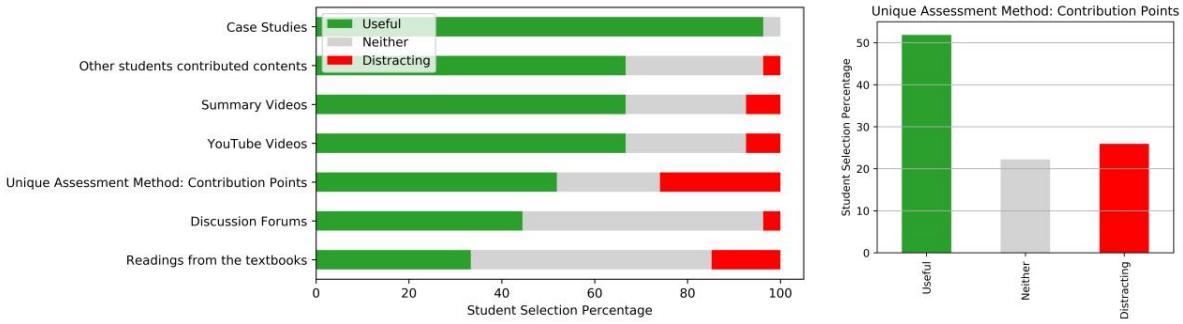
- **ADQ1 (Motivation Factor):** The contribution point system was a motivating factor for me to meaningfully participate in the class discussion (whether synchronous or asynchronous).
- **ADQ2 (Forcing Factor):** The contribution point system often created a situation that I felt that I had to say something even though what I had to say in my own mind was not a worthwhile contribution to the class.
- **ADQ3 (Comfort Factor):** The contribution point system created a culture that made contributing to the class more comfortable.
- **ADQ4 (Overall Opinion):** Overall, I think the contribution point system is an excellent substitute for the traditional grading system for graduate-level classes.
- Explain your experience with the contribution point system. Some talking point suggestions:
    - Helpful/distracting
    - Motivating/distracting
    - Created a culture that made contributing more/less comfortable
    - Lead to better learning the course objectives
    - Any other relevant item

The rate of participation in this second survey was 84% (27 out of the 32 students). Figure 3 b) and Figure 4 show the summary of students' responses. The first factor that stands out from Figure 3, specifically from comparing the results of Week 3 and week 10 surveys is that many students changed their minds in favor of the new assessment method after having had 7 more weeks of engaging with it. One might notice a subtle change between part a) and b) in Figure 3 – “synchronous class discussion” and “other students contributed contents” are not part of both surveys. This is simply a result of inconsistency in carrying out the surveys. Second, while the student body seems not to have changed their mind about “reading from the textbook”, but they also changed mildly in favor of “discussion forums.” Third, the other “students contributed contents” is one of the highly supported aspects of the class in the week 10 survey. That shows even students who may have found the contribution point system distracting found its aftermath valuable and useful for learning.

The response to Agree or Disagree Quesitons (ADQ) also confirms the students' overall positive experience with the new assessment method. First, more than 90% of the students agreed that the contribution point system was a motivating factor for them to meaningfully participate in the class discussion (whether synchronous or asynchronous). Second, only 30% of the students agreed the contribution point system often created a situation that they felt that they had to say something even though what they had to say in their own mind was not a worthwhile contribution to the class. Third, almost 80% of the students agreed that the contribution point system created a culture that made contributing to the class more comfortable. Lastly, 70% of the students agreed that the contribution point system is an excellent substitute for the traditional grading system for graduate-level classes.



a) Week 3 Anonymous Feedback



b) Week 10 Anonymous Feedback

Figure 3 - Week 3 & 10 Anonymous feedback analysis

To better understand the results of the students' opinions, a clustering analysis on the data from the week 10 anonymous survey was performed and four meaningful clusters of students emerged from the data. To perform the clustering analysis, the following columns of data were used: binary columns showing if the student finds each aspect of the class useful or not (seven columns), binary columns showing if the student finds each aspect of the class distracting or not (seven columns), binary columns showing if the student agreed or disagreed with the four ADQs. K-Means clustering algorithm was used to group the student based on their attitude toward the class and the new assessment method. After running K-means with different Ks (number of clusters), it was apparent that four is the best number of clusters using the silhouette goodness of clustering measure. This measure captures both the tightness of in-cluster members and the farness of out-of-cluster members. To study and profile these four clusters, the columns that were meaningfully distinguishing the clusters from one another were found by

calculating the variance of the average value of each data column over the 4 clusters. The columns that had variance greater than 0.1 were used to illustrate Figure 5 which shows each cluster's opinion tendency and also contrast that with other clusters. As Figure 5 Suggests, the students in clusters have meaningful similarity to one another, and also meaningful dissimilarities from the members of the other clusters. To help with the analysis of the results each cluster has been given a name. The four clusters are namely New System Opposers (NSOs), Discussion Lovers + Online Learners (DLOLs), New System Appreciators (NSAs), and New System Neutrals (NSNs).

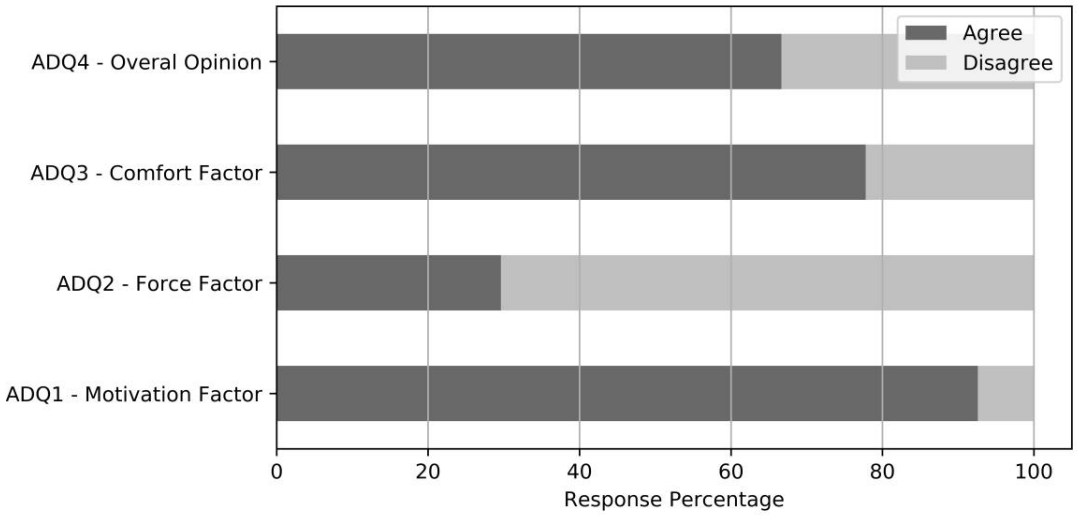


Figure 4 - week 10 Anonymous Survey – Responses to ADQ

The New System Opposers (NSOs) comprise one-fourth of the class that did not find the contribution point system useful for their learning. Interestingly, NSOs also did not find the content provided by other students very useful either. Some of the NSOs did find the discussion forums useful, and also some of them found the new assessment method helpful in creating a more comfortable culture to contribute to class discussions. On the other hand, Discussion Lovers + Online Learners (DLOLs) found the new assessment method useful and successful in creating a culture that facilitates making contributions. DLOLs make up the majority of the class, 37 percent. DLOLs were the cluster of students that found the discussion forums and the new assessment method the most useful. What separates, DLOLs from New System Appreciator (NSAs) is that unlike DLOLs, NSAs did find the new system feeling forceful. It is interesting that NSAs still had a favorable opinion for the new assessment method even though they found it forceful. Moreover, NSAs did not find the youtube video useful which is in dire contrast with DLOLs. The smallest cluster of students who were also very different from NSOs was New System Neutrals (NSNs). NSNs did not find the new assessment method useful or distracting, but overall agreed it would be a suitable assessment method for graduate-level courses. Again, unlike DLOLs, NSNs found the content provided by other students the most useful.

The students' verbatim comments to the last question in the week 10 Anonymous survey are provided in the appendix for insight and consideration.

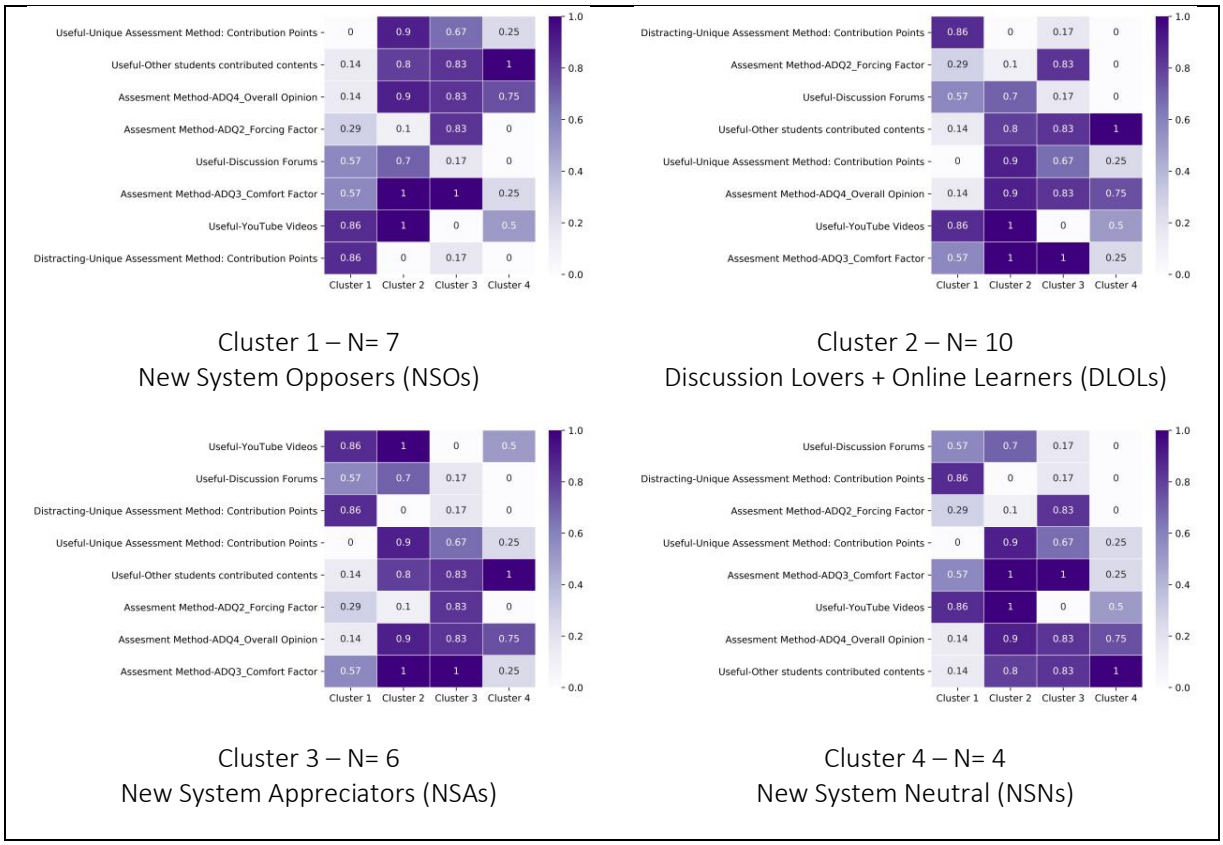


Figure 5 – Clustering and profiling of students based on week 10 anonymous feedback

Lessons Learned

As this was the first time the Contribution Point System was implemented, many insights and lessons emerged during the quarter. First, while the new assessment method matches well with some students' learning styles, it is in contrast with some others. While the system was designed to accommodate a wide range of different learners as far as the way the students were able to contribute to the class in different forms, there was one student that felt extreme anxiety even having to think about presenting to class for the majority of their grade. The student in question ended up making excellent contributions to the class only through the discussion forums and was able to pass the course, but the stress they felt was real. It is important to pay attention that a new assessment method such as the one presented in this paper will likely increase the difficulty of getting good grades for students who feel less included in the class or have not found their unique voice among their cohort. To assuage this possible concern, I recommend instructors repeat their commitment to fairness, and openness throughout the class. It will be also helpful if students can trust the instructor will be open to forming individual assessment methods for students who feel a high level of stress because of the newness of the assessment method.

The second lesson that emerged from this experiment was the instructor must do calculations in advance to avoid situations that the majority of the class reach their required level of contribution before the course ends. If this happens students will stop participating and the course dynamic will become out of balance. This was a hard lesson for this experiment, as because of this mistake,

I had to give students a chance to earn more contribution points instead of having to take the final exam. This was a difficult but necessary decision that had to be made. I strongly recommend that instructors have some ideas of how much contribution points students will be making and come up with policies that keep them engaged throughout the quarter. There are some interesting ideas among students' comments in the appendix.

CONCLUSION & FUTURE WORK

This paper posed a multi-dimensional research question about the effect of extrinsic motivation for class discussion on the intrinsic motivation for learning and student performance. While the paper does not offer empirical research to answer the question, however, it does showcase a real class experiment that gives valuable insight for designing empirical research to answer the question. The experiment, however, suggests extrinsic motivation does positively influence perceived usefulness and learning.

REFERENCES

Ajjawi, R., Bearman, M., & Boud, D. (2019). Performing standards: a critical perspective on the contemporary use of standards in assessment. Teaching in Higher Education, 1-14.

Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Intrinsic motivation, extrinsic motivation, and academic achievement among Indian adolescents in Canada and India. Social Psychology of Education, 14(3), 427.

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. Journal of Educational psychology, 102(4), 773.

Carless, D. (2015). Excellence in university assessment: Learning from award-winning practice: Routledge.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. Psychological bulletin, 140(4), 980.

Chen, C.-A. (2014). Nonprofit managers' motivational styles: A view beyond the intrinsic-extrinsic dichotomy. Nonprofit and Voluntary Sector Quarterly, 43(4), 737-758.

Covington, M. V., & Müeller, K. J. (2001). Intrinsic versus extrinsic motivation: An approach/avoidance reformulation. Educational psychology review, 13(2), 157-176.

Daouk, Z., Bahous, R., & Bacha, N. N. (2016). Perceptions on the effectiveness of active learning strategies. Journal of Applied Research in Higher Education.

Dunn, K. (2014). Why wait? The influence of academic self-regulation, intrinsic motivation, and statistics anxiety on procrastination in online statistics. Innovative Higher Education, 39(1), 33-44.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410-8415.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. Journal of educational psychology, 95(1), 148.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and emotion, 24(3), 175-213.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. Proceedings of the first ACM conference on Learning@ scale conference. 41-50.

Hancock, D. R. (2007). Effects of performance assessment on the achievement and motivation of graduate students. Active learning in higher education, 8(3), 219-231.

Hegarty, N. (2010). Application of the academic motivation scale to graduate school students. The Journal of Human Resource and adult learning, 6(2), 48.

James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. International Journal of Educational Technology in Higher Education, 13(1), 19.

Kyei-Blankson, L. (2009). Enhancing student learning in a graduate research and statistics course with clickers. Educause Quarterly, 32(4).

Lee, J. Q., McInerney, D. M., Liem, G. A. D., & Ortiga, Y. P. (2010). The relationship between future goals and achievement goal orientations: An intrinsic–extrinsic motivation perspective. Contemporary Educational Psychology, 35(4), 264-279.

Matsumoto, D., & Sanders, M. (1988). Emotional experiences during engagement in intrinsically and extrinsically motivated tasks. Motivation and emotion, 12(4), 353-369.

McEvoy, G. M. (2011). Increasing intrinsic motivation to learn in organizational behavior classes. Journal of Management Education, 35(4), 468-503.

Mercer, J., Pisutova, K., & Rogers, R. (2018). From mystery to mastery: Creating & enhancing the ultimate virtual classroom experience. 2018 COIL Conference Proceedings. 112-124.

Rettinger, D. A., Jordan, A. E., & Peschiera, F. (2004). Evaluating the motivation of other students to cheat: A vignette experiment. Research in Higher Education, 45(8), 873-890.

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. Fifth International AAAI Conference on Weblogs and Social Media.

Sadler*, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. Assessment & evaluation in higher education, 30(2), 175-194.

Sansone, C., & Harackiewicz, J. M. (2000). Intrinsic and extrinsic motivation: The search for optimal motivation and performance: Elsevier.

Suskie, L. (2018). Assessing student learning: A common sense guide: John Wiley & Sons.

Van Dijk, L., Van Der Berg, G., & Van Keulen, H. (2001). Interactive lectures in engineering education. European Journal of Engineering Education, 26(1), 15-28.

Wise, R. A. (2004). Dopamine, learning and motivation. Nature reviews neuroscience, 5(6), 483-494.

APPENDIX

The students' answer to the last question in the Week 10 anonymous feedback was insightful. Their answers are reported here verbatim, however, they are presented under five categories: Positive (8), Positive with minor frustration (9), Positive Critique (4), Ambivalent (1), Negative Critique (5). The number in the parentheses is the number of comments under each category of comments. The comments were categorized subjectively by the content of the comments themselves. The student cluster presented in Table 1 is the cluster of the commenting students based on their other responses to the same survey (See Figure 5).

Table 1 – Students responses to the last question of Week 10 anonymous feedback

Category	Comment	Student Cluster
Positive	I found the contribution point system to be motivating and created a much better environment in the classroom. Students were more inclined to share opinions and contribute to meaningful discussions.	NSA
Positive	I think that the contribution points made it more comfortable to contribute and share ideas in class. Typically, students do not want to say too much or ask too many questions in fear of being judged as a nerd or teacher's pet, but the contribution points eliminate this fear as we all must contribute and discuss so no one is judging. In addition, the points motivated me to stay up to date on class lessons and material so I could contribute valuable opinions. Overall, I really liked the contribution points and would recommend this method of grading for most graduate classes.	DLOL
Positive	I felt that the contribution points system allowed me the flexibility to engage in the class and explore other students' points of view. I very much enjoyed the relaxed nature of the class and did not feel overly pressured to get a "grade". I think this allowed me to enjoy the topics better and furthermore it reminded me that I am in this class to learn and better myself and not just get a grade. This made me try harder and contribute more because I was not as afraid of being wrong.	DLOL

Positive	The contribution system made me a lot more motivated to be ready and prepared for class, I wanted to have good contributions especially when sharing in front of my peers. I felt the discussion boards got very competitive and most were completed a week in advance. I really enjoyed the student contributed topics, I felt that I learned a lot about very relevant technology.	NSNs
Positive	I found the contribution point system extremely helpful for my learning. I chose to only do online discussions and I think it made me pay a lot more attention to what I was reading/watching/writing because I knew my writing would be visible to the entire class so I wanted it to be better quality than if I was just turning it in for homework. The only thing that wasn't so good about the contribution points is that it would have been really easy to completely ignore certain parts/chapters of the class if you already had a certain number of contribution points. Having a high number of contribution points to shoot for was good because it motivated me to continue to contribute. The one thing I really appreciate about online courses is that it has, in many cases, forced professors to be more creative in their assessment methods. Based on my experience at Cal Poly, many of my professors give tests that are very memorization based, but I feel like I've learned so much more in all my classes because the assessment methods have been more about learning and understanding than about answer questions correctly on a test.	DLOL
Positive	I thought this system was very helpful because I am the kind of student who normally sits back and listens to discussions rather than actively participating. It made each class a good exercise for challenging myself to be more articulate and think critically about each subject. Additionally, the amount of contribution points necessary for each letter grade seemed reasonable in that students were dominating discussions in order to get an A but each student had to contribute a decent amount to get a good letter grade.	DLOL
Positive	The contribution point system makes me more willing to read the discussion board after class to see the responses of the students and put forward my own opinions. Based on the rich content on the discussion board, I also learned many different cases from it.	DLOL

Positive	I liked the contribution point system a lot because it put less pressure on the class and made it more enjoyable. It allowed me to be more present in class. It also made me feel a little more motivated to participate in class. I like that there were varying levels of points for quality of contribution because I think it forced us to think more deeply and make sure we were not contributing for the sake of it. For improvement: More small group discussions - having ~5 gives everybody the chance to be each role In discussion forum on Canvas, create a response that is just a statement of the question and have people build off that so there are no repeats	DLOL
Positive with minor frustrations	The contribution system created a great culture of sharing! I can tell you poured an immense effort into this course and my learning was a reflection. While every online class is "you get what you put into it," I felt that this course was especially engaging and accommodating to different learning styles. I did feel that students would jump on the discussion forums very quickly, as in, each question in a discussion post would be fully answered and exhausted within an hour or two of its release to the forum. This was frustrating at times, but I found ways to get around this.	NSA
Positive with minor frustrations	I enjoyed the contribution points especially when it was for people teaching apps and topics. It was very insightful and informative when people would provide learning for apps and topics. We were able to get information from different points of view which was nice. However, when it came to discussion forums online. I didn't think they were as informational or useful for myself. I found that I sometimes would post and then never check back in with the discussion and they didn't seem to foster as much debate or personal insight. Overall, I enjoyed the contribution points because it kept me engaged throughout the quarter. I also appreciated how there were many different ways to earn points. I would also like to add that I think they were given out fairly.	NSA

Positive with minor frustrations	I found the contribution point system helpful in that it encouraged me to participate in the class and have meaningful discussion with the class. I would much rather earn contribution points through synchronous class discussions and discussion forums than take a final or write a paper. For me, the former is more enriching than the latter. However, when it was announced that you could opt out of the final based on the second calculation of contribution points, contributing became less comfortable. I think this had a lot to do with timing, I felt it was announced a little too late and the class started scrambling for points, which made things feel very competitive rather than collaborative.	NSNs
Positive with minor frustrations	I thought the contribution point system made me more comfortable to participate in class more than I normally would, I liked hearing other's interpretations and thoughts on the readings. However, my favorite part of the class was learning about the new apps and discussion topics form my peer's presentations. I think if it were required that every person do at least one of those presentations, people would have been more comfortable contributing overall also all the topics would have been covered over the course of the class. Overall, I really enjoyed this class and feel like I have gained a lot of useful skills and knowledge on different applications I can use in industry as well as day to day life.	DLOL
Positive with minor frustrations	The contribution point system, though slightly uncomfortable at first, was helpful in getting me out of my shell to participate in class more often.	NSA
Positive with minor frustrations	I thought that the contribution point system created a better dynamic for how classmates were contributing to the class. There was good discussion and people brought up some very interesting talking points to be discussed. People were more creative with how they thought about the subject at hand. One downside to the contribution points is that it often created a very repetitious discussion because people were all trying to convey similar ideas to get points. This is just a result of many people in the class and a small number of questions to be answered. The other options to get contribution points, like presenting an app or a topic helped alleviate this a little bit. Also, with the limited time that I had I felt that it was better to focus most of my efforts on contribution points. I think that I will go back and read through interesting sections of the textbook after this quarter so that I can gain some more knowledge from there.	NSNs

Positive with minor frustrations	I enjoyed the contribution grading system because it made me think about things much deeper than I initially would if I was just reading. The discussions were very helpful because I could build off other points I heard and use new knowledge to apply in different scenarios. I do wish that the 8 EC method had been announced earlier because I felt that it in the first few weeks I started to relax more when I had thought that I had hit the maximum amount of ECs.	DLOL
Positive with minor frustrations	I thought that it was fairly easy to get the contribution points early on, especially during the synchronous discussions. The cap on points in week 4 and week 8 were frustrating because it lowered my motivation to participate in class. I would have found it more engaging if there were categories of contributions that I had to achieve. Like if I had to contribute at least 3 points towards online discussions, and 3 points to synchronous discussions, and 2 points of contributed topics or discussions. I did like that we could move at our own pace though because there are times in the quarter where other classes can be overwhelming and if you can select when you want to contribute it can be helpful. The contribution points prevented me from reading the textbook due to time constraints which was difficult because I found the text very interesting. It might have been interesting to give students the opportunity to earn contribution points by creating a presentation video about one of the chapters.	DLOL
Positive with minor frustrations	Because of the large amount of resources (discussions, case studies, videos, readings, etc) it became hard to balance it all - especially with other demands of the graduate program. I found myself focusing on doing activities that related directly to contribution points rather than reading the book. However I did find this to be more impactful and relevant than reading a textbook. It was very interesting learning from other students as well.	DLOL

Positive Critique	I think the contribution point system is a move in the right direction, but it is difficult to implement. Obviously, there were multiple methods of grading considered throughout the class, and I feel like at the end, the two options were good ones. The largest difficulty however is the consistency in earning points. When creating a discussion presentation or video, a student may put 5+ hours of work in for a EC. However that same student can simply talk in class twice and earn the same EC given that they bring up a good point, which takes far less time and effort. This is a frustration that I know was shared by many other students, but I don't know if I can offer a solution for it. At the end of the day, any grading system is going to have flaws, but at least using "contribution points" encouraged more participation and engagement.	NSA
Positive Critique	The contribution grading system was helpful in encouraging open participation and critical thinking. I liked how it motivated peers to build off of each other's points and have an ongoing dialogue. It created one of the most comfortable cultures around participating that I have seen in an academic class online. I did not like the way in which the contribution system was distracting because of how long it took to adjust to it, understand how we are getting graded on specific types of contribution, and overall lack of clarity at times on how many contribution points we had in total because of discussions that had not yet been graded or reviewed.	NSO
Positive Critique	Contribution system is good to motivate students to participate, but it does not really excite me to learn more about a topic. It creates a competition around contributing, but not a constructive learning environment where everyone helps each other to get through an assignment for example. I feel we learn better when we help our peers understand something better in a hands-on environment. I appreciate the thought process behind getting everyone to contribute and create a comfortable environment in an online format though. Personally, information systems is a very interesting to me and I would have loved to learn more about different technologies and perform labs to actually learn about implementing these technologies. Data is the future and this class can be structured to be a little more practical. It was a fun experience for an online class though and hope this feedback helps in a constructive manner.	NSO

Positive Critique	I felt that the contribution points system was effective in gathering participation. However, I do not feel that it is an appropriate system for grading a graduate level course. I feel like students should be put through more rigorous evaluation methods. If a student wanted to contribute the bare minimum and receive 0.25 contribution points per online and in class discussion, he or she could easily pass through the course without a firm understanding. I feel like the bar should be placed higher for what graduate students are expected to accomplish. Completely basing the course on a contribution points system would be too much. I would prefer you stick to a partial contribution, partial final exam/paper option. That way, the contribution points can only get you so far, but it still encourages participation.	NSNs
Ambivalent	It certainly made me engage and share with the class, but it felt forced. It felt like people (including myself) were commenting just for the sake of getting a contribution point. At times I felt like the whole point of attending class was to get a contribution point, and I would fixate on what I was going to say rather than listen and learn.	NSO
Negative Critique	I felt a lot of pressure to try and beat other students to commenting quickly to make sure that I could get the points for the session. Also, I felt that the in-class contribution was inconsistent, as sometimes I would say lots of meaningful comments and get 0.25, while other times I would say 1 thing and still get 0.25. I liked the small group discussions and thought they were great ways to have real conversations about the readings.	NSA
Negative Critique	I think the contribution point system led to increased participation, but lower quality of content. Many of the contributions were surface-level, preventing us from going into the level of depth that I would want from a graduate-level class. I appreciate how you accommodated people's different learning styles (discussion forum, synchronous discussion, etc.), but did not learn much from any of the available options.	NSO
Negative Critique	The system was not necessarily bad. It did not align with “learn by doing” in my opinion though. I personally prefer more project-based classes.	NSO

Negative Critique	I would say ultimately the motive behind contribution points is good, but it ended up being mostly distracting. This was a result of how it always became a competition to contribute whether in lecture or on a discussion forum. It encouraged learning just enough to answer a few questions. Instead, it may be more useful to have free response case study quizzes such as in the group work. The group work overall provided a better learning experience than the contribution system.	NSO
Negative Critique	I think it made the overall class a little forced and made me feel like I always needed to contribute and say something to get grades. I wish the class was a little more technical instead of discussion based.	NSO

DECISION SCIENCES INSTITUTE

System Analysis and Design of Smart Assistive Service for Tire Assistance

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Yiyun (Cindy) Fei
Georgia Institute of Technology
Email: yfei38@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Pan Zou
Georgia Institute of Technology
Email: zoupzp123@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Shaojie (Emily) He
American Heritage School
Email: pl241111@ahschool.com

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

The tire safety of private vehicles heavily relies on the on-time inspection of the tires. The vast market volume of tire aftersales market implies an enormous potential of a service system to serve the driver. This paper develops a smart assistive service system for the tires. This smart assistive service system has a mobile application for interaction with customers, a tire pumping robot, as well as a software solution to link the convenience store and tire service center. This smart assistive service system envisions an ecology to serve drivers and acts as a platform role to integrate the service resources.

KEYWORDS: Smart assistive service system, Mobile Application, Robot

INTRODUCTION

The objective of this smart assistive service (SAS) system is to automate the process of filling the air in automobile tires. A substantial proportion of automobile tires are improperly inflated, leading to significant safety, economic and environmental issues (Udompetaikul et al., 2011). According to NHTSA, only 19% of consumers properly inflate their tires in the US, and there are approximately 400 deaths and 78000 car crashes due to tire blowouts caused by underinflation every year (NHTSA, 2019). By inflating tires properly and driving at the prescribed tire air pressure, not only can consumers save costs, but they can also reduce a tremendous amount of carbon dioxide emission (Johnson, 2018). Several automatic pumping appliances have been put into the market, but a systematical solution is a lack to serve the private customers (Hendricks, 2011). Moreover, several practitioners have developed smart tires with embedded sensors, and one of them even developed a mobile application linking to the sensors. However, their solution has not been linked to the smart assistive service system (Askari et al., 2019).

Therefore, a smart assistive service system has been proposed as a systematical method. The system has three elements: pump robot, mobile application, and service software solution. The pumping robot is an automated tire inspecting and pumping appliance, which contains the loading dock, elevator, end-effector, and the base structure. A machine vision module is installed at the end-effector to identify the nozzle, and an auto connected pumping nozzle to adjust the pressure. Moreover, a mobile application is developed to provide customer access to the system, which can show instructions, product recommendations, and purchasing options. These data will be sent to the smart assistive service system to find the appropriate business entity, which includes convenience stores, tire service centers, to name but a few.

USER ANALYSIS

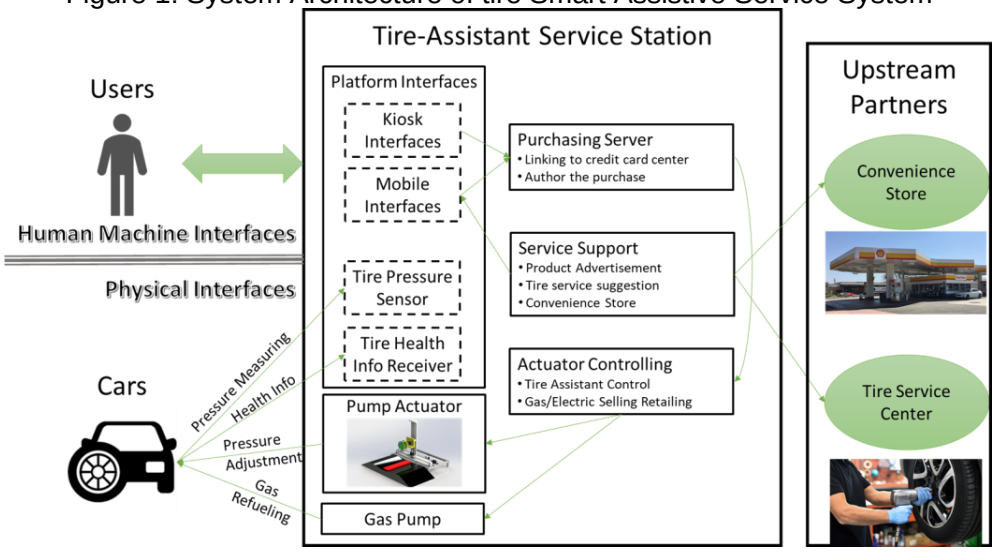
This system can be used in parking garages, car dealerships, gas and EV charging stations, and mostly all automobile stops, as shown in Table 1. The solution proposed in this work targets on the gas stations and EV charging station as the primary use cases and sets private car drivers as our targeted consumers. First, gas and EV charging stations can be exposed to a large number of consumers. The increase in the number of inspected vehicles leads to more chances of the system being used to adjust tire pressure. Thus, the death rate and the number of car accidents caused by underinflation will be reduced. Unlike in parking garages, more consumers stay at the gas or EV charging stations. Both the position and period are suitable for the usage of this system. Moreover, customers have a higher chance of receiving further service suggestions, including tire product recommendations, purchasing options, and appointments.

Table 1 User Analysis		
	Users	Justifications
01	Common people and non-professional drivers	Save labor efforts and time Low using frequency
02	Disabled and elderly people	Eliminate physical movement High medical safety requirements
03	Car dealership and workshop	Save labor time High productivity and efficiency
04	Gas and EV charging station	Service while energy refueling Extend to a merchandising spot
05	Public parking space	Service while parking Low using frequency

SYSTEM ARCHITECTURE

Our systematical solution is shown in Figure 1. The entire system starts with the consumer. Once the customer approaches the gas station or EV charging station, they can drive the car on the loading dock to initiate the service process. The customer interacts with the system by mobile application or kiosk interfaces. These two interfaces will ask permission to check the tire pressure. Once it is permitted, the actuator control module will ask the pump actuator to approach the tire and engage the valve. According to the vehicle information from the mobile application registration and the tire information the sensor retrieved, the pump can adjust the pressure to an ideal level. Once the process is completed, the system restarts with the next consumer. The mobile application will show recommended relevant products that can be purchased and link the services from both convenience stores and tire service centers.

Figure 1: System Architecture of tire Smart Assistive Service System

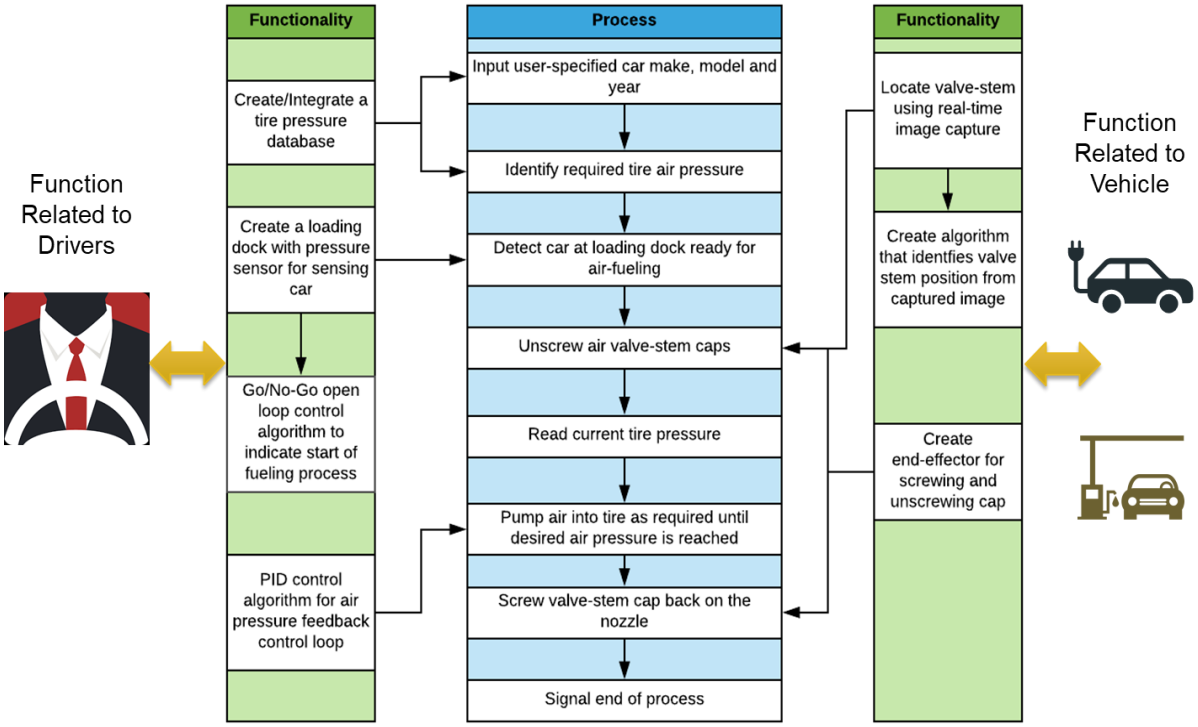


USE CASE ANALYSIS FOR FUNCTIONS OF PUMPING ROBOT

As an essential physical module of tire product-service systems, the pumping robot should serve the driver and vehicle, respectively. The use case analysis has been widely adopted as an effective way to derive the function of the system and the corresponding processes (Fantechi et

al., 2003). The function analysis is based on the use cases from driver and vehicle. Thus, the service process can be derived in the following. The use case analysis is shown in Figure 2.

Figure 2: Use Case Analysis for Functions and Service Process



The first function the driver needs is the creation and integration of a tire pressure data base. This function acts as a registrational functionality, which requires the system to allow the input of user-specific car make, model, and year. This information paves the way for identifying the required tire air pressure. The following functionality is the creation of a loading dock with a pressure sensor to sense the position of the car. This functionality can lunch the third process of the system, which is the detection of the readiness of the car at a loading dock for air-fueling. After that, the driver can select the service or not start the service loop.

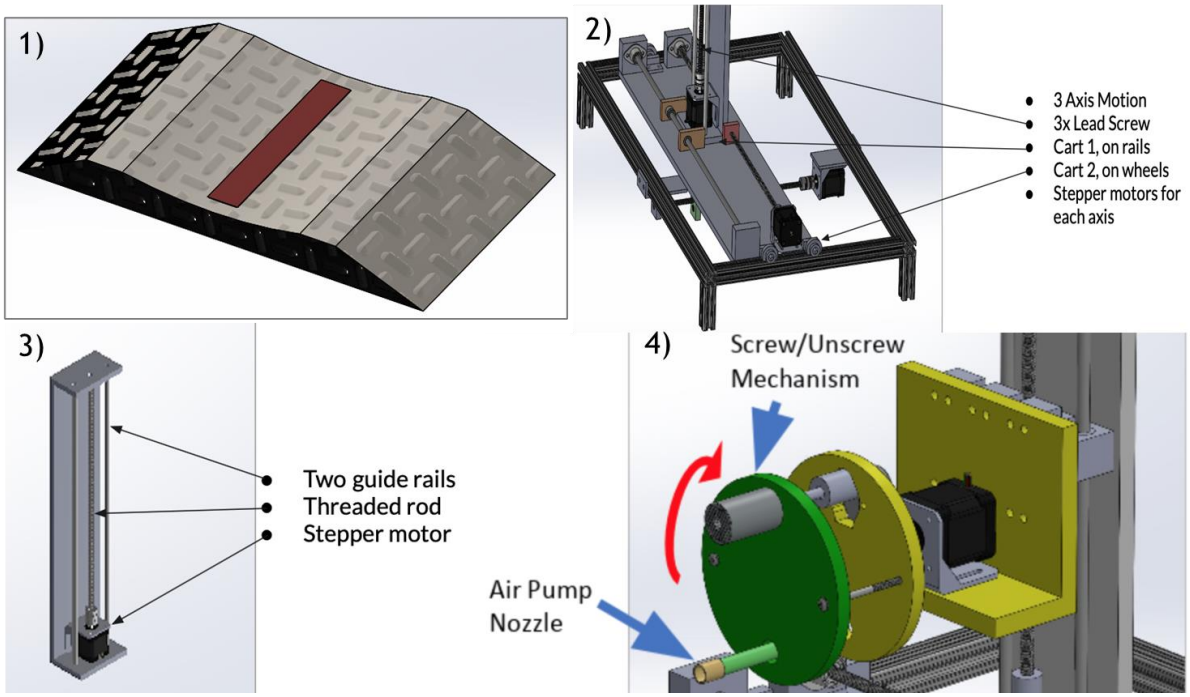
If the driver decides to start the service loop, the functions will serve the vehicles by locating the valve-system using real-time image capture. This functionality enables the unscrewing of the valve-stem cap, reading the tire pressure data, pumping the air into the tire to the required level, and screw the valve-stem cap back on the nozzle. This loop is controlled by the PID control algorithm for the air pressure feedback loop. Also, this loop creates an end-effector for screwing and unscrewing the cap. The last process is the signaling end of the service.

DESIGN EMBODIMENT OF AUTOMATIC PUMPING ROBOT

The design embodiment of the automatic pumping robot includes the design of subsystems and finite element analysis (FEA) for verification.

This system is composed of four main subsystems, as shown in Figure 3. The first one is a loading dock, which detects and starts the whole air-filling process when a car drives onto it. Once the pressure sensor on the bottom is pressed, actions of the next subsystem triggers. The second subsystem is a base system, which contains two carts, three stepper motors, and three lead screws driving X, Y, and Z axes motion, respectively. It also supports the two carts, an elevator, and a head system. The third subsystem is an elevator, which consists of two guide rails, one threaded rod, and a stepper motor. The threaded rod supports the Z-axis motion of the Head System to get to the tire valve-stem location. This elevator supports the end effectors and the camera. The fourth subsystem is the head system, which mainly consists of two end effectors. One of the end effectors is a screwdriver to unscrew and screw tire valves stem caps before or after adjusting the tire pressure. The other end effector is for tire air filling. Both end effectors are mounted on a head plate that can rotate 360 degrees so that this system works for all possible locations of a tire valve stem. A stepper motor is used to facilitate this rotation, depending on whether the tire valve stem is in the upper or lower hemisphere.

Figure 3: Four Subsystems

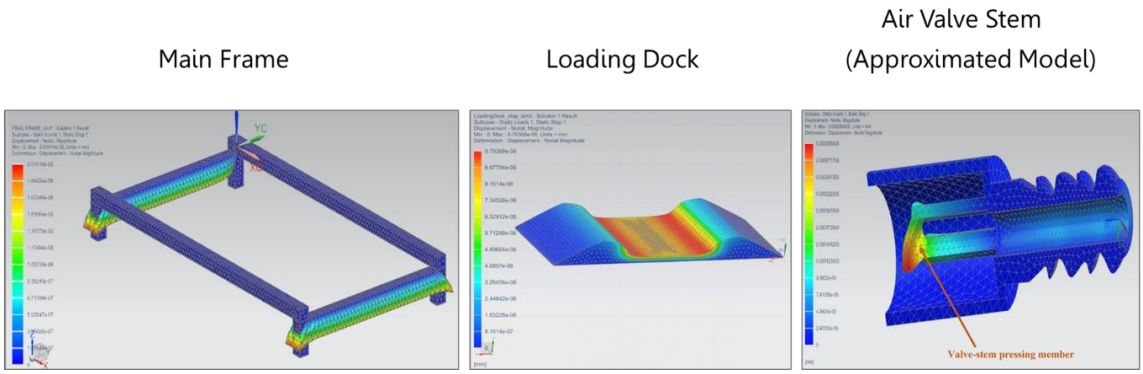


FEA has been applied to critical parts of our system to identify design and material failures, which is shown in Figure 4. The first critical part is the mainframe of the device, which consists of majorly 80/20 aluminum. The maximum displacement is 2.04E-6 millimeters, which can be approximated to nearly negligible. This result suggests that mainframe material and design are satisfactory. The second critical part is the loading dock, which is aluminum support for wheel placement. It will contain a pressure sensor to detect the presence of a wheel. The maximum displacement is 9.79E-6 millimeters, which can be approximated to nearly negligible. This result suggests that loading dock material selection and design is satisfactory.

The air valve system is the third critical part, which is used ASTM SAE AISI 1045 carbon steel to engage the wheel. The results of maximum Von-Mises Stress of 10.6 MPa and maximum

displacement of 0.3 millimeters are minimal, which suggests the air valve design and material selection, is satisfactory

Figure 4: FEA Analysis Results



The final design can achieve the following service. It can automatically sense the arrival of the vehicle. It can align with and unscrew the air valve cap. The end effector can sense the tire pressure while the pump can adjust the air until full. After pumping, the robot can re-attach the air valve cap and return to the initial position. The automatic engagement process to the tire is shown in Figure 5.

Figure 5: Automatic Tire Engagement

DISCUSSIONS AND CONLUSIONS

The proposing smart assistive service system achieves the customer services by detecting the valve, design a mechanism to reach the valve, and pumping the air to the tire while refueling and charging. The proposed solution has implemented advanced software analysis methods, which include the PID control model and finite element analysis.

Moreover, the introduction of mobile applications and back-end software solutions envisions a product service ecology in the center of the tire of a private vehicle. The application can provide interactive access to the customer, which enables a cyber platform to integrate back-end suppliers and service providers. This system can give access to the convenience store in the gas station and EV charging station, as well as the tire service center. It not only enhances the user experience but also extends the traditional gas station and EV charging station to a potential purchasing site.

ACKNOWLEDGMENTS

Anviksha Busa, Christopher Chao, Haider Jamal, Achint Lehal, Chukwuanugo Mojekwu, and Ricardo Riveraare acknowledged for their valuable contributions to prototype development and running the experiments.

REFERENCES

Askari, H., Hashemi, E., Khajepour, A., Khamesee, M.B. and Wang, Z.L., 2019. Tire condition monitoring and intelligent tires using nanogenerators based on piezoelectric, electromagnetic, and triboelectric effects. Advanced Materials Technologies, 4(1), p.1800105.

Fantechi, A., Gnesi, S., Lami, G. and Maccari, A., 2003. Applications of linguistic techniques for use case analysis. Requirements Engineering, 8(3), pp.161-170.

Hendricks, T.E., Hendricks Sr Todd E, 2011. Robotic tire spraying system. US Patent 7,943,201.

Johnson, D., 2018. Gas Prices Are Going Up. Here's How Much More You'll Pay This Year Based On Your Exact Car. Time, https://time.com/5306658/gas-prices-calculator/

National Highway Traffic Safety Administration, 2020. Tires, https://www.nhtsa.gov/equipment/tires

Udompetaikul, V., Upadhyaya, SK and Vannucci, B., 2011. The effect of tire inflation pressure on fuel consumption of an agricultural tractor operating on paved roads. Transactions of the ASABE, 54(1), pp.25-30.

DECISION SCIENCES INSTITUTE
When Managers Meet Models: Integrating Human Judgment and Analytics

Rebekah Brau
University of Arkansas
rbrau@walton.uark.edu

John Aloysius
University of Arkansas
jaloysius@walton.uark.edu

Enno Siemsen
University of Wisconsin
esiemsen@wisc.edu

ABSTRACT

Although analytics and machine learning are increasingly prevalent in emerging forecasting practice, practitioners continue to employ human judgment to incorporate contextual information. Our research uses an experiment to examine how human judgment and analytical models may be best integrated to improve forecast accuracy. Our findings suggest that human judgment provides a significant benefit to forecasting dependent on the method of integration. Interactive machine learning and human-guided machine learning are the most effective methods of integration in comparison to other methods commonly used in practice and studied in the academic literature.

KEYWORDS: Machine Learning, Human Judgment, Operations Forecasting, Laboratory Experiments

DECISION SCIENCES INSTITUTE

The role of risk avoidance mechanisms on media-reported environmental supply chain sustainability risks

Ivana Mateska
Swiss Federal Institute of Technology Zurich (ETH Zurich)
Email: imateska@ethz.ch

ABSTRACT

As the complexity of supply networks increases and the sustainability awareness spreads globally, firms are exposed to pressures for superior environmental performance along their supply chains by various media stakeholders. Grounded in the agenda-setting theory and stakeholder theory, this study takes into account the complex interplay between buying firm's organizational characteristics and environmental sustainability levels and evaluates the effectiveness of risk avoidance mechanisms on media-reported supply chain sustainability risks. The findings suggest that for buying firms which belong to vulnerable categories, media-reported wrongdoings could be attenuated by pro-active (supplier selection) whereas augmented by re-active (contract termination) risk avoidance mechanisms.

KEYWORDS: Supply Chain Risk Management, Environmental Management

DECISION SCIENCES INSTITUTE

When Worse Is Better: Strategic Choice of Vendors with Differentiated Capabilities in a Complex Co-Creation Environment

Shubham Gupta
Fox School of Business, Temple University
Email: gupta.shubham@temple.edu

Abhishek Roy
Fox School of Business, Temple University
Email: abhishek.roy@temple.edu

Subodha Kumar
Fox School of Business, Temple University
Email: subodha@temple.edu

Ram Mudambi
Fox School of Business, Temple University
Email: ram.mudambi@temple.edu

ABSTRACT

The increasing complexity of consulting, research and development, and information technology projects has resulted in close collaboration between clients and vendors, who co-create value through co-design and co-production. In such co-creation environments, the output is contingent on the positive efforts exerted by all parties involved. These collaborative settings often involve a client engaging with more than one vendor. However, existing research on collaborative value co-creation has mainly focused on the one-vendor scenario, while research on multi-vendor sourcing has rarely considered value co-creation. We study a scenario involving a client and two vendors, examining the trade-offs that arise in the strategic interactions among the three firms. We derive novel and interesting insights about how these interactions differ from those under one-vendor co-creation, and from those under traditional multi-vendor sourcing. Specifically, we determine the conditions under which the client prefers the two-vendor arrangement over one-vendor co-creation, and the effects of such a preference on the primary vendor. Several counter-intuitive results emerge from our model: The client can sometimes be better off by adding a new vendor less efficient than the primary one; and the primary vendor becoming more expensive can sometimes make the client less inclined to add the new vendor. We also show that the collaborative nature of the interaction results in the primary vendor benefiting from the addition of a new vendor whenever it is beneficial for the client to add this vendor. The conceptual and managerial insights drawn from our research contribute to a better theoretical and practical understanding of strategic decision-making by firms in multi-vendor co-creation environments that are increasingly prevalent.

KEYWORDS: Collaboration, co-production, co-creation, multiple vendors, game theory, analytical modeling

DECISION SCIENCES INSTITUTE

Where's My Package: Vehicle Routing with Home Delivery Attempt Predictors

Stanley Lim
University of San Diego
stanleylim@sandiego.edu

Qingchen Wang
University of Hong Kong
qcwang@hku.hk

ABSTRACT

About 20% of business-to-consumer deliveries fail on first attempt. Failed deliveries carry not only cost-implications, but also incur a significant damage to retailers' brand reputation. Despite its economic significance, extant literature tends to prescribe recommendations to improve delivery operations based on the implicit assumption that attempted deliveries are successful. We address this gap by first identifying a set of drivers influencing delivery outcomes via a fixed-effects regression. Using these variables, we develop a machine learning algorithm, and then incorporate it into routing optimization to explore the value of accounting for failed attempts in routing models. We discuss several managerial implications.

KEYWORDS: Home delivery attempt, Online retail, Econometrics, Machine learning, Vehicle routing

DECISION SCIENCES INSTITUTE
Role of Social Media in Marketing – A Bibliometric Analysis of Literature

Mehmood Khan
Abu Dhabi University
Email: mehmood.khan@adu.ac.ae

ABSTRACT

Social media has attracted wide-ranging interest over the last two decades, both in academia and in industry, mainly due to its global reach to consumers. This paper takes a bibliometric approach, reviewing literature drawn from the Scopus database published in 2005–2020. In all, 2780 documents were located and analyzed using Biblioshiny (RStudio) and VosViewer. The findings indicate that the US, Australia, and China are active in social media research. Furthermore, the Journal of Marketing and Business Horizons are the highly-cited journals in the field of social media.

KEYWORDS: Social Media; Marketing; RStudio; Bibliometrics; Biblioshiny; VosViewer

INTRODUCTION

Social media marketing is an amazing tool that organizations of all sizes can use to reach their customers, most of whom connect with brands through social media. Ignoring this huge community of spectators on Facebook, Twitter, Instagram, and Pinterest, would mean losing market share – Schweidel and Moe (2014); Busca and Bertrandias (2020). Smart advertisements delivered through web-based networks can have a significant effect on the success of a business, promoting dedicated brand advocates and driving leads and deals – Kannan, PK (2017).

Social media advertising is a type of web promotion that communicates to specific target markets to accomplish marketing objectives. It may include posts, pictures, and recordings that drive interest and commitment. The way the web content is created and how much is delivered depends on the advertiser's goals, intended interest groups, and social media chosen – Campbell *et al.* (2014). For instance, a web-based travel business, which often creates profoundly visual marketing materials, can produce a great deal of significant worth from advertisement on Instagram or Pinterest. A business-to-business organization may find more response from Twitter or LinkedIn – Huang *et al.* (2019).

Social media marketing is used to increase site traffic, enhance brand profile, promote brand loyalty, and improving its correspondence and cooperation with interest groups. The closer a business is to its intended audience, the easier it is to achieve promotion objectives – Majumdar and Bose (2019). Examples of marketing using social media include the following:

- Attracting the attention of a target group through the arrangement of content
- Drawing clients in and maintaining the relationships though live online promotions
- Showcasing one's brand image in online networking
- Promoting new products or services through webpages and blogs
- Building on trust and quality by curating outside sources and providing connection to them
- Monitoring what networking channels are being tapped by competitors

The dominance of social media advertising is evident from the millions of posts that go up daily – Nair *et al.* (2019). The vast range of content appearing on social networks entails asymmetrical data and challenges in dealing with purchasers – Jovicic (2019). The contemporary consumer is more inclined to rely on client-created content, such as photographs, recordings, and remarks – Kim, W. G. and Park (2017). This social content provides an exceptional avenue for comprehending shoppers' inclinations – Jin and Phua (2014). For example, TripAdvisor and VirtualTourist help service providers learn how visitors compare tour operators, hotels, and airlines – Slattery *et al.* (2019).

The exploration of social media content through empirical research can help industry experts keep abreast of developments in their fields and allow specialists and professionals to stay current with the most recent customer trends in their field – Kim, Y. *et al.* (2019). Because a single model cannot reliably predict all preferences, it is vital to routinely audit customers' tendencies and inclinations in different regions to anticipate future patterns – De Caigny *et al.* (2019).

Although few publications have provided a general overview of the use of social media in marketing, none has given a broad picture of the current state and evolution of the literature in social media using bibliometric indicators. Bibliometric analysis, enabled by the widespread availability of bibliometric software, is receiving increasing appreciation in the research community. the graphics software VosViewer can generate maps of bibliographic coupling, co-authorship, citation, co-citation, and co-occurrence of keywords This methodology, which originated in library and information science, has found use in a variety of areas, such as. in health and security – Jalali *et al.* (2019); entrepreneurship – Baier-Fuentes *et al.* (2019); and international business – Rialp-Criado and Rialp (2020).

This paper provides detailed insight into the research on social media marketing, derived from the analysis of a set of articles obtained from the Scopus database. This identifies the most influential and productive research in the field, as well as providing the most significant authors, institutions, keywords, and countries. Some topics are showing increasing popularity, while others are losing attention share.

The reminder of the paper is organized as follows: section 2 introduces the idea of bibliometric analysis, section 3 describes the research methodology, section 4 presents the results of the analysis, section 5 provides the research implications, and section 6 concludes the paper.

BIBLIOMETRIC ANALYSIS

Background

Bibliometric analysis describes the scope of research in a specific area of interest both quantitatively and qualitatively – Ellegaard and Wallin (2015). Developed for library and information science, bibliometrics is used to classify research according to publications, authors, and journals – Merigó and Yang (2017); Potter (1981); Diodato and Gellatly 2013. In other words, bibliometrics augments the scientific literature with an improved understanding of research mechanisms – Osareh (1996). Stevens (1953) divided bibliometric studies into two large areas and several subareas, as follows:

1. Productivity count or descriptive area:
    a) Country or geographical location
    b) Period of time
    c) Discipline or subject area

2. Evaluative area:
   a) Reference
   b) Citation

Bibliometric studies can be descriptive or evaluative – Potter (1981). Descriptive studies categorize publications by country, contributing authors, journal, year, and discipline. The second category ranks publications using how much they are citated and where. This analysis allows publications to be evaluated (Wei (2019) in given subject areas, whether in health studies – Ekundayo and Okoh (2018); criminal studies – Sweileh (2018); or social research – Kannan, P. and Thanuskodi (2019) or in identifying early trends in research – Ellegaard and Wallin (2015).

The basic structure of specific leading journals in certain disciplines have been analyzed using these tools as well. For example, Tang *et al.* (2018) performed a bibliometric overview of publications in *sustainability* between 2009 and 2018 and found that it has published more each year, and researchers in the US and South Korea countries contributed the most to this journal. Further, the keywords *sustainability*, *management*, and *China* were observed the most often. To celebrate the tenth anniversary of the journal *Forests*, Uribe-Toril *et al.* (2019) analyzed all of its published research to determine leading trends and challenges, concluding that researchers in the US, China, and Canada contributed the most it. Authors from 105 countries published in *Forests*, making it a truly global journal. Its publications and citations were found to be increasing annually.

Bibliometrics

Bibliometrics uses references between documents to establish statistical models of scholarly communication streams, adopting diverse indicators and measures, e.g., co-citation analysis, which quantitatively investigates the structure of publications in a particular area of interest – Wei (2019). Where citation is the acknowledgment that a document receives from another one – Osareh (1996); co-citation is the frequency with which two documents are cited together – Small (1973). An analysis of citations indicates the relative influence of a study, author, or journal in a field of interest – Meho and Rogers (2008). Likewise, the impact of an author or article can be determined from co-citation analysis by counting the number of times it is cited with other articles – Mattsson *et al.* (2011).

Another popular quantitative technique in bibliometrics is keyword analysis – Ellegaard and Wallin (2015). This type of analysis describes what keywords co-occur in publications in a specific area and how often – Ding *et al.* (2001). For example, co-word analysis identifies patterns and trends in large sets of publications – Assefa and Rorissa (2013). Finally, bibliometrics in a general sense identifies scientific collaboration through author collaborations among researchers, institutions, and countries – Khan *et al.* (2016).

Tools in Bibliometric Analysis

Bibliometric methods are used in quantitative analyses of publications. When first developed, this involved bibliographic overviews of scientific productions or commonly cited publications. A number of new tools have become available to produce more publication data and present a wide range of indicators, as listed in Table 1. For a comparison between these tools and a description of their strengths and weaknesses, the reader is referred to Harinarayana (2019).

Table 1. Summary of Popular Tools in Bibliometric Analysis	
Element	Description
BibExcel	Olle Persson
Authors	Authors' frequency distribution
Pajek	Vladimir Batagelj and Andrej Mrvar
CiteSpare	Chaomei Chen
VosViewer	Nees Jan van Eck and Ludo Waltman
R-Bibliometric Package	Massimo Aria and Corrado Cuccurullo

METHODOLOGY

RStudio (with Biblioshiny) and VosViewer were employed to explore the features of a collection of documents published in 2005–2020 and taken from Scopus database. The search term *social media* was used with the keywords *marketing*, *service*, *education*, *healthcare*, *hospitality*, and *tourism/tourist*. This produced 2780 documents.

The Bibliometrix package in RStudio contains a suite of tools for accurate data processing. For example, term extraction, file conversion, merging, duplicate matching, descriptive analysis, and similarity normalization for network analysis form part of the package. Matrices are built from the publication dataset (e.g., words, keywords, authors, countries, and references) for coupling – Darvish (2018). The Biblioshiny package is a Web application framework for R that transforms analysis performed by the R script into interactive Web applications. Utilizing R and Biblioshiny together can allow positive development of the bibliometric community through co-citation, collaboration, conceptual framework, and multiple correspondence analyses – Liu and Li (2016).

It should be noted that the Scopus data contains several elements, including authors' names, titles, keywords, and other information, for each paper in the set. These elements together constitute the bibliographic attributes of a document, termed metadata. The standard output from Biblioshiny, i.e., tables and graphs, cannot be modified or changed – Aria and Cuccurullo (2017). Table 2 describes the elements that appear in the outputs and results with Biblioshiny – Aria and Cuccurullo (2017).

Table 2. Description of Elements in Biblioshiny	
Element	Description
Articles	Total number of publications
Authors	Authors' frequency distribution
First Authors	First author of each publications
Affiliation	Frequency distribution of affiliations (of all co-authors for each document)
CO	
Sources	Affiliation country of first author
DE	Frequency distribution of the sources (journals, books, others)
ID	Frequency distribution of the authors' keywords
	Frequency distribution of keywords associated to the document by Clarivate Analytics Web of Science and Scopus databases
TotalCitation	Number of times each document has been cited
TCperYear	Yearly average number of times each document has been cited

To produce a detailed view of authors and citations, a graphical analysis of the bibliometric material is carried out using VosViewer – Van Eck and Waltman (2017). This tool generates graphic maps, developed from data on bibliographic coupling, co-authorship, citation, co-citation, and co-occurrence of keywords – Merigó and Yang (2017). Table 3 describe the elements used in the outputs and results with VosViewer.

Table 3. Description of Elements in VosViewer	
Element	Description
Bibliographic Coupling	Two articles citing the same third article
Co-citation	Two articles receiving a citation with the same third article
Co-authorship	Extent of collaboration between the most productive sources
Citation Analysis	
Co-occurrence of Author Keywords	Level of citations between two articles
	Frequency of common keywords

RESULTS AND IMPLICATIONS

Performance Analysis

Main Results

First, I present an overview of the acquired data developed in Biblioshiny. Table 4 presents 2780 documents by 6529 authors from 2005 to 2020.

Table 4. Overview of Main Results

Description	**Results**
Documents	2780
Sources (Journals, Books, etc.)	1556
Keywords Plus (ID)	6008
Author's Keywords (DE)	4517
Period	2005 – 2020
Average citations per documents	11.28
Authors	6529
Author Appearances	7979
Authors of single-authored documents	578
Authors of multi-authored documents	5951
Single-authored documents	661
Documents per Author	0.426
Authors per Document	2.35

Category-wise Classification of Documents

The data were then categorized into 13 types of documents. Most were articles, followed by conference papers. Table 5 presents the distribution of document types.

Table 5. Main Document Types

Type	No. of documents	Percentage
Article	1525	54.88%
Conference Paper	574	20.65%
Book Chapter	274	9.86%
Review	165	5.94%
Note	71	2.55%
Editorial	62	2.23%
Book	31	1.12%
Letter	28	1.01%
Article in Press	16	0.58%
Erratum	15	0.54%
Short Survey	12	0.43%
Conference Review	4	0.14%
Data Paper	2	0.07%

Year-wise Distribution of Documents

Then, the data were distributed by year from 2005 to 2020 to determine progress in the productivity in research on social media. Table 6 shows year-wise publication statistics.

Table 6. Year-wise Distribution of Documents

Year	Publications	Year	Publications	Year	Publications
2005	1	2011	116	2016	346
2007	1	2012	153	2017	387
2008	5	2013	213	2018	429
2009	15	2014	263	2019	486
2010	46	2015	302		

The earliest publication found appeared in 2005. There has been a remarkable growth in this subject area, with 486 documents published in 2019 alone.

Geographical Distribution of Contributors

This analysis investigated countries that appeared most often in the dataset to determine its international status. Biblioshiny uses the country of the affiliation of the first author for this purpose. Ultimately, publications came from 93 countries, according to this metric. Authors from the US contributed the most publications (1711), followed by the UK (470) and Australia (359). Table 7 presents the top 20 most productive countries in terms of publishing on the subject area. These results indicate the global impact of social media.

Table 7. Countries' Ranking Based on Publications

Rank	Region	No. of publications	Rank	Region	No. of publications
1	USA	1711	11	Taiwan	100
2	UK	470	12	Indonesia	87
3	Australia	359	13	Malaysia	80
4	Canada	248	14	Netherlands	73
5	China	248	15	Finland	71
6	India	188	16	Turkey	61
7	Spain	148	17	Greece	59
8	Germany	116	18	France	58
9	South Korea	111	19	Saudi Arabia	57
10	Italy	101	20	South Africa	48

Authors' Ranking and Contribution

It was investigated which authors contributed the most to the subject area. Figures 1 and 2 show the ranking of the top 20 authors and contributors of documents in this field and their publications over the last 15 years. A total of 2528 authors contributed to 2780 publications, Fernandez Luque from Qatar was the most prolific author, with 12 publications, followed by De Choudhury from the US, with 11 publications.

Figure 1. Highest-Ranked Authors by Number of Publications

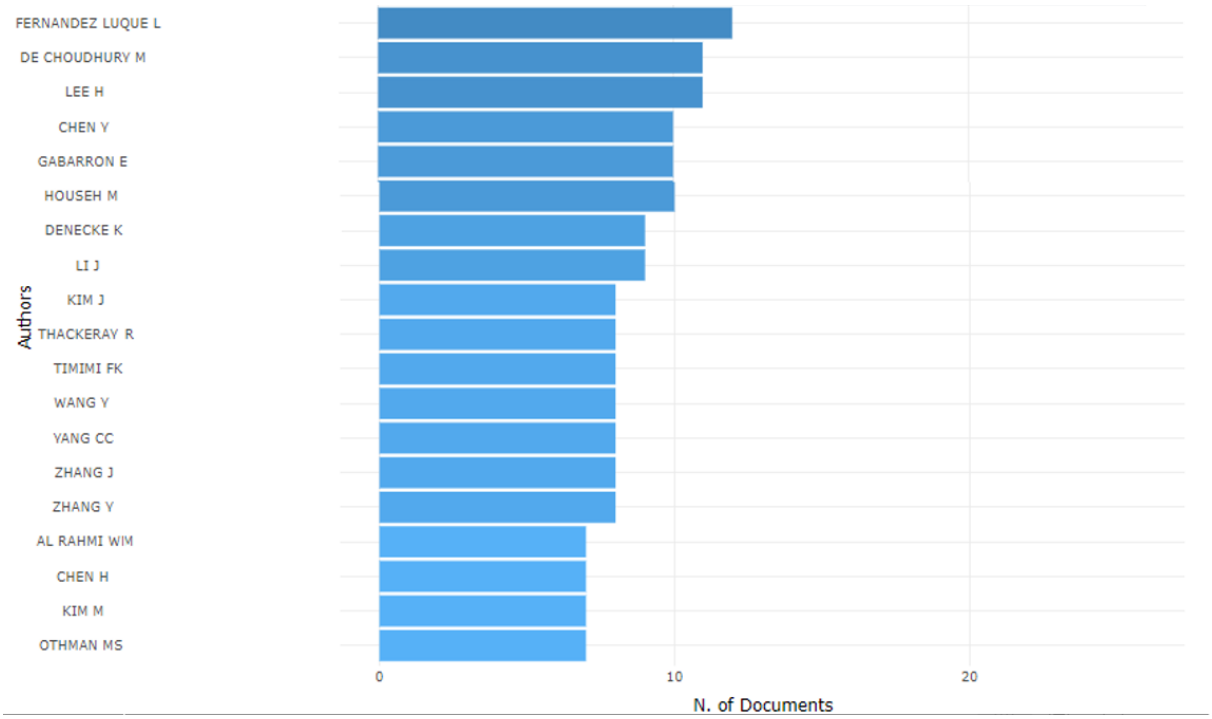
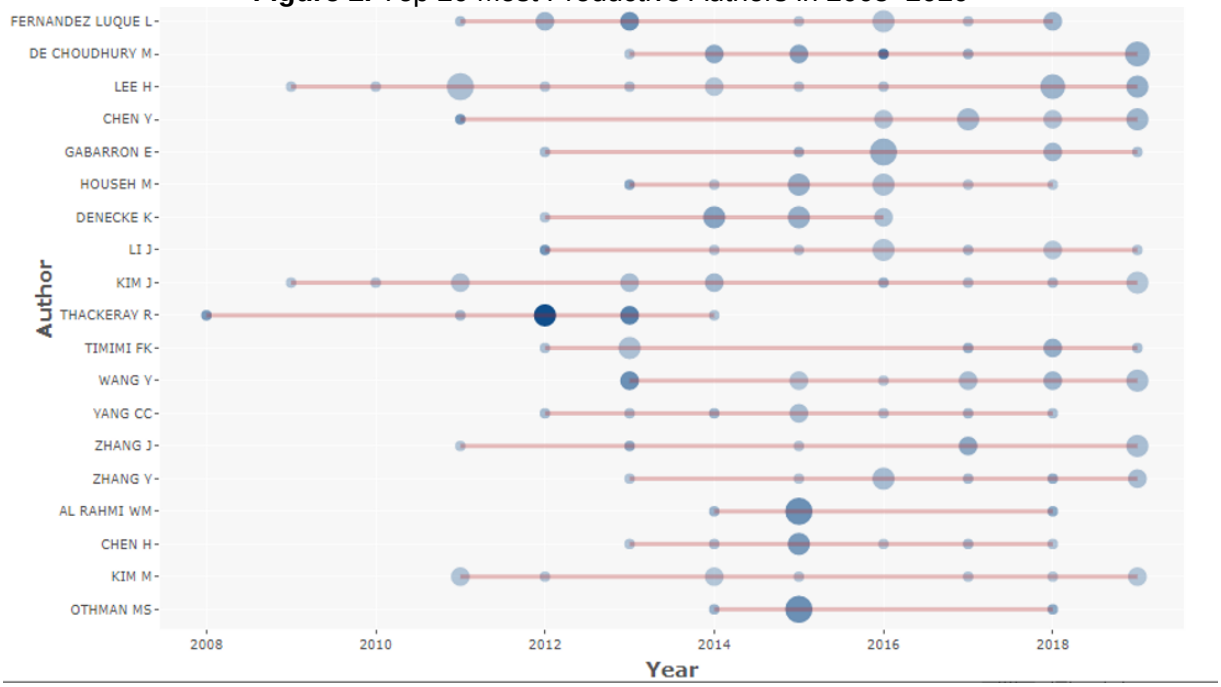


Figure 2. Top 20 Most Productive Authors in 2005–2020



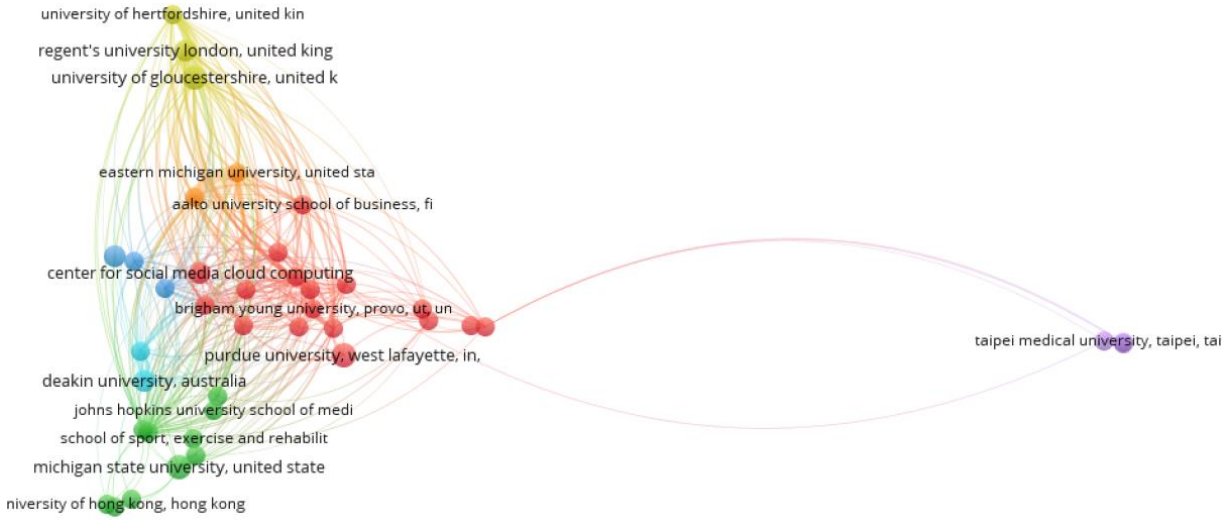
Institution-wise Distribution of Documents

The results showed that authors at 2647 institutions contributed to this subject area, since 2005. The University of Ottawa in Canada provided the highest number of publication (28), followed by the University of California, in the US (23). This analysis is based on the institutional affiliation of the first authors at time of publication. The top 10 institutions are presented in Table 8. Figure 3 also shows a bibliographic coupling between author's institutional affiliation and publication on social media. The results indicate that social media has become a significant research interest of institutions in Western countries in the past few years. However, the impact of social media on businesses is in fact a global phenomenon.

Table 8. Top 10 Institutions Contributing to Social Media Research

Affiliation	Articles	Percentage (%)
University of Ottawa	28	7.41%
University of California	23	6.08%
Deakin University	22	5.82%
Harvard Medical School	22	5.82%
University of Pennsylvania	22	5.82%
Monash University	21	5.56%
Mayo Clinic	20	5.29%
University of Kentucky	20	5.29%
Florida State University	18	4.76%
Michigan State University	18	4.76%

Figure 3. Bibliographic Coupling for Institutions and Publication on Social Media



Analysis of Collaboration

Degree of Correspondence between Authors and Countries

The correspondences between authors from different countries were analyzed to investigate the extent of international collaboration on social media research. Table 9 provides a list of first authors' countries with their single-country and multiple-country publications. Here, the US was the leading country, with a total of 527 publications (470 SCP, 57 MCP, and a 0.1082 MCP ratio). China exhibited a significant MCP ratio of 0.4259. That is, a considerable number of its publications (23 from 31) were MCPs.

Table 9. Top 10 Corresponding First Authors' Countries and Publications.

Rank	Country	Publications	SCP*	MCP*	MCP ratio
1	USA	527	470	57	0.1082
2	Australia	100	69	31	0.3100
3	United Kingdom	94	73	21	0.2234
4	Canada	64	46	18	0.2812
5	Korea	62	45	17	0.2742
6	India	60	59	1	0.0167
7	China	54	31	23	0.4259
8	Germany	47	42	5	0.1064
9	Spain	41	30	11	0.2683
10	Indonesia	37	34	3	0.0811

*Note: SCP = Intra-country collaboration; MCP = Intercountry collaboration

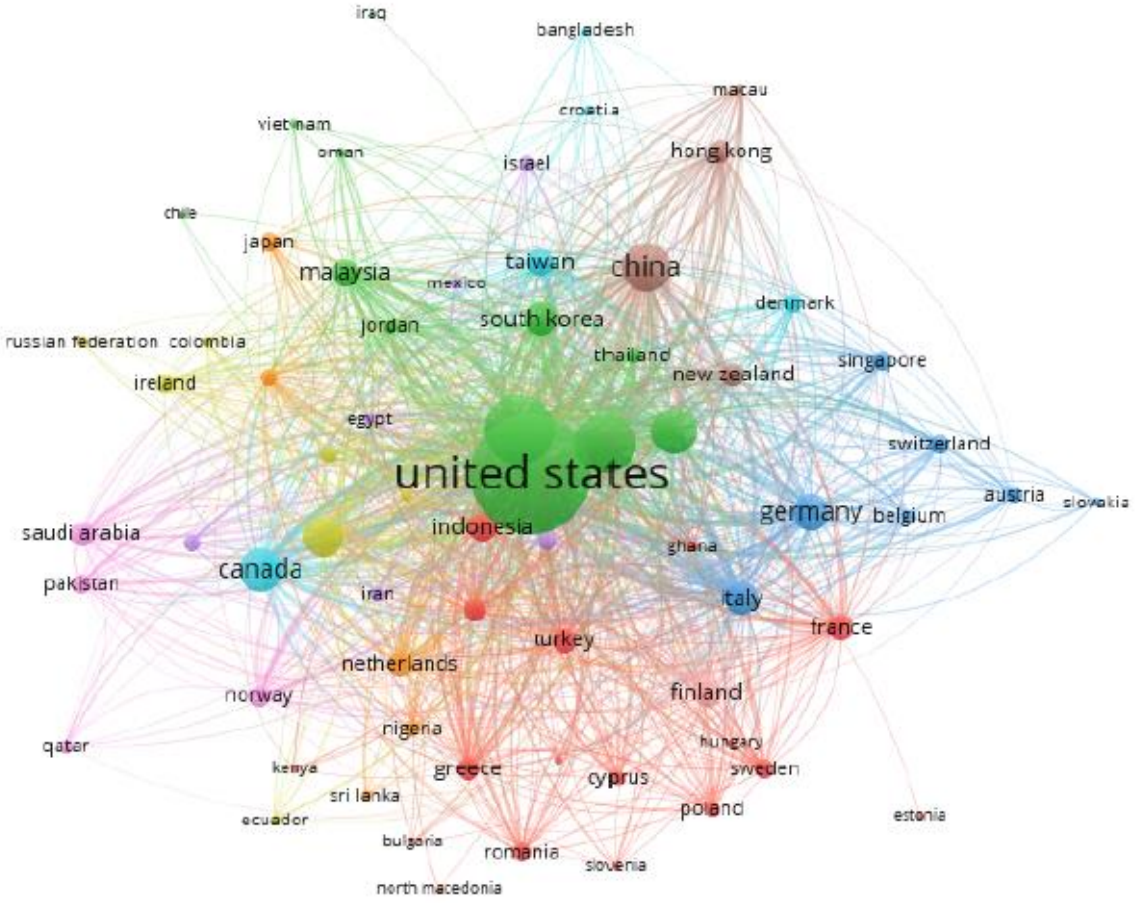
Collaboration between Countries

This analysis explored which countries collaborate more often in the subject area. Table 10 shows country-wise collaboration, where an author from country a collaborates with an author from country b. Australia showed outstanding performance here. It topped the list with 27 documents in collaboration with authors from the US. The next most common collaboration for Australia was with authors from the UK, with 25 publications. Figure 4 presents a bibliographic coupling between countries that research on social media. This provides another representation of the global recognition of the subject.

Table 10. Country Collaboration on Social Media Research

Rank	From country 'a'	To country 'b'	Contributions
1	Australia	USA	27
2	Australia	United Kingdom	25
3	Australia	New Zealand	9
4	Australia	Taiwan	6
5	Australia	Canada	5
6	Australia	France	4
7	Australia	Saudi Arabia	4
8	Australia	Brazil	3
9	Australia	Norway	3
10	Argentina	Australia	2

Figure 4. Bibliographic Coupling of Countries in Publications on Social Media



Citation Analysis

For the 2780 publications investigated, the average citation was 11.3 (citations per document). Local citations were calculated using Biblioshiny, and global citations were determined using a Scopus search. Table 11 indicates the high level of quality of the research on social media, with numerous publications cited often. The highest-ranking publication was Mangold (2009), which described the hybrid nature of promotion over social media. The author provided a conceptualization of the use of social media, that was also used in other studies. This was followed by De Vries (2012), where drivers for brand popularity were determined.

The citation network presented in Figure 5 traces the connection between two documents or more in the top 20 most cited publications to establish a co-citation network on social media. Kaplan and Haenlein (2010) had the highest count in this group. Figure 6 presents the co-citation network of the 50 cited journals with research on social media.

Table 11. Top 10 Cited Publications between 2005 and 2020						
Rank	Author	Source	Year	Local Citations (LCs)	Global Citations (GCs)	Ratio
1	Mangold Wg	Business Horizons	2009	180	1318	0.14
2	De Vries L	Journal of Interactive Marketing	2012	56	612	0.09
3	Moorhead Sa	Journal of Medical Internet Research	2013	76	567	0.13
4	Chou Wys	Journal of Medical Internet Research	2009	65	521	0.12
5	Leung D	Journal of Travel and Tourism Marketing	2013	64	391	0.16
6	Berthon Pr	Business Horizons	2012	39	357	0.11
7	Michaelidou N	Industrial Marketing Management	2011	42	331	0.13
8	Munar Am	Tourism Management	2014	32	272	0.12
9	Hays S	Current Issues in Tourism	2013	39	260	0.15
10	Korda H	Health Promotion Practices	2013	36	262	0.14

Figure 5. Co-citation Network of Top 20 Cited Publications between 2005 and 2020

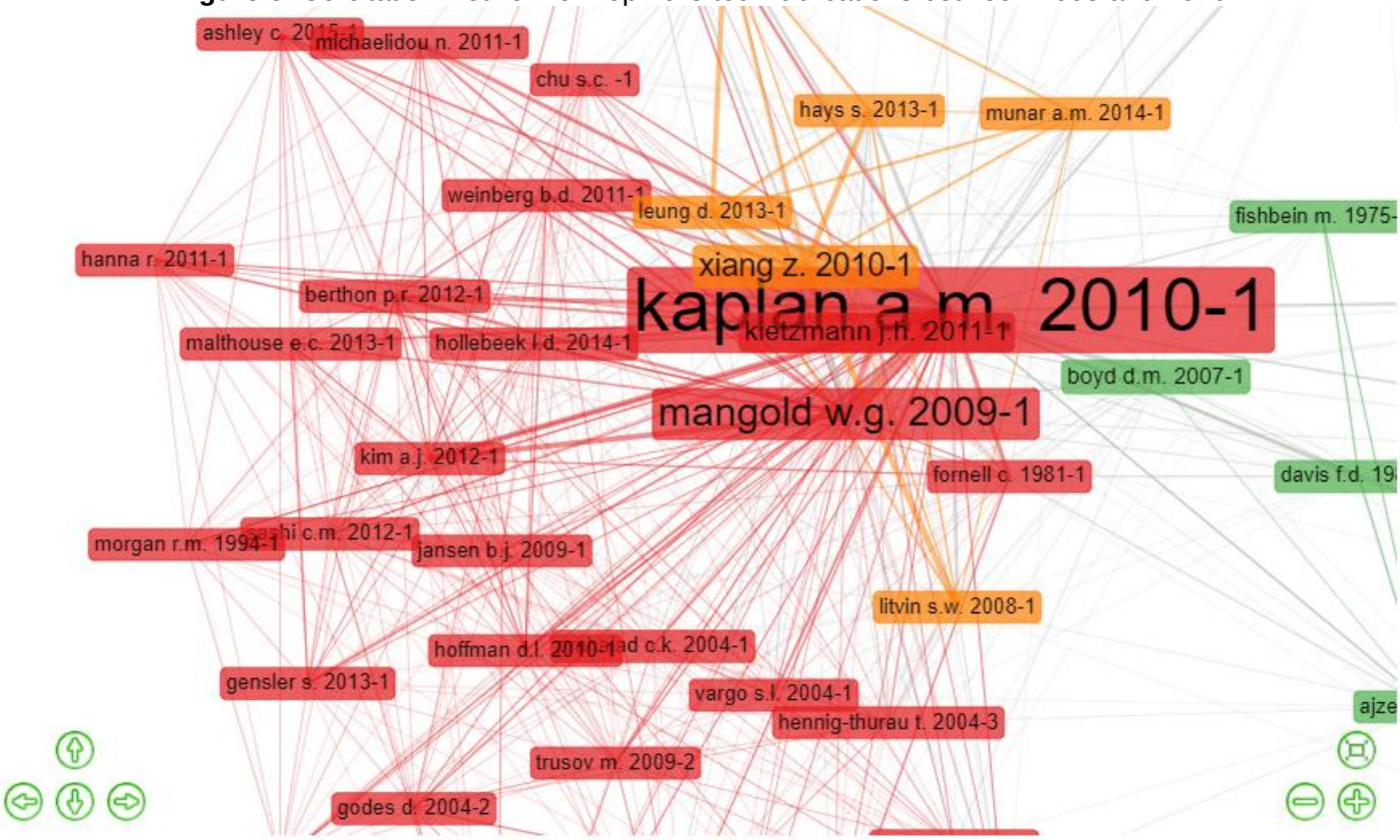
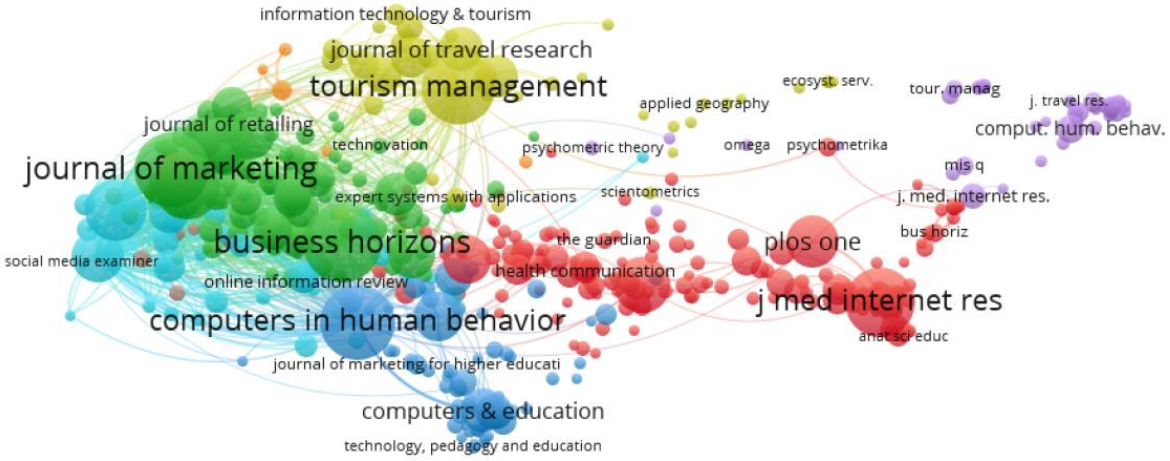


Figure 6. Co-citation Network of Top 50 Journals between 2005 and 2020



Analysis of Keywords

I analyzed both keyword plus (ID) and authors' keyword (DE), as given previously in Table 2, to illustrate trends in research on social media. It was found that 6008 keywords (measureed using ID) were used in the set from 2005 to 2020. The keyword *social media* was the most common in this subject area, with 1916 occurrences, followed by *social networking online* (689 occurrences) and *human* (666 occurrences). Table 12 provides the list of the top 10 most-used keywords on social media.

Table 12. Top 10 Keywords Plus on Social Media between 2005 and 2020

Rank	Keyword Plus (ID)	Occurrences	Percentage (%)
1	social media	1916	26.27%
2	social networking online	689	9.45%
3	human	666	9.13%
4	humans	491	6.73%
5	internet	347	4.76%
6	female	326	4.47%
7	male	307	4.21%
8	article	304	4.17%
9	education	275	3.77%
10	adult	269	3.69%

Finally, I analyzed the most common keywords used by authors to describe their research. Figure 7 presents the co-occurrence of author keywords, with a threshold of five occurrences.

Social media, Twitter, marketing, and mental health were the most common keywords used in this subject. To better view the leading author keywords, Table 13 presents a list of the top 20 keywords of all-time. The results are in accordance with Figure 7.

Figure 7. Co-occurrence of Author Keywords of Documents Published between 2005 and 2020

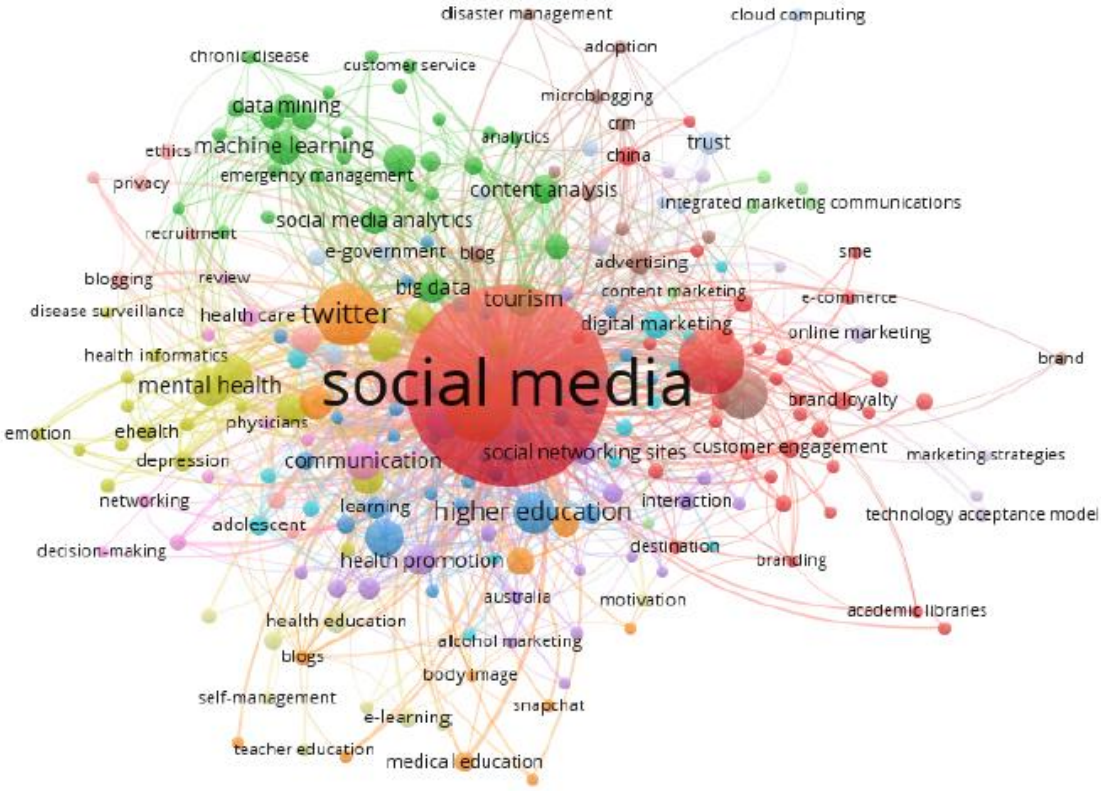


Table 13. Most Common Author Keyword Occurrences in Social Media Research										
R	ID	Keyword	Occurrences	Total Link Strength		R	ID	Keyword	Occurrences	Total Link Strength
1	3608	social media	1262	2021		11	3066	public health	45	105
2	1284	Facebook	141	390		12	536	communication	42	116
3	3670	social media marketing	141	194		13	1538	health	37	90
4	4274	Twitter	121	359		14	3773	social networks	37	86
5	2172	Marketing	82	187		15	2138	machine learning	35	88
6	1918	Internet	71	225		16	4151	tourism	35	67
7	4466	web 2.0	61	154		17	3389	sentiment analysis	33	55
8	1685	higher education	58	127		18	3763	social networking	33	104
9	2289	mental health	46	110		19	4046	technology	33	103

10	1105	Education	45	107		20	1568	health communication	32	97

RESEARCH IMPLICATIONS

The degree of prominence of social media in marketing strategies depends on the local markets present in a society; specifically, patterns of use of social media depend on technological infrastructure, culture of internet use, and local regulations – Berthon *et al.* (2012). These factors both drive the motivation to create and exchange content over social networks and the methodologies for successful social marketing – Carins and Rundle-Thiele (2014). The expressions of a culture depend predominantly on its intrinsic nature, whether individualist/independent/separate or collectivist/interdependent/connected – Okazaki and Mueller (2007).Interdependent societies, for example, may be more promising sites for social networking than individualist ones. Furthermore, based on the agency-communion theory – Zarbatany *et al.* (2004); the scope of a social marketing network varies with the type of friendship between individuals, i.e., communal or agentic. The former is based on mutual past experience, intimacy, and loyalty, while the latter is governed by mastery, validation, and ego.

- Can a conceptual model be developed that prescribes the role of individual and collective cultures in social media marketing?
- How can agentic and communal groups be targeted in social media advertising?
- What brand images are most popular among customers in the context of self-schema theory?
- Can we develop an empirical model for advertising through games, with factors that influence youth from different cultures?
- What social, economic, and technological factors are most suited to fostering social media advertising through mobile applications?

The results of this study may have practical implications for industries that use social media platforms such as:

- *Facebook* – where users unwind and recollect with their companions. The promotor must adopt a lighter and inviting tone here because natural reach can be very restricted.
- *Google+* – where users share their photographs, video and sound recordings, and connections on. This may enable a business to fragment their supporters into little gatherings for the sake of imparting data to certain adherents.
- *Pinterest* – which offers a picture-focused stage that is perfect for retail promotions. It enables organizations to create eye-catching brand character through pinboards.
- *Twitter* – where organizations spread internet-based live advertisements and exchange the communications that can help them establish and sustain a following. Pursuing tweets in their industry or in related fields ensures an increased flow of adherents.
- *LinkedIn* – that provides an extraordinary setting for expert exchange of content with individuals in similar businesses. The suggestions provided by this platform can cause your business to show up more extensively.
- *YouTube* – which is a popular medium for making and sharing video content. The visual clips it hosts can be sensational if they are informative, valuable, and up-to-date.

CONCLUSIONS

A general bibliometric overview of social media marketing in several industries is presented. The results are in line with the observations and opinions of the academic community. Most numerical and graphical results on journals, authors, and articles are quite significant for understanding the state of research into social media.

The subject area has increased exponentially: in 2005, 1 paper was published, and in 2019, 486 were. The US led in research output. The most prolific author, Fernandez Luque, worked at Hamad Bin Khalifa University. The University of Ottawa in Canada had the most contributions. Most affiliation between authors on the subject was seen among Western countries, and the US led the list with most intra-country collaborations. Australia led the intercountry collaborations.

Mangold and Faulds (2009) received the most citations, and Kaplan and Haenlein (2010) led the research network with the highest co-citations. *Journal of Marketing* and *Business Horizons* were the top cited journals in this subject are. *Social media* and *social networking* online were the most common keywords in the documents while most authors used the keywords as *social media*, *Twitter*, *marketing*, and *mental health* with their abstracts.

REFERENCES

Aria, M. and Cuccurullo, C. (2017), "bibliometrix: An R-tool for comprehensive science mapping analysis", *Journal of Informetrics,* Vol. 11, No. 4, pp. 959-975.

Assefa, S.G. and Rorissa, A. (2013), "A bibliometric mapping of the structure of STEM education using co-word analysis", *Journal of the American Society for Information Science and Technology,* Vol. 64, No. 12, pp. 2513-2536.

Baier-Fuentes, H., Merigó, J.M., Amorós, J.E. and Gaviria-Marín, M. (2019), "International entrepreneurship: a bibliometric overview", *International Entrepreneurship and Management Journal,* Vol. 15, No. 2, pp. 385-429.

Berthon, P.R., Pitt, L.F., Plangger, K. and Shapiro, D. (2012), "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy", *Business horizons,* Vol. 55, No. 3, pp. 261-271.

Busca, L. and Bertrandias, L. (2020), "A Framework for Digital Marketing Research: Investigating the Four Cultural Eras of Digital Marketing", *Journal of Interactive Marketing,* Vol. 49, pp. 1-19.

Campbell, C., Ferraro, C. and Sands, S. (2014), "Segmenting consumer reactions to social network marketing", *European Journal of Marketing,* Vol. 48, No. 3/4, pp. 432-452.

Carins, J.E. and Rundle-Thiele, S.R. (2014), "Eating for the better: A social marketing review (2000–2012)", *Public health nutrition,* Vol. 17, No. 7, pp. 1628-1639.

Darvish, H. (2018), "Bibliometric analysis using Bibliometrix an R Package", .

De Caigny, A., Coussement, K., De Bock, K.W. and Lessmann, S. (2019), "Incorporating textual information in customer churn prediction models based on a convolutional neural network", *International Journal of Forecasting,* .

Ding, Y., Chowdhury, G.G. and Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis", *Information processing & management,* Vol. 37, No. 6, pp. 817-842.

Diodato, V.P. & Gellatly, P. 2013, *Dictionary of bibliometrics,* Routledge.

Ekundayo, T.C. and Okoh, A.I. (2018), "A global bibliometric analysis of Plesiomonas-related research (1990–2017)", *PloS one,* Vol. 13, No. 11, pp. e0207655.

Ellegaard, O. and Wallin, J.A. (2015), "The bibliometric analysis of scholarly production: How great is the impact?", *Scientometrics,* Vol. 105, No. 3, pp. 1809-1831.

Harinarayana, N. (2019), "Data Sources and Software Tools for Bibliometric Studies", *Online].Available: https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/library_and_information_science/informetrics _&_scientometrics/data_sources_and_software_tools_for_bibliometric_studies/et/333_et_m2.pdf.Accessed on,* Vol. 12, .

Huang, L., Clarke, A., Heldsinger, N. and Tian, W. (2019), "The communication role of social media in social marketing: a study of the community sustainability knowledge dissemination on LinkedIn and Twitter", *Journal of Marketing Analytics,* Vol. 7, No. 2, pp. 64-75.

Jalali, M.S., Razak, S., Gordon, W., Perakslis, E. and Madnick, S. (2019), "Health care and cybersecurity: bibliometric analysis of the literature", *Journal of medical Internet research,* Vol. 21, No. 2, pp. e12644.

Jin, S.A. and Phua, J. (2014), "Following celebrities' tweets about brands: The impact of twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities", *Journal of Advertising,* Vol. 43, No. 2, pp. 181-195.

Jovicic, D.Z. (2019), "From the traditional understanding of tourism destination to the smart tourism destination", *Current Issues in Tourism,* Vol. 22, No. 3, pp. 276-282.

Kannan, P. and Thanuskodi, S. (2019), "Bibliometric Analysis of Library Philosophy and Practice: A study based on Scopus Database", *Library Philosophy and Practice,* pp. 1-13.

Kannan, P. (2017), "Digital marketing: A framework, review and research agenda", *International Journal of Research in Marketing,* Vol. 34, No. 1, pp. 22-45.

Kaplan, A.M. and Haenlein, M. (2010), "Users of the world, unite! The challenges and opportunities of Social Media", *Business horizons,* Vol. 53, No. 1, pp. 59-68.

Khan, A., Choudhury, N., Uddin, S., Hossain, L. and Baur, L. (2016), "Longitudinal trends in global obesity research and collaboration: a review using bibliometric metadata", *Obesity Reviews,* Vol. 17, No. 4, pp. 377-385.

Kim, W.G. and Park, S.A. (2017), "Social media review rating versus traditional customer satisfaction: which one has more incremental predictive power in explaining hotel performance?", *International Journal of Contemporary Hospitality Management,* Vol. 29, No. 2, pp. 784-802.

Kim, Y., Kim, C., Lee, D.K., Lee, H. and Andrada, R.I.T. (2019), "Quantifying nature-based tourism in protected areas in developing countries by using social big data", *Tourism Management,* Vol. 72, pp. 249-256.

Liu, Y. and Li, M. (2016), "BibeR: A Web-based tool for bibliometric analysis in scientific literature", *PeerJ PrePrints,* Vol. 4, pp. e1879v1.

Majumdar, A. and Bose, I. (2019), "Do tweets create value? A multi-period analysis of Twitter use and content of tweets for manufacturing firms", *International Journal of Production Economics,* Vol. 216, pp. 1-11.

Mangold, W.G. and Faulds, D.J. (2009), "Social media: The new hybrid element of the promotion mix", *Business horizons,* Vol. 52, No. 4, pp. 357-365.

Mattsson, P., Sundberg, C.J. and Laget, P. (2011), "Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position", *Scientometrics,* Vol. 87, No. 1, pp. 99-105.

Meho, L.I. and Rogers, Y. (2008), "Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science", *Journal of*

the American Society for Information Science and Technology, Vol. 59, No. 11, pp. 1711-1726.

Merigó, J.M. and Yang, J. (2017), "A bibliometric analysis of operations research and management science", *Omega,* Vol. 73, pp. 37-48.

Nair, S., Iamnitchi, A. and Skvoretz, J. (2019), "Promoting Social Conventions across Polarized Networks: An Empirical Study", .

Okazaki, S. and Mueller, B. (2007), "Cross-cultural advertising research: where we have been and where we need to go", *International Marketing Review,* Vol. 24, No. 5, pp. 499-518.

Osareh, F. (1996), "Bibliometrics, citation analysis and co-citation analysis: A review of literature I", *Libri,* Vol. 46, No. 3, pp. 149-158.

Potter, W.G. (1981), "Introduction to bibliometrics", *Library trends,* Vol. 30, No. 5, .

Rialp-Criado, J. and Rialp, A. (2020), "Speed of Use of Social Media as an Antecedent of Speed of Business Internationalization", *Journal of Global Information Management (JGIM),* Vol. 28, No. 1, pp. 142-166.

Schweidel, D.A. and Moe, W.W. (2014), "Listening in on social media: A joint model of sentiment and venue format choice", *Journal of Marketing Research,* Vol. 51, No. 4, pp. 387-402.

Slattery, P., Finnegan, P. and Vidgen, R. (2019), "Creating compassion: How volunteering websites encourage prosocial behaviour", *Information and Organization,* Vol. 29, No. 1, pp. 57-76.

Small, H. (1973), "Co-citation in the scientific literature: A new measure of the relationship between two documents", *Journal of the American Society for information Science,* Vol. 24, No. 4, pp. 265-269.

Sweileh, W.M. (2018), "Bibliometric analysis of peer-reviewed literature in transgender health (1900–2017)", *BMC international health and human rights,* Vol. 18, No. 1, pp. 16.

Tang, M., Liao, H., Wan, Z., Herrera-Viedma, E. and Rosen, M. (2018), "Ten years of sustainability (2009 to 2018): A bibliometric overview", *Sustainability,* Vol. 10, No. 5, pp. 1655.

Uribe-Toril, J., Ruiz-Real, J.L., Haba-Osca, J. and de Pablo Valenciano, J. (2019), "Forests' First Decade: A Bibliometric Analysis Overview", *Forests,* Vol. 10, No. 1, pp. 72.

Van Eck, N.J. and Waltman, L. (2017), "Citation-based clustering of publications using CitNetExplorer and VOSviewer", *Scientometrics,* Vol. 111, No. 2, pp. 1053-1070.

Wei, G. (2019), "A bibliometric analysis of the top five economics journals during 2012–2016", *Journal of Economic Surveys,* Vol. 33, No. 1, pp. 25-59.

Zarbatany, L., Conley, R. and Pepper, S. (2004), "Personality and gender differences in friendship needs and experiences in preadolescence and young adulthood", *International Journal of Behavioral Development,* Vol. 28, No. 4, pp. 299-310.

DECISION SCIENCES INSTITUTE

Semantic Search in a Hyperlinked Environment

Rakesh Kulkarni.
Indian Institute of Technology Bombay.
Address: #601, MANAS, Hill Side,
IIT Campus, Powai,
Mumbai – 400076
India
E-mail address: kulkarni.rakesh1@gmail.com
Phone: +91-908 219 9091

ABSTRACT

The perfect search engine should return results based on the meaning inherent in the documents to be able to give the best possible results to the user. This goes beyond simple word matching and includes accounting for polysemy and synonymy, which link based algorithms such as PageRank fail to do. The meaning or the semantic content of documents and their importance in a corpus must be considered in search. I propose novel algorithms and run experiments on a comprehensive, real world dataset to compare results with the PageRank algorithm and provide empirical evidence for the supremacy of my algorithms.

KEYWORDS: Text Analytics, Search Engines, Semantic Analysis, Web Search, and Information Retrieval.

INTRODUCTION

With the rapid explosion of data and data storage & processing in the information age, information retrieval (IR) or “Search” assumes great significance. Text mining being a major aspect of IR, this work focuses on this particular aspect. Text mining algorithms are being used in a plethora of applications both within academic and industrial research. The importance of quality and relevance of the output of these algorithms cannot be stressed enough.

The most profound application of text mining in recent years has been the search engine. Poor quality of results from web search engines are frustrating and time consuming for the users and are the biggest problem facing such users (Brin and Page 2012). Google uses information contained in the hyperlinks and exploits the link structure of the web in producing matching results for a query. PageRank of a web page is an 'importance' metric that stems from this link structure and helps in ranking the results. The web is treated as a Markov Chain and its stationary distribution calculated in ranking the web pages. While most of the popular web search engines use algorithms based on rules such as proximity of words and the link structure of the web, there exists on the web tremendous amounts of information in the text of the web pages apart from hypertext, link structure and link text. Importance of web pages with respect to a query may also be evaluated in the absolute sense, and not based on their relative

importance as done by Google's PageRank algorithm, based on how closely their contents match the meaning intended in the search query. The PageRank algorithm as defined (Page et al., 1999) uses a rule based model to match a query with previously ranked web pages. The rule based models suffer from their inability to cope with two classic problems arising in natural language processing: synonymy and polysemy. Synonymy is two or more different words in a language having the same meaning (say, car and automobile). Polysemy refers to multiple meanings attributed to the same term (for example, 'charge' could mean electric charge or the amount charged for a product or service). Here the computed similarity is overestimated (Manning et al., 2008).

Most Natural Language Processing (NLP) approaches do not consider the semantic relationships within the grammatical information. It has been demonstrated that by relying on semantic discourse, pairs of sentences with different keywords can be similar and those with almost identical keywords can be very different. This semantic discourse level analysis is necessary in answering complex questions (Galitsky, 2017). Web Search can be improved by using Semantic knowledge along with linguistic knowledge (Conesa et al., 2008). Purely text based search engines are not focused on domain knowledge and hence do not understand the inherent relationships between a query's terms or the knowledge latent in a knowledge base (Fang, Wei-Dong, et al., 2005). Similarity based on the semantic content of whole documents, rather than merely author chosen titles and key words is an exciting prospect. Semantic models have been used to build recommendation systems based on content, since user feedback is not available on newly introduced service websites (Liu et al., 2013). Stock returns have been predicted by determining the sentiments tweeted (Sul et al., 2017). The dictionary based approach in mining twitter, although, is imperfect in measuring sentiment since it does not detect the meaning inherent in the semantic groups of words. The concept of Semantic Web is expected to bring structure to the meaning contained in Web pages. The highest ideal of the Semantic Web would be the evolution of human knowledge (Berners et al., 2001).

Concepts of Semantic Web or Semantic Networks still involve a lot of preprocessing using experts' opinions. After results for a query, such as documents containing the search words are obtained, lexical analysis techniques are used to filter the results so that only documents which use the search terms, according to the selected word sense are returned mimicking a dictionary (Leyba, 2016). A fully automated Semantic Web search, without manual intervention is thus the need of the hour. The means to this end can be found in the broad area of probabilistic topic modeling.

This study uses a novel methodology to exploit the advantages of both the probabilistic topic models and the graph based models while addressing the drawbacks of both. I propose a Semantic PageRank algorithm and study its performance on a real world anthology of 18,164 papers from the Association of Computational Linguistics (ACL). I compare the results of this algorithm with the basic PageRank algorithm and show that the semantic PageRank algorithm gives outstanding results. I begin by describing probabilistic topic modeling in the following section.

LITERATURE REVIEW

Topics or semantic structures that occur in a collection can be modeled by experts in a specific area or by the use of automated techniques like Probabilistic topic modeling. In enterprise search, a hierarchy of subject metadata terms is available and this is used with the text search

process to improve the semantic search performance of an enterprise search engine (Schymik, Gregory, et al., 2015). Probabilistic topic modeling has been defined by Blei et al as providing "methods for automatically organizing, understanding, searching, and summarizing large electronic archives" (Blei et al., 2003). Probabilistic topic models include Latent Semantic Analysis (LSA), an extension of LSA - PLSA (Probabilistic LSA) (Hofmann 1999) and Latent Dirichlet Allocation (LDA) (Blei et al., 2003).
LSA has advantages over PLSA to the extent that while the Singular Value Decomposition (SVD) is computed exactly, PLSA uses an iterative Expectation Maximization algorithm which is limited to finding the local maximum of the likelihood function (Hofmann 2001). The Gibbs Sampling algorithm is used to approximate the LDA model. In LDA, all the topics do not have equal probability. LDA is known to become very expensive computationally on large data sets (Mikolov et al., 2013).
LSA is a probabilistic topic model within which a low rank approximation is used due to which terms/documents are clustered tightly together with co-occurring terms along each of the k-dimensions. In LSA, documents from a given topic become closely clustered together based on the similarity of words in a transformed eigenspace, irrespective of the exact same words occurring in each of the documents. Thus, the retrieval quality may actually improve with the loss in the data being composed of noise and not information. Using literature from academic journals (3207 abstracts) related to the Operations Management, LSA was used to determine trends and patterns exhibited by the area from 1980 to 2012 (Kulkarni et al., 2014). Dumais (Dumais, 1993) and Berry et al., (Berry et al., 1995) conducted experiments with LSA on Text Retrieval Conference (TREC) data. Their algorithm performed at or better than the median on metrics like precision. In close to a fifth of TREC topics, they were the leaders among all participants. Their conclude that LSA addresses the challenges of synonymy and polysemy.
The PageRank algorithm is now briefly described followed by a description of the steps involved in performing LSA.

The PageRank (PR) Algorithm

The PageRank of a page u is given as follows (Page et al., 1999):
Let, Fu = the set of pages u points to and
Bu = the set of pages that point to u.
Nu = $\left|Fu\right|$ = the number of links from u
c = a factor used for normalization.

Begin by defining which is a slightly simplified version of PageRank R:

$$R(u) = c \sum_{v \in Bu}^{n} R(v)/Nv \quad (1)$$

The n equations can be written by treating R as a vector over web pages. A being the adjacency matrix, we have R = cAR. So R is an eigenvector of A with eigenvalue c. The goal being to get the dominant eigenvector of A, which can be computed by repeatedly applying A to any nondegenerate initial vector, with d being a damping factor which is usually set to 0.85. Let S be any vector over Web pages (for example E). Then PageRank can be computed using the following iterative algorithm:

$R0 \leftarrow S$

loop :

$R_{i+1} \leftarrow AR_i$

$d \leftarrow \|R_i\|_1 + \|R_{i+1}\|_1$

$R_{i+1} \leftarrow R_{i+1} + dE$

$\delta \leftarrow \|R_{i+1} - R_i\|_1$

while $\delta > \varepsilon$

The d factor increases the rate of convergence and maintains $\|R_i\|_1$ (Page et al., 1999).

The drawback of the PageRank algorithm is that it does not consider the rich semantic information contained within the text. Much emphasis is laid on exploiting the link structure after using a simple text search. The PR algorithm thus fails to deal with Polysemy and Synonymy. The use of meaning inherent within text in conjunction with the link structure is highly recommended in producing quality search results (Page et al., 1998, Kleinberg, 1998, Moran and Lempel, 2001). A third problem facing the IR community on the web is one of Search Engine Persuasion whose genesis lies in the way the PR algorithm works. Search engine persuasion can be defined as inflating the rank of a web page with respect to certain key words by manipulating its content to and by increasing its inbound links.

Having discussed one aspect of my formulation – the graphical model, let us now look at the second aspect, which is topic modeling. LSA is a probabilistic topic modeling technique and is useful in finding the meaning latent in the text. LSA is discussed briefly in the following section.

LSA

Latent Semantic Analysis as used in this approach involves the following steps:

Step 1: Compilation of Term Frequency Matrix (Vector Space Model)

A weight is assigned to each term in a given document based on the frequency of the term *t* in the document *d*. This enables us to calculate a score between a query term and a document, based on this. The weighed terms are known as term frequency denoted by *tf. df* or document frequency is the count of documents in the corpus where *t* occurs. *N* being the total number of documents in a collection, the inverse document frequency (*idf*) of a term *t* can be defined as:

$$idf = \log \frac{N}{df} \tag{2}$$

A composite weight for each term *t* in a document *d* using *tf* and *idf looks as under*:

$$tf - idf_{t,d} = tf_{t,d} X\, idf_t \tag{3}$$

Multiple variants of term frequency can be created. The logarithm of the term frequency can be used, and assign the weight as:

$$Weight = \begin{cases} 1+\log tf_{t,d} \; if \; tf_{t,d} > 0 \\ 0 \; otherwise \end{cases} \quad (4)$$

Step 2: Dimensionality Reduction Using Reduced Singular Value Decomposition (RSVD)

Singular Value Decomposition (SVD) transforms correlated variables into a set of uncorrelated ones while exposing the relationships among the original data items in an enhanced manner. It is also a method used in identifying and ordering the dimensions along which maximum variation in the data is present. SVD as a method for data reduction (Baker 2005), is particularly useful in text mining applications where we can ignore the variation below a threshold to massively reduce the data while assuring ourselves that the main relationships of interest have been preserved.

Dimensionality is reduced by restricting the transformed matrix to the first 'k' singular values. This is called Reduced Singular Value Decomposition (RSVD). We lose noise and not information from the original data as we delete dimensions along which little meaningful variation is explained. That is, similar items from the original data become more similar to each other and dissimilar items become more dissimilar leading to documents from a given topic becoming closely clustered together based on the similarity of words in this eigen space.

Step 3: Choice of k

Selecting the right number of dimensions is an open problem. The elbow point from the singular values scree plot as an indicator for the number of singular values to be considered is often suggested (Kulkarni et al., 2014). Other options include looking at the proportion of the variation explained by the corresponding eigen values, in trying to determine the "cutting point".

The term frequency matrix is an index of documents. LSA applied in this manner in search is simply called Latent Semantic Indexing (LSI). A new low-rank representation is given for each document in the collection by the reduced matrix from the previous step. Search queries are now matched with documents in this reduced space.

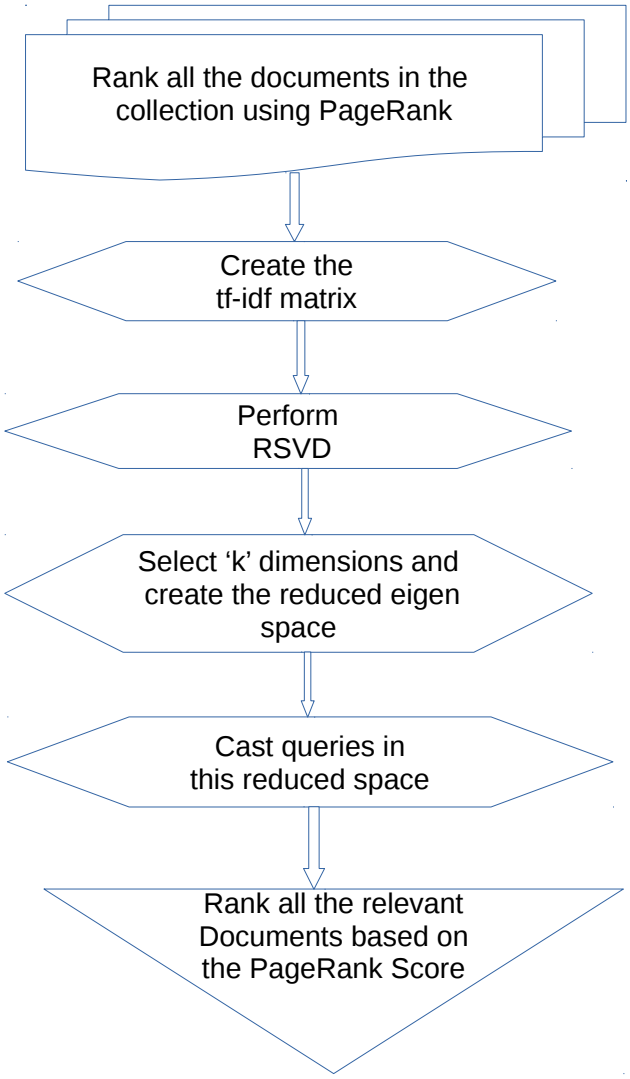
LSI solves the problems of Synonymy and Polysemy, but assigns the same relevance score to multiple pages. A ranking of these relevant results is needed. This can be achieved based on the notion of "importance". The importance of pages is determined by using the link structure of the web. By relying on importance, the PageRank (PR) algorithm assigns an objective score to the page as opposed to the subjective ranking based on relevance. Intuitively, a combination of these two approaches should help us rank search results better. An algorithm that uses LSA in conjunction with the PR algorithm can be called the *Semantic PageRank algorithm* and the same is described below.

Semantic PageRank algorithm

The implementation of this algorithm is as follows:

I use the PR algorithm described in the previous sections to rank all the documents. The LSI procedure is then applied on the corpus of documents to force the similar words/documents to be closest to each other by using fewer dimensions in a reduced eigenspace. The query is cast into this reduced eigenspace and the results are the top 't' highest values in the principal eigen vector. I use this score in conjunction with the PR score determined earlier. Thus, we are able to rank the *relevant* search results based on the importance score (PR Score). This process is better explained in the flowchart in Figure 1:

Figure 1: Flowchart of the proposed algorithm



By using Semantic PR, we are simultaneously able to:
1. Solve the problems of Synonymy and Polysemy using LSI which are not solved using a purely text based PR algorithm.
2. Rank the multiple relevant pages resulting from LSI using PR.
3. Handle the Search engine persuasion problem since PR alone takes care of unimportant pages based on the links and the semantic aspect from LSI ensures that misleading pages with innumerable hyperlinks alone are not given high relevance scores.
The performance of both the base PR algorithm and the Semantic PR algorithm was tested by me comprehensively on real world data.

EXPERIMENTAL EVALUATION

This section discusses the results from experiments conducted using the PageRank algorithm with pure text based search and the semantic PageRank algorithm. It is important to note that I

compare this algorithm with the PageRank algorithm described by Page et al. (Page et al., 1999) since it is the only published resource describing the algorithm behind Google. I also do not consider other improvements to the PageRank algorithm such as those proposed by Nie et al., (Nie et al., 2006) or by Jardine and Teufel (Jardine and Teufel, 2014). The LDA method used in both these papers is known to become very expensive computationally on large data sets (Mikolov et al., 2013) and its applicability to web search thus becomes questionable. The ACL anthology network corpus (Radev et al., 2009) has been used for this study and is described below.

DATA DESCRIPTION

The ACL Anthology Network (AAN) is a corpus of all papers published by ACL, its related organizations and the Computational Linguistics journal over 40 years. I use this dataset since it contains both the corpus of papers which can be looked at as web pages and the citation network between the papers which has been created manually by ACL, to be used as the Markov Chain. The paper citation network is a directed network in which each node represents a paper labelled with an ACL ID number and edges represent citations between papers. The paper citation network consists of 18,164 papers (Nodes), 14,799 authors, 341 venues, 110,975 paper citations (Edges). Many papers include references to papers outside of the network. These edges have been dropped and disconnected nodes have not been included.

RESOURCES USED

For Coding and Analysis, I relied on VBA, R and SAS. 3 servers were used for processing the code, each with the following configuration:

Table 1: Configuration of the computing resources used

CPU	NUMBER OF CORES	RAM (GB)
Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	20	128

To replicate the original PR algorithm, and to make my results comparable with it, I use the Paper Titles as documents to match my queries with.

NO GOLD STANDARD

There is consensus in the Information Retrieval community of a lack of a gold standard for measuring search results (Kleinberg, 1998, Page et al., 1998, Moran and Lempel, 2001, Manning et al., 2008). The search results must be evaluated subjectively.

The top 8 results of the PR algorithm for the search term “Summarization” are given in Table 2. Summarization is one of the 3 most important key words (phrases) in NLP (Radev et al., 2009), the other two being “Machine Translation” and “Dependency Parsing”.

The first column in Table 2 gives the results from the PR algorithm matching the query term “Summarization”. Relevance Score here is the ratio of the number of occurrences of the search term in the title to the total number of words in the title. The PR score in column 3 is the global score of importance given to the web page by the PR algorithm. Column 4 gives the PageRank

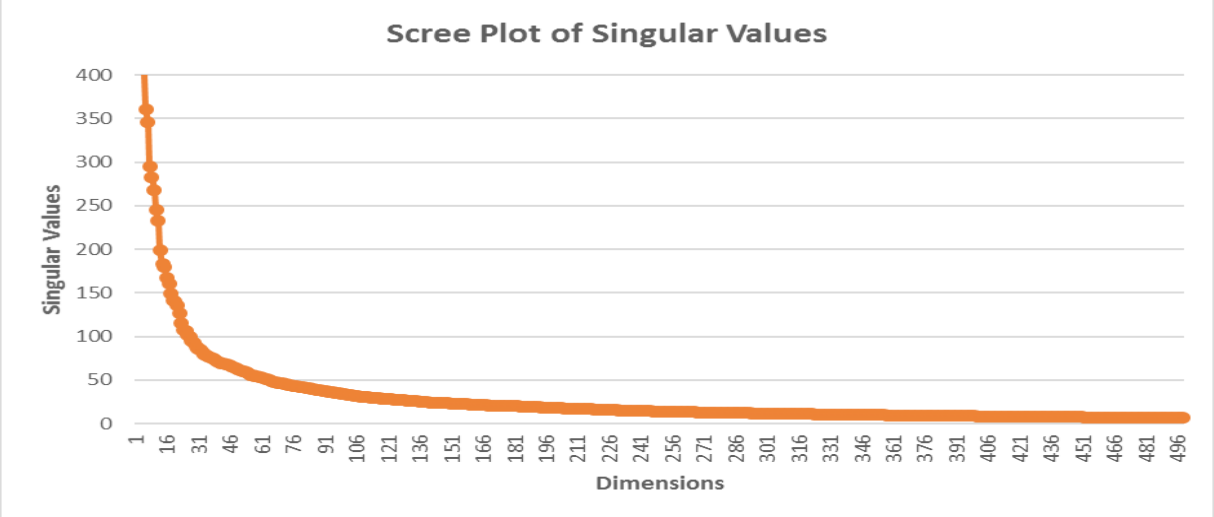
of each relevant page sorted by the PR Score. It can be seen that all the results except the first one have the same relevant score and the PR Score helps us rank them in the order of importance.

Table 2: Top 8 results for the query “Summarization” from the PR algorithm

PAPER TITLE	RELEVANCE SCORE	PR SCORE	PAGERANK
Automatic Summarization	0.5	0.0000128	1
Bayesian Query-Focused Summarization	0.33	0.0000883	2
Event-Based Extractive Summarization	0.33	0.0000605	3
Multi-Document Biography Summarization	0.33	0.0000471	4
Template-Filtered Headline Summarization	0.33	0.0000159	5
Multi-Topic Multi-Document Summarization	0.33	0.0000143	6
Automated Text Summarization	0.33	0.0000137	7
Helpfulness-Guided Review Summarization	0.33	0.0000128	8

To build semantic models, reduced singular value decomposition (RSVD) needs to be carried out on the adjacency matrix (A) pertaining to the AAN graph. To determine the optimal number of singular values to use in RSVD, we analyze the Scree plot of singular values of the square matrix (AA^T) as seen in Figure 2. The reduced form of the scree plot in Figure 2 which shows the elbow being formed at 24 dimensions.

Figure 2: Reduced Scree Plot of Singular Values



Apart from the Scree plots, I also look at the proportion of the variation explained by the corresponding eigen values, in trying to determine the “cutting point”. An important insight I note here is that this is a very iterative and a subjective process and any single methodology is not generalizable across datasets. The number of Cutting Points used by me to reduce 11952

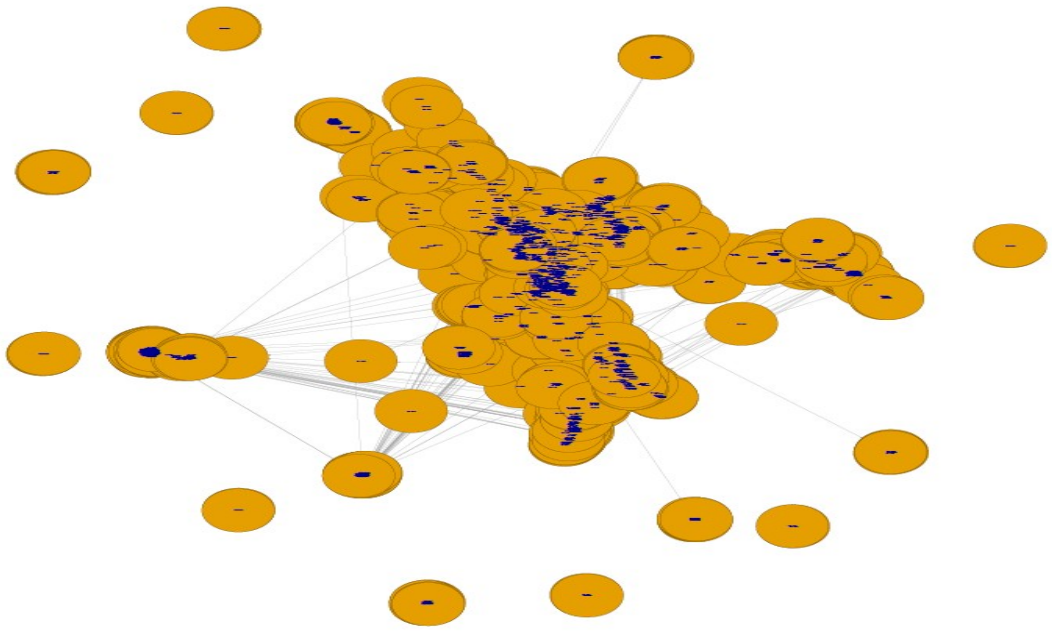
dimensions were as given in Table 3. The cutting points were determined by looking at the elbow formation, the point of decay and different levels of proportion of variation explained.

Table 3: Determining the Cutting Points

# DIMENSIONS	PROPORTION OF VARIATON EXPLAINED
24 (Corresponding to the elbow)	32.49%
47	40.24% (>40)
97	50.10% (>50)
147 (Corresponding to the point of decay)	56.04%
192	60.03% (>60)
377	70.01% (>70)
772	80.00% (>80)

I have performed RSVD with each of the cutting points referred to in Table 3 and have implemented the Semantic PR algorithm for all such combinations. I find that while 24 dimensions are too few, thereby introducing much noise while being lossy, 772 would be too many dimensions considering 80% of the variation is explained and the original dataset was mimicked almost entirel y. The results presented here are only for models with 47 dimensions. Term stemming or stop word removal were not used in my LSI models, to enable comparison with the pure text based PR model. The graph after RSVD and with 47 dimensions is plotted in Figure 3.

Figure 3: The reduced graph with 47 dimensions



The top 8 results from my Semantic PR algorithm (with 47 dimensions) for the query “Summarization” are given in Table 4.

Table 4: The top 8 results from the Semantic PR algorithm for the query “Summarization”

PAPER TITLE	SEMANTIC RELEVANCE	PR SCORE	SEMANTIC PAGERANK
Text Summarization Challenge 2 - Text Summarization Evaluation At NTCIR Workshop 3	7.44	0.000156	1
Hierarchical Summarization: Scaling Up Multi-Document Summarization	7.30	0.0000128	2
Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality	6.19	0.0000274	3
Extractive email thread summarization: Can we do better than He Said She Said?	5.94	0.0000128	4
Using the Amazon Mechanical Turk to Transcribe and Annotate Meeting Speech for Extractive Summarization	5.87	0.0000128	5
Empirical analysis of exploiting review helpfulness for extractive summarization of online reviews	5.80	0.0000128	6
Extractive Summarization and Dialogue Act Modeling on Email Threads: An Integrated Probabilistic Approach	5.77	0.0000128	7
Subtree Extractive Summarization via Submodular Maximization	5.65	0.0000145	8

It can be seen from Table 2 and Table 4 that the two algorithms give different results but all of the results have the query word occurring at least once. It is interesting to note that 367 documents contain the search term in their titles and all of these occur among results in the base version of the PR algorithm and in the Semantic PR algorithm with just 47 dimensions. This is a testament to the underlying premise of RSVD in LSI – that we lose noise due to dimensionality reduction and not information. In addition to the 367 documents returned by both algorithms, my algorithm also returns documents which do not contain the search query but are close to it in meaning. Although, the ones containing the query term are ranked at the top.

There may be instances where the search term may not occur in the corpus at all. One way of presenting the results in this case would be a “Did you mean this?” format, where the search term does not occur in the results but a synonym or a closely associated term does. I test the performance of the models on this count by removing the 545 documents containing the root word “summar”. While the PR algorithm returns zero results in this case, the top 8 results from the Semantic PR algorithm with 47 dimensions are given in Table 5.

Words such as “compression”, “abstractive” and “entailment” are very closely related in meaning to “summarization” and are returned by the Semantic PR model, as seen in Table 5. This is of extreme importance in delighting a user and in broadening the user’s knowledge.
Thus far, I have assumed the person searching is an expert in the area of NLP and is thus using the keyword “summarization”. Since I am more interested in helping the novice user, there is a need to build a search engine robust to noisy queries. I now test the models on the query “Summarize”.

Table 5: Top 8 documents with words synonymous to Summarization.

PAPER TITLE	SEMANTIC RELEVANCE	PR SCORE	SEMANTIC PAGERANK
An extractive supervised two-stage method for sentence **compression**	0.94	0.0000205	1
Using the Omega Index for Evaluating **Abstractive** Community Detection	0.63	0.0000133	2
Focused **Entailment** Graphs for Open IE Propositions	0.62	0.0000128	3
A New Sentence **Compression** Dataset and Its Use in an Abstractive Generate-and-Rank Sentence Compressor	0.55	0.0000134	4
Global Learning of Focused **Entailment** Graphs	0.55	0.0000249	5
A System Of Verbal **Semantic Attributes** Focused On The Syntactic Correspondence Between Japanese And English	0.52	0.0000235	6
You Are What You Say: Using Meeting Participants' Speech To **Detect** Their Roles And Expertise	0.49	0.0000251	7
Contradiction-focused qualitative evaluation of textual **entailment**	0.48	0.0000136	8

Only 11 results are returned for the query “summarize” by the PR algorithm from 367 titles containing the word “summarization”. I have not presented the results here. The top 8 results from the Semantic PR algorithm (with 47) dimensions for the query “Summarize” are given in Table 6.

We can see from Table 6 that the Semantic PR model not only retrieves documents with the word “Summarization”, it also retrieves documents that do not contain the word, but contain text closest to the meaning of “Summarize”. Documents ranked 2, 6 and 8 are cases in point. There is no mention of a synonym to the query term in these documents, but the meaning latent in the complete sentence is matched with the query word. I summarize my findings and bring up the conclusion in the next section.

Table 6: The top 8 results from my Semantic PR algorithm for the query “Summarize”

PAPER TITLE	SEMANTIC RELEVANCE	PR SCORE	SEMANTIC PAGERANK
Summarize What You Are Interested In: An Optimization Framework for Interactive Personalized Summarization	0.35	0.0000213	1
Tell Me What You Do And I'll Tell You What You Are: Learning Occupation-Related Activities For Biographies	0.34	0.000024	2
Can You Summarize This? Identifying Correlates of Input Difficulty for Multi-Document Summarization	0.29	0.0000298	3
CBSEAS a Summarization System – Integration of Opinion Mining Techniques to Summarize Blogs	0.26	0.0000128	4
What Are The Points? What Are The Stances? Decanting For Question-Driven Retrieval And Executive Summarization	0.23	0.0000143	5
You Are What You Say: Using Meeting Participants' Speech To Detect Their Roles And Expertise	0.23	0.0000251	6
Extractive email thread summarization: Can we do better than He Said She Said?	0.20	0.0000128	7
Learning Abstract Concept Embeddings from Multi-Modal Data: Since You Probably Can't See What I Mean	0.20	0.0000128	8

CONCLUSION

The PR algorithm exploits the link structure of the web and uses a rule based model to match a query with previously ranked web pages. It suffers from an inability to cope with two classic problems arising in natural language processing: synonymy and polysemy. Search Engine Persuasion is another problem it has to contend with. I have presented a novel approach for searching documents based on the meaning inherent in their content and call it Semantic PR. By using Semantic PR, problems of Synonymy and Polysemy using LSI are solved which are not solved using PR alone. Multiple relevant pages result from LSI which are ranked based on the notion of importance using PR. The Search engine persuasion problem is taken care of in my approach since PR alone takes care of unimportant pages based on the links and the semantic aspect ensures that misleading pages with hyperlinks alone are not given high relevance scores. In testing the performance of both the base PR algorithm and the Semantic PR algorithm comprehensively on real world data, it is seen that all the documents containing a query word occur among results in the base version of the PR algorithm and in the Semantic PR algorithm

built using a fraction of the total number of dimensions. This is a testament to the underlying premise of RSVD in LSI – that we lose noise due to dimensionality reduction.
In instances where the search term may not occur in the corpus at all, the PR algorithm returns zero results, but the Semantic PR algorithm with 47 dimensions returns documents with words very closely related in meaning to the query. This increase in the user's knowledge is due to the semantic network. Since I am more interested in helping the novice user, i test the models' performance on noisy or part queries. I find that only a few results are returned for such a query by the PR algorithm from a large number of possible results. The Semantic PR model not only retrieves all possible results, but it also retrieves documents containing text closest to the meaning of the query.
Future work would include searching for multiple words in a query string. It would indeed be very interesting to see this algorithm generalized to the world wide web.

REFERENCES

Baker, Kirk. "Singular value decomposition tutorial." *The Ohio State University* 24 (2005).

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 34-43.

Berry, Michael W., Susan T. Dumais, and Gavin W. O'Brien. "Using linear algebra for intelligent information retrieval." *SIAM review* 37.4 (1995): 573-595.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3. Jan (2003): 993-1022.

Brin, Sergey, and Lawrence Page. "Reprint of: The anatomy of a large-scale hypertextual web search engine." *Computer networks* 56.18 (2012): 3825-3833.

Conesa, Jordi, Veda C. Storey, and Vijayan Sugumaran. "Improving web-query processing through semantic knowledge." *Data & Knowledge Engineering* 66.1 (2008): 18-34.

Fang, Wei-Dong, et al. "Toward a semantic search engine based on ontologies." *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*. Vol. 3. IEEE, 2005.

Galitsky, Boris. "Matching parse thickets for open domain question answering." *Data & Knowledge Engineering* 107 (2017): 24-50.

Hofmann, Thomas. "Probabilistic latent semantic indexing." *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM*, 1999.

Hofmann, Thomas. "Unsupervised learning by probabilistic latent semantic analysis." *Machine learning* 42.1-2 (2001): 177-196.

Jardine, James, and Simone Teufel. "Topical PageRank: A model of scientific expertise for bibliographic search." *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. 2014.

Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment."In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms.* 1998.

Kulkarni, Shailesh S., Uday M. Apte, and Nicholas E. Evangelopoulos. "The use of latent semantic analysis in operations management research." *Decision Sciences* 45.5 (2014): 971-994.

Lempel, Ronny, and Shlomo Moran. "The stochastic approach for link-structure analysis (SALSA) and the TKC effect1." *Computer Networks* 33.1-6 (2000): 387-401.

Leyba, Todd. "Semantic search by means of word sense disambiguation using a lexicon." U.S. Patent No. 9,317,589. 19 Apr. 2016.

Liu, Liwei, Freddy Lecue, and Nikolay Mehandjiev. "Semantic content-based recommendation of software services using context." *ACM Transactions on the Web (TWEB)* 7.3 (2013): 17.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. Introduction to information retrieval. Vol. 1. No. 1. Cambridge: *Cambridge university press*, 2008.

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint* arXiv:1301.3781 (2013).

Nie, Lan, Brian D. Davison, and Xiaoguang Qi. "Topical link analysis for web search." *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2006.

Page, Lawrence, et al. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, 1999.

Radev, Dragomir R., Pradeep Muthukrishnan, and Vahed Qazvinian. "The ACL anthology network corpus." Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries. *Association for Computational Linguistics*, 2009.

Schymik, Gregory, et al. "The Benefits and Costs of Using Metadata to Improve Enterprise Document Search." *Decision Sciences* 46.6 (2015): 1049-1075.

Sul, Hong Kee, Alan R. Dennis, and Lingyao Ivy Yuan. "Trading on twitter: Using social media sentiment to predict stock returns." *Decision Sciences* 48.3 (2017): 454-488.

DECISION SCIENCES INSTITUTE

A Comparative Study of Self-congruency, Privacy and Trust Models for Self-disclosure on Social Networking Sites

Anjee Gorkhali
Susquehanna University
Email: gorkhali@susqu.edu

Asim Shrestha
Arkansas Tech University
Email: ashrestha2@atu.edu

ABSTRACT

This study analyzes the responses from 380 individuals to analyze the factors that impact disclosure of information on a social networking sites. It explores the similarities/differences regarding the impact that different self-concepts, privacy and trust have across different domains of self disclosure. Social networking sites survive and thrive based on the information that users disclose.

KEYWORDS: Social Networking, Self Disclosure, Privacy Calculus, Self Congruency, Trust in Social Networking

INTRODUCTION

The earliest known definition of self-concept was provided by James, 1890 as “all that we call our own, and with who or with which we share a bond of identity”. It was coined in the field of psychology however the major research in self-concept has been conducted under the term “symbolic interactionism” in the field of sociology. The basis of this theory is that the self-arises in social interaction with others through symbolic communication. There is a symbolic value associated with people or product that interacts with the self-concept of the individual. Depending on whether the symbol enhances, distorts, or has no effect on the individual’s self-concept, the individual is motivated to approach, avoid, or remain apathetic to the product (Denzin, 2016).

Self-theorists have defined self-concept as an attitude one holds about or towards one’s person (self) (Ross, 1971). This attitude consists of the following components: Cognitive: knowledge, belief; Affective: evaluations; and Behavioral-motivational: predispositions or tendencies to respond. There have been different approaches to the operationalization of self-concept in the literature. This abundance of definitions and categorizations has sometimes been cited as a problem (Claiborne & Sirgy, 1990). We have adopted four different types of self-concept listed below for this research. These are also the most widely used categories in the literature (M Joseph Sirgy, 1982).

- Actual self-concept: This refers to the actual self-image that a person has about oneself.
- Ideal self-concept: This refers to the self-image that an individual would like to be. This is desired self of the individual.
- Social self-concept: This refers to the self-image that an individual believes others have of them. It has also been referred to as looking glass self or presenting self.
- Ideal social self-concept: This refers to the self-image that the individual desires others to have about them.

Self-congruency is the level of match or mismatch between the self-concept of an individual and the image of the product, brand, or service that the individual consumes or has the intention of consuming. A product-user image interacts with the consumer's self-concept and generates a subjective experience referred to as self-image/product image congruency or simply self-congruency (M. Joseph Sirgy et al., 1997). Product and services have personality images just as people do (M. Joseph Sirgy, 1985), (Aaker, 1999). A variety of factors may contribute to the construction of these images including the physical characteristics of the products, advertising, price, and stereotypes of a generalized user of that product or service (M. Joseph Sirgy, 1982), (M. Joseph Sirgy, Grewal, & Mangleburg, 2000). Self-congruency theory proposes that consumer behavior is partially determined by the congruency resulting from the psychological comparison involving the product-user image and the consumer's self-concept. The cause for this effect is explained by the consumer's self-concept motives of self-esteem and self-consistency, i.e. a desire to enhance or maintain one's self-concept (M Joseph Sirgy, 1986).

Obar and Wildman (2015) have used the following commonalities to define a social networking site:

- Social networking sites are (currently) Web 2.0 internet-based applications.
- User generated content is the lifeblood of social networking sites.
- Individuals and groups create user-specific profiles for a site or app designed and maintained by a social networking site.
- Social networking sites facilitate the development of social networks online by connecting a profile with those of other individuals and/or groups.

Users share their personal information through their profiles, status updates, group and private chats to other members of the social networking sites. Hence, the information content consumed in the social networking sites is the product of self-disclosure by individuals and groups using the service. Although research on self-disclosure is not a new phenomenon, research that focuses on self-disclosure in the context of social networking sites is a relatively recent development (Varnali & Toker, 2015). With origins in verbal communication research, self-disclosure has been defined as the process of making the self-known to others (Jourard & Lasakow, 1958). It is an act of revealing personal information including thoughts, feelings, and experiences to others (Derlega et al., 1993). Another definition describes self-disclosure as any information about himself that Person A communicates to Person B (Cozby, 1973). Different factors that may potentially vary self-disclosure include the duration, accuracy, intimacy, intent of disclosure, positive or negative information, and relevance to other topics under discussion (Wheeless & Grotz, 1976). Previous research shows that the different factors affecting self-disclosure in social networking sites are self-congruency Shrestha, A. (2017), internet trust and personal interest (Dinev & Hart, 2006), perceived publicness of social networking site (Pike, Bateman, & Butler, 2009), trust and perceived control (Krasnova et al., 2010), relationships development, social validation, and self-expression (Yang and Tan, 2012), perceived benefits and social influence (Cheung et al. 2015). In addition to that, self congruency have a positive direct impact on user's perception of trust and privacy towards social networking platform.

According to Westin (1967), "privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means" (p. 7). Krasnova et al. (2010) analyze the privacy calculus theory from the context of SNS and establish that users who reported having higher perceived privacy risks had a less comprehensive Facebook profile and users who reported getting more benefits had a more comprehensive profile. Hence, asserting that user's privacy concern will impact their self-disclosure behavior in SNS. Similarly, Joinson et al. (2008)

explain that user's trust in the social networking platform plays an important role in user's self disclosure.

The theory of reasoned action proposes behavioral intention as the most important determinant of an individual's behavior (Fishbein & Ajzen, 1977), (Ajzen & Fishbein, 1980). The theory of planned behavior (Ajzen, 1991) was developed as an extension to the theory of reasoned action by adding an additional construct of perceived behavioral control that predicts behavioral intention and the actual behavior. Thus, attitude towards the behavior, subjective norm, and the perceived behavioral control are the three antecedents to the behavioral intention that leads to the actual behavior.

Drawing from the self-congruency theory, the theory of planned behavior and the privacy calculus theory, we have developed a theoretical model that explains the effect of self-congruency on trust, privacy, attitude, intention and behavior towards self-disclosure on social networking sites. We tested this model for four different types of self-congruencies emanating from the four different types of self-concepts mentioned above. We attempted to seek answers to following research questions:

- To what extent does different type of self-congruencies affect the attitude, intention, and behavior regarding self-disclosure on social networking sites?
- To what extent does trust on a social networking site and user's perception of privacy affect the attitude, intention and behavior regarding self-disclosure on social networking sites?
- To what extent the different self-congruency models are similar or different?

The remainder of this paper is organized as follows: the next section presents related work in self-congruency and self-disclosure on social networking sites. In the third section, we present and discuss the research model and our research hypotheses. In the sections that follow, we discuss the method, results, contributions, future work and limitations of our study.

RELATED WORKS

Self-congruency

Some of the earlier works that shifted the focus in marketing research from a purely economic and utilitarian perspective were (Gardner & Levy, 1955), (Newman, 1957), and (Levy, 1959). They suggested that an effort needs to be applied to understand the consumer needs and buying decisions by using behavioral science rather than just the economic rationality and sales statistics. Ever since then, the concept of self-congruency has been applied as a predictor for things ranging from brand relationship quality (Nyffenegger, Krohmer, Hoyer, & Malaer, 2015), attitude and intention to visit a tourism destination (Pratt & Sparks, 2014), to attitude and intention to adopt e-books (Anton et al., 2013). Although, self-congruency based on actual self-concept dominates the literature, there is still a significant number of work on different types of self-concepts. (Hosany & Martin, 2012) showed ideal self-congruency as a predictor of satisfaction from a cruise trip, (Nam, Ekinci, & Whyatt, 2011) found positive effect of ideal self-congruency on emotional brand attachement. (Hyun Ju & Mira, 2013) used the ideal social self-concept and found its positive effect on the intention to join a cause. Along with other self-concepts, (Bosnjak, Sirgy, Hellriegel, & Maurer, 2011) found the positive effect of ideal social-self congruency on post visit loyalty for a tourist destination.

In the context of social networking sites, researchers have explored the role of self-congruency on perceived usefulness, perceived enjoyment and continuance intention of the social networking

sites (Kang et al., 2009), (Kang et al., 2013). Shrestha, A. (2017), showed the positive effect of self-congruency on intention and behavior regarding self-disclosure on social networking sites.

Privacy on social networking sites

According to Burgoon (1982), it is possible to distinguish *physical privacy* (freedom from surveillance and unwanted intrusions upon one's physical space), *interactional privacy* (control over social encounters), *psychological privacy* (protects from intrusions upon one's thoughts, feelings, attitudes, and values), and *informational privacy* (the ability to control the aggregation and dissemination of information). As privacy in SNSs is largely about the dissemination and retention of personal information, we hence focus on aspects of informational privacy. To the best of our knowledge that there has not been any attempt towards examining the role of self congruency on and that of privacy on self-disclosure on social networking sites except for Shrestha, A. (2017). Our aim in this study is to explore and compare the effect of self-congruency on user's privacy perception and its impact on their self-disclosure on social networking sites based on the four major self-concepts.

Trust on social networking sites

Trust is critical in understanding when we choose to share personal information with others and when we choose secrecy. Altman (1977) describes a self boundary (the boundary around the person) that is modified by self-disclosure; and a dyadic boundary that ensures the discloser's safety from leakage of information to uninvited third parties. The self boundary may be open or closed depending on such interpersonal factors as the level of trust in a disclosure target (Altman, 1973). Within e-commerce, trust has been identified as a key factor in determining purchasing behavior: ''if the web site does not lead the consumer to believe that the merchant is trustworthy, no purchase decision will result'' (Ang & Lee, 2000, p. 3). Metzger (2006) found that trust (conceptualized as reputation) predicted disclosure behavior to a mock music CD e-commerce site. To the best of our knowledge that there has not been any attempt towards examining the role of self congruency on trust and that of trust on self-disclosure on social networking sites except for Shrestha, A. (2017). Our aim in this study is to explore and compare the effect of self-congruency on user's trust and its impact on their self-disclosure on social networking sites based on the four major self-concepts.

Self-disclosure on social networking sites

Before the mainstream popularity of social networking sites, (Dinev & Hart, 2006) found that the influence of internet trust and personal interests outweigh the privacy risk perceptions in individual's decision to disclose personal information in the context of electronic transactions. Some of the motivators for self-disclosure on social networking sites as indicated by past research are convenience of maintaining and developing relationships, platform enjoyment, trust and perceived control of social networking site, positive social influence, reciprocity, tendency towards collectivism, user commitment, user satisfaction, social validation, and self-expression among others (Krasnova et al., 2010), (Posey, Lowry, Roberts, & Ellis, 2010), (Yang and Tan, 2012), (Xu, Visinescu, & Kim, 2013).

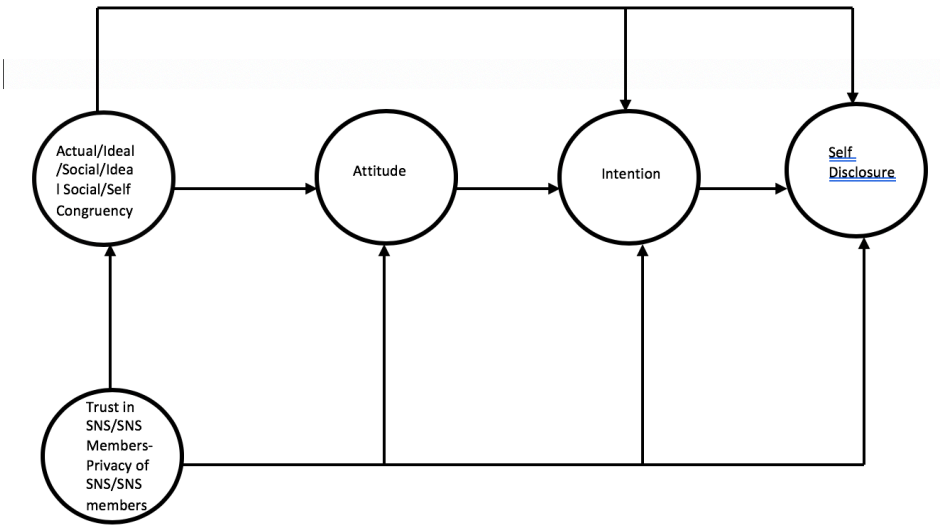
To the best of our knowledge that there has not been any attempt towards examining the role of self-congruency on self-disclosure on social networking sites except for Shrestha, A. (2017). Our aim in this study is to explore and compare the effect of self-congruency on self-disclosure on social networking sites based on the four major self-concepts.

RESEARCH MODEL

Synthesizing the self-congruency theory and the theory of planned behavior, we developed a theoretical model to evaluate the effect of self-congruency on self-disclosure on social networking sites for four major self-types.

Self-congruency with a social networking site is the match between the users' self-concept and the image of the social networking site. It has been shown that self-congruency with a product, service, or activity has an effect on the attitude, intention, as well as behavior associated with the usage of that product, service, or activity (M. J. Sirgy, 2015), (Pratt & Sparks, 2014), (Schoenmueller et al., 2013), (Anton et al., 2013), (D. Kim et al., 2015), (Ryu & Lee, 2013), (Ying & Hailin, 2015). Furthermore, it has been shown that the attitude and intention of disclosure on social networking site is impacted by the privacy concerns which results from the concerns about possible loss of privacy as a result of self disclosure (Xu et al. 2008, p.4). In addition to privacy another factor impacting the intention and attitude to disclosure is trusting beliefs (McKnight et al. 2002, p.303). Accordingly, we argue that self-congruency with a social networking site will have a direct positive effect on the attitude towards self-disclosure, self-disclosure intention, and self-disclosure behavior on the social networking site. In addition to that, trust in a social networking site as well as its members will have a direct positive effect on the attitude towards self disclosure, self disclosure intention and self-disclosure behavior on the social networking site. Finally, privacy concern in a social networking site as well as with its members will have a direct negative effect on the attitude towards self-disclosure, self-disclosure intention and self disclosure behavior on the social networking site.

Figure 1: Research Model



H1 (a-d): Self-congruency (Actual/Ideal/Social/Ideal Social) with a social networking site has a positive effect on attitude towards self-disclosure on social networking site.

H2 (a-d): Self-congruency (Actual/Ideal/Social/Ideal Social) with a social networking site has a positive effect on self-disclosure intention towards self-disclosure on the social networking site.

H3 (a-d): Self-congruency (Actual/Ideal/Social/Ideal Social) with a social networking site has a positive effect on self-disclosure on the social networking site.

Attitude represents user's favorable or unfavorable feelings of disclosing information on the social networking site. Theory of planned behavior suggests that attitude towards a behavior directly influences the behavioral intention and behavioral intention is directly linked to the actual behavior (Ajzen, 1991). This leads us to the following hypotheses:

H4 (a-d): Attitude towards self-disclosure on a social networking site has a positive effect on self-disclosure intention on the social networking site for all types of self-congruency (Actual/Ideal/Social/Ideal Social) models.

H5 (a-d): Self-disclosure intention towards a social networking site has a positive effect on self-disclosure on the social networking site for all types of self-congruency (Actual/Ideal/Social/Ideal Social) models.

H6 (a-b): Trust with a social networking site/members of a self networking site has a positive effect on attitude towards self-disclosure on social networking site.

H7 (a-b): Trust with members on a social networking site/members of a self networking site has a positive effect on self disclosure intention towards self disclosure on a social networking site.

H8(a-b) Privacy concern with a social networking site/members of a self networking site has a negative effect on attitude towards self-disclosure on social networking site.

H9 (a-b): Privacy concern with members on a social networking site/members of a self networking site has a negative effect on self disclosure intention towards self disclosure on a social networking site.

Research Method

We administered a survey among undergraduate students at a large university in the US South. There were 380 total participants: 199 females and 181 males. Average age of the participants was 20.73 years. We choose Facebook as the prototype social networking site for this research. The average number of 'friends' of the participants on Facebook was 853.77.

We adopted all the scales used in measuring the constructs in our research model from previous studies. All the constructs were assessed using a 5-point Likert scale. Sample items used are shown in Table 1.

Table 1: List of Constructs and Sample Items

Construct	Construct Definition	Sample Items	Source
Self-congruency (SC)	The congruence resulting from a psychological comparison involving the product-user image and the consumer's self-concept.	SC1. The image of the typical user of Facebook is consistent with how I am. (Actual) SC5. The image of the typical user of Facebook is consistent with how I would like to see myself. (Ideal) SC8. The image of	(M. Joseph Sirgy et al., 1997) (M. Joseph Sirgy & Su, 2000)

		the typical user of Facebook is consistent with how others see me. (Social) SC10. The image of the typical user of Facebook is consistent with how I would like others to see me. (Ideal Social)	
Attitude towards self-disclosure (ATT)	User's favorable or unfavorable feelings of disclosing information on the social networking site.	ATT1. I think disclosing my information on Facebook is good for me.	(Hsu, Yen, Chiu, & Chang, 2006)
Self-disclosure intention (SDI)	The behavioral intention to disclose personal information on the social networking site.	SDI2. I do not hesitate supplying my personal information to my Facebook friends.	(Beldad, van der Geest, de Jong, & Steehouder, 2012)
Self-disclosure (SD)	The extent to which information about the self is disclosed on the social networking site.	SD3. I often post about myself on Facebook.	(Krasnova et al., 2010) (Sawyer et al., 2011)
Privacy Risk from SNS (PRSNS)	The expected losses related to self-disclosure due to SNS.	PRSNS1. In general, it would be risky to give information to Facebook.	(Jarvenpaa, Tractinsky, & Saarinen, 1999) (Naresh K Malhotra, Kim, & Agarwal, 2004)
Privacy Risk from SNS members (PRSNSM)	The expected losses related to self-disclosure due to SNS members.	PRSNSM1. In general, it would be risky to give information to my Facebook friends.	(Jarvenpaa et al., 1999) (Naresh K Malhotra et al., 2004)
Trust in SNS (TRSNS)	User's trust on the social networking site.	TRSNS1. Facebook is open and receptive to the needs of its members.	(McKnight, Choudhury ,& Kacmar, 2002) (Jarvenpaa et al., 1999)
Trust in SNS members (TRSNSM)	User's trust on his/her connections on the	TRSNSM1. My Facebook friends will	(Chiu, Hsu, & Wang, 2006) (McKnight et al., 2002)

	social networking site.	do their best to help me.	

We checked the validity of the constructs, assessed the models, and performed path analysis. The results are analyzed and discussed in the next section.

Analysis and Results

The measurement model's consistency reliability was established using Cronbach's alpha, composite reliability and average variance extracted as reported in table 2. The first step in measuring model reliability is to measure the internal consistency of the model. It is the indicator of how well the items on the test measure the same construct. For this purpose, we used Cronbach's alpha value that provides an estimate of the reliability based on the inter-correlations of the observed indicator variables. Since, Cronbach's alpha tends to underestimate the internal consistency reliability, we also used composite reliability to measure the internal consistency of our model. The values of Cronbach's alpha and composite reliability between 0.7 and 0.9 are considered satisfactory (Nunnally & Bernstein, 1994). In the next step we assessed convergent validity. It is the extent to which a measure correlates positively with alternative measure of the same construct. To have convergent validity, the indicators or the items that measure a construct should converge or share a high proportion of variance. The average variance extracted (AVE) is used as a measure of convergent validity. This criterion is defined as the grand mean value of the squared loadings of the indicators associated with the construct (Hair Jr, Hult, Ringle, & Sarstedt, 2016). AVE values of 0.5 or higher indicate that on average, the construct explains more than half of the variance of its indicators and are considered satisfactory (Hair Jr et al., 2016). All the metrics were satisfactory for all the constructs.

Table 2: Reliability and Validity Measures for all Constructs

Construct	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
Actual Self-congruency	0.932	0.91	0.822
Ideal Self-congruency	0.942	0.919	0.845
Social Self-congruency	0.942	0.924	0.844
Ideal Social Self-congruency	0.956	0.934	0.878
Privacy Risk from SNS	.886	.885	.613
Privacy Risk from SNS members	.940	.931	.758

Trust in SNS	.889	.906	.582
Trust in SNS members	.896	.914	.636
Attitude	0.919	0.90	0.79
Intention	0.856	0.809	0.671
Self-disclosure	0.849	0.784	0.583

After the evaluation and refinement of the measurement model, we estimated the structural model. All the four models had a good fit as reflected by normed chi square value of less than 3 and the RMSEA value of less than 0.8. The results are summarized in Table 3.

Table 3: Model Fit for all four Models.

Model	CMIN/DF	RMSEA	AIC
Actual Self-congruency	2.932	.071	193.699
Ideal Self-congruency	2.936	.071	190.592
Social Self-congruency	2.953	.072	186.389
Ideal Social Self-congruency	2.941	.072	202.503

AIC value is used as metric to compare different models Kline (2005). The model with lower AIC value is considered better. The AIC values listed in table 3, indicate that social self-congruency model with AIC value of 186.389 outperformed the other models.

The path coefficients for hypothesized relationships and R^2 values for dependent variables were estimated by applying bootstrapping with no sign changes and 5000 subsamples. We did bootstrapping as Partial Least Squares – Structural Equation Modeling does not assume normality of data distribution meaning that the parametric significance test used in regression analyses cannot be directly applied (Hair Jr et al., 2016).

The R^2 and adjusted R^2 values for the dependent variables viz, Attitude towards self-disclosure (ATT), Self-disclosure intention (SDI), Self-disclosure (SD), Privacy risk from SNS (PRSNS), Privacy risk from SNS members (PRSNSM), Trust in SNS (TRSNS) and Trust in SNS MEMBERS (TRSNSM(across the four models are shown in Table 4. The adjusted R^2 values suggest that the four models explain 13.1, 13.2, 13.8, and 12.7 percent of the variance for the overall model.

Table 4: R^2 and Adjusted R^2 Values of Dependent Variables for all four Models.

Model	R^2	Adjusted R^2
Actual Social Congruency	SD: .148	SD: .131
Ideal Self Congruency	SD: .149	SD: .132
Social Self Congruency	SD: .154	SD: .138
Ideal Social Self Congruency	SD: .144	SD: .127

Table 5 lists the path coefficients and their significance for all the hypothesized relationships across the four models. All the hypotheses are supported meaning that self-congruency based on all four types of self-concepts has a positive effect on attitude towards self-disclosure, intention of self-disclosure, and actual self-disclosure behavior on social networking sites. Also, there is a significant direct positive effect from attitude to intention, and from intention to self-disclosure across all the models.

Table 5: Path Coefficients, Significance and Support for Hypotheses for all four Models

Model	Path	Significance	Path	Significance	Path	Significance
Actual	SC→ATT	.105(**) H1(a)(-)	SC→INT	.063(**) H2(a)(-)	SC→SD	.233(***) H3(a)(+)
	SC→PRSNS	.105(**) H8(a)(-)	SC→PRSNSM	.063(**) H9(a)(-)	SC→TRSNS	.233(***) H6(a)(+)
	SC→TRSNSM	.312(***) H7(a)(+)				
	ATT→INT	.251(***) H4 (a) (+)	INT→SD	.107(***) H5(a)(+)		
Ideal	SC→ATT	.142(*) H1(b)(+)	SC→INT	.265(***) H2(b)(+)	SC→SD	.146(*) H3(b)(+)

	SC→PRSNS	.104(not sig) H8(b)(+)	SC→PRSNSM	.067(not sig) H9(b)(+)	SC→TRSNS	.242(***) H6(b)(+)
	SC→TRSNSM	.295(***) H7(b)(+)				
	ATT→INT	.326(***) H4 (b) (+)	INT→SD	.107(not sig) H5(b)(+)		
Social	SC→ATT	.143(**) H1(c)(+)	SC→INT	.219(***) H2(c)(+)	SC→SD	.171(***) H3(c)(+)
	SC→PRSNS	.163(**) H8(c)(+)	SC→PRSNSM	.050(**) H9(c)(+)	SC→TRSNS	.233(**) H6(c)(+)
	SC→TRSNSM	.293(**) H7(c)(+)				
	ATT→INT	.250(***) H4 (c) (+)	INT→SD	.110(***) H5(c)(+)		
Ideal Social	SC→ATT	.137 (*) H1(d)(+)	SC→INT	.212 (***) H2(d)(+)	SC→SD	.133(*) H3(d)(+)
	SC→PRSNS	.167 (***) H8(d)(+)	SC→PRSNSM	.086 (not sig) H9(d)(+)	SC→TRSNS	.213 (***) H6(d)(+)
	SC→TRSNSM	.254 (***) H7(d)(+)				
	ATT→INT	.253 (***) H4 (d) (+)	INT→SD	.119(*) H5(d)(+)		

Contributions

This study validates the concept of self-congruency, trust and privacy as a predictor for self-disclosure on social networking sites. We have shown that self-congruency emanating from four major self-concepts are equally valid for this purpose. This is a novel contribution to the literature. To the best of our knowledge, the effect of different types of self-congruencies has never been

studied in the context of social networking sites. In addition that, including trust and privacy variables to the model gives additional insight into the extent to which users tend to disclose when busing social media. The extent depends greatly on trust and privacy of the users on both the platform as well as the other users using the platform.

We were able to compare the four models based on different types of self-concepts and conclude that they are equally good in terms of model fit and the significance of path estimates of all the hypothesized relationships. We believe this has led to a better understanding of different self-concepts – in terms of their role in explaining self-disclosure on social networking sites, they all perform good.

Limitation and Future Studies

Since we collected data using a survey for our study, it consequently inherits the limitations of this method. A longitudinal study method could have captured the changes in the parameters of the research model. The temporal distinction between independent and dependent variables is absent. This means that it is hard to argue with full confidence that the causal relationship assessed in the study have the same direction as postulated in the research model. We used a single social networking site: Facebook, for this research. While it is the most popular social networking site ("Social Networking Use," 2015), it cannot be argued that self-disclosure phenomenon on Facebook would be same as on other platforms. The differences that exist among social networking sites would raise questions on any attempts towards generalization of the results of this study. A convenient sample of undergraduate students also contributes towards the lack of generalizability.

People use different social networking sites for different purpose. An interesting research for the future would be to assess how different our model would perform for platforms other than Facebook. This will shed light on the effect of different self-concepts according to the characteristics of the platform itself. Longitudinal and qualitative methods could yield results that could challenge or strengthen the validity of the causal relationships assessed in this study. Future studies involving different groups such as older adults, non-students, professional etc. will do the same. We have used self-reported data for this study, participants could have underreported their self-disclosure behavior on the survey. This could have been the reason for slightly lower validity measures for the construct self-disclosure. Using actual behavioral data collected from the social networking site for measuring self-disclosure might yield more accurate results.

Conclusion

We were able to formulate and test four different models for self-disclosure on social networking sites. We have been able to show that different self-concepts: actual, ideal, social, and ideal social behave almost the same way when it comes having an impact on sharing of information about self on social networking sites. We take it as a starting point to explore the similarities/differences regarding the impact that different self-concepts have across different domains of our interest. Social networking sites survive and thrive based on the information that users disclose. It is the willingness of users to disclose their information that drives the economies of these sites (A. Joinson et al., 2011). An in-depth understanding of self-disclosure process is of immense importance for the maintenance, promotion, and growth of social networking sites. By testing and reporting the impact of different types of self-concept on self-disclosure, we believe that we have added new insights into that understanding.

REFERENCES

Aaker, J. L. (1999). The Malleable Self: The Role of Self-Expression in Persuasion. Journal of Marketing Research (JMR), 36(1), 45-57-57.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.

Altman, I. (1975). The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.

Ang, L., & Lee, B. C. (2000). Influencing perceptions of trustworthiness in Internet commerce: A rational choice framework. In *Fifth CollECTer Conference on Electronic Commerce* (pp. 1-12). CollECTeR.

Anton, C., Camarero, C., & Rodriguez, J. (2013). Usefulness, Enjoyment, and Self-Image Congruence: The Adoption of e-Book Readers. Psychology & Marketing, 30(4), 372384-384. doi: 10.1002/mar.20612

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Bosnjak, M., Sirgy, M. J., Hellriegel, S., & Maurer, O. (2011). Postvisit Destination Loyalty Judgments: Developing and Testing a Comprehensive Congruity Model. Journal of Travel Research, 50(5), 496-508-508. doi: 10.1177/0047287510379159

Burgoon, J. K. (1982). Privacy and communication. *Communication Yearbook,* 6, 206–249.

Cheung, C., Lee, Z. W., & Chan, T. K. (2015). Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence. Internet Research, 25(2), 279299.

Claiborne, C. B., & Sirgy, M. J. (1990). Self-Image Congruence as a Model of Consumer Attitude Formation and Behavior: A Conceptual Review and Guide for Future Research. In B. J. Dunlap (Ed.), Proceedings of the 1990 Academy of Marketing Science (AMS) Annual Conference (pp. 1-7): Springer International Publishing.

Cozby, P. C. (1973). Self-disclosure: a literature review. Psychological bulletin, 79(2), 73.

Denzin, N. K. (2016). Symbolic interactionism. *The international encyclopedia of communication theory and philosophy*, 1-12.

Derlega, V. J., Metts, S., Petronio, S., & Margulis, S. T. (1993). Self-disclosure: Sage Publications, Inc.

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1), 61-80.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research.

Gardner, B. B., & Levy, S. J. (1955). The product and the brand. Harvard Business Review, 33(2), 33-39.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A primer on partial least squares structural equation modeling (PLS-SEM): Sage Publications.

Hosany, S., & Martin, D. (2012). Self-image congruence in consumer behavior. Journal of Business Research, 65(5), 685-691-691. doi: 10.1016/j.jbusres.2011.03.015

Hyun Ju, J., & Mira, L. (2013). The Effect of Online Media Platforms on Joining Causes: The Impression Management Perspective. Journal of Broadcasting & Electronic Media, 57(4), 439-455-455. doi: 10.1080/08838151.2013.845824

James, W. (1890). The consciousness of self, principles of psychology.

Joinson, A., Houghton, D., Vasalou, A., & Marder, B. (2011). Digital Crowding: Privacy, SelfDisclosure, and Technology. In S. Trepte & L. Reinecke (Eds.), Privacy Online (pp. 3345): Springer Berlin Heidelberg.

Joinson, A. N. (2008). *Looking at, looking up or keeping up with people?: motives and use of facebook.* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Jourard, S. M., & Lasakow, P. (1958). Some factors in self-disclosure. The Journal of Abnormal and Social Psychology, 56(1), 91.

Kim, D., Jang, S., & Adler, H. (2015). What drives cafe customers to spread eWOM? International Journal of Contemporary Hospitality Management, 27(2), 261-282-282. doi: 10.1108/ijchm-06-2013-0269

Kline, R. B. (2005). Principles and practice of structural equation mod.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: why we disclose. J Inf technol, 25(2), 109-125.

Levy, S. J. (1959). Symbols for sale. Harvard Business Review, 37(4), 117-124.

Nam, J., Ekinci, Y., & Whyatt, G. (2011). Brand equity, brand loyalty and consumer satisfaction. Annals of Tourism Research, 38(3), 1009-1030. doi: http://dx.doi.org/10.1016/j.annals.2011.01.015

Metzer, D. (2006). Modern silence. *The Journal of Musicology*, *23*(3), 331-374.

Newman, J. W. (1957). New Insight, New Progress, for Marketing. Harvard Business Review, 35(6), 95-102.

Nyffenegger, B., Krohmer, H., Hoyer, W. D., & Malaer, L. (2015). Service Brand Relationship Quality: Hot or Cold? Journal of Service Research, 18(1), 90-106-106. doi: 10.1177/1094670514547580

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. Psychometric theory, 3(1), 248-292.

Pratt, M. A., & Sparks, B. (2014). Predicting Wine Tourism Intention: Destination Image and Self-congruity. Journal of Travel & Tourism Marketing, 31(4), 443-460-460. doi: 10.1080/10548408.2014.883953

Obar, J. A., & Wildman, S. S. (2015). Social Media Definition and the Governance Challenge An: Introduction to the Special Issue. Available at SSRN 2663153.

Pike, J. C., Bateman, P. J., & Butler, B. S. (2009). I didn't know you could see that: The effect of social networking environment characteristics on publicness and self-disclosure. AMCIS 2009 Proceedings, 421.

Posey, C., Lowry, P. B., Roberts, T. L., & Ellis, T. S. (2010). Proposing the online community self-disclosure model: the case of working professionals in France and the UK who use online communities. European Journal of Information Systems, 19(2), 181-195.

Pratt, M. A., & Sparks, B. (2014). Predicting Wine Tourism Intention: Destination Image and Self-congruity. Journal of Travel & Tourism Marketing, 31(4), 443-460-460. doi: 10.1080/10548408.2014.883953

Rogers, C. (1951). A theory of personality and behavior. C. Rogers, Client-Centered Therapy: Its Current Practice, Implications, and Theory, 481-533.

Ross, I. (1971). Self-Concept and Brand Preference. The Journal of Business, 44(1), 38-50. doi: 10.2307/2351834

Schoenmueller, V., Bruhn, M., Walther, E., & Schaefer, D. (2013). The Impact of Viral Message Self-Congruity on the Perception of the Message and the Intention to Share It. AMA Winter Educators' Conference Proceedings, 24.

Sirgy, M. J. (1982). Self-Concept in Consumer Behavior: A Critical Review. Journal of Consumer Research, 9(3), 287-300. doi: 10.2307/2488624

Sirgy, M. J. (1986). Self-congruity: Toward a theory of personality and cybernetics: Praeger Publishers/Greenwood Publishing Group.

Sirgy, M. J. (2015). The Self-Concept in Relation to Product Preference and Purchase Intention. In V. V. Bellur (Ed.), Marketing Horizons: A 1980's Perspective (pp. 350-354): Springer International Publishing.

Sirgy, M. J., Grewal, D., Mangleburg, T. F., Park, J.-O., Kye-Sung, C., Claiborne, C. B., . . . Berkman, H. (1997). Assessing the Predictive Validity of Two Methods of Measuring Self-Image Congruence. Journal of the Academy of Marketing Science, 25(3), 229.

Sirgy, M. J., Grewal, D., & Mangleburg, T. (2000). Retail Environment, Self-Congruity, and Retail Patronage: An Integrative Model and a Research Agenda. Journal of Business Research, 49(2), 127-138-138.

Shrestha, A. (2017). An Integrated Framework for Self-disclosure on Social Networking Sites.

Social Networking Use. (2015). http://www.pewresearch.org/data-trend/media-andtechnology/social-networking-use/

Varnali, K., & Toker, A. (2015). Self-disclosure on social networking sites. Social Behavior & Personality:an international journal, 43(1), 1-13. doi: 10.2224/sbp.2015.43.1.1

Westin A (1967) Privacy and freedom. Atheneum, New York

Wheeless, L. R., & Grotz, J. (1976). Conceptualization and measurement of reported selfdisclosure. Human Communication Research, 2(4), 338-346.

Xu, C., Visinescu, L., & Kim, D. (2013). Disclose Intimately, Honesty, Heavily, Positively and Intentionally: An Exploration of Self-Disclosure in Social Networking Sites.

Yang, L., & Tan, B. C. (2012). Self-disclosure on online social networks: Motives, context feature, and media capabilities.

Ying, Q. U., & Hailin, Q. U. (2015). Nonutilitarian Tourism Destination Positioning: A Case Study in China. International Journal of Tourism Research, 17(4), 388-398-398. doi: 10.1002/jtr.2005

DECISION SCIENCES INSTITUTE

Visual Analytics for Storytelling: Transitioning from Analysis Skillset to Business Mindset

Dr. Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

Visual analytics is known for its easy-to-use, interactive, and intuitive data analysis capabilities and more organizations use them in their decision situations. As such, we are in need of a teaching method to guide students to learning not only the technical skillsets of the visual analytics but developing a business mindset. In this paper, the author will present one such approach by adopting a storytelling technique. It covers: 1) the essential features distinctively unique in the visual analytics technology; and 2) how the series of technical analyses can turn into a business story where its entirety represents business insights.

KEYWORDS: Visual analytics, Storytelling, business insights, Scene-based inquiry-answering process

BACKGROUND

Visual analytics is known for its easy-to-use, interactive, and intuitive data analysis capabilities. That's why software applications such as Tableau and Power BI have been widely adopted in various industry sectors in recent years. These are powerful analysis tools that can be used for various types of business analysis and more organizations use them in their decision situations. As a result, teaching these tools in an educational setting has become more important as the demand for college graduates who can use the technology effectively grows. However, the current teaching approach in this area largely depends on reference-type materials where instructors and students go over from one feature to another; or students are taught based on multiple unrelated modules. Teaching various technical capabilities of the visual analytics technology is essential but teaching it in a business context is even more important. It's because all the technical analysis results should be converted to the business information the decision makers would appreciate. As such, we are in need of a teaching method with which instructors guide students to learning not only the technical skillsets related to the visual analytics but at the same time developing a business mindset by connecting their analysis outcomes with the business insights.

OBJECTIVES

In this paper, the author will present one such approach based on multiple years of teaching Tableau in the college classes and industry training sessions. To do it, the author will introduce an operational improvement context and cover: 1) the essential features distinctively unique in the visual analytics technology; and more importantly 2) how the series of technical analyses can turn into a business story where its entirety represents business insights.

CONTEXT & SCENARIO

To provide a business context for the data analysis, the author will use a case study based on a fictional scenario involving sales challenges. In the scenario, the visual analytics is used as a main tool to find root causes of an incapable system and to deliver the findings to non-technical business decision makers.

Scenario

A recent college graduate, John, gets a task of analyzing business data from his boss, Paul. The following is an interactive analytics operations John performed with Paul.

"Hi, John, we get the sales data of the past 4 years (Table 1). I have to present the sales challenges and potential solutions to the Board. So I need your help."

Lesson Focus

With this small business scenario and the specific data given at the very beginning, students start operating in a business context. However, instead of jumping directly into the analytics storytelling session, students are given an Excel-based exercise questions first, such as "What's the total sales in 2014?" This warm-up serves two purposes: 1) Students get to know the data better before they start analytics operation; and 2) They get the solution based on what they used to use using the spreadsheet. The latter is important because it is an opportunity for the instructor to reminds them of information granularity in the spreadsheet. In other words, in Excel all the information by default is presented at a cell level. In order for students to get the answer, they have to filter all the sales cells for 2014 and then add up these cell values to get the total sales for the year. This will soon contrast with how Tableau works in the upcoming analyses below.

Table 1. Sales Data Used in Case Study

Order ID	Order Date	Customer N...	City	State	Country	Region	Product...	Category	Sub-Cate...	Sales	Profit
CA-2014-AB10...	11/11/2014	Aaron Bergman	Oklahom...	Oklahoma	United St...	Central US	TEC-PH-5...	TC	PNSZ	221.98	62.15
IN-2014-JR16...	2/5/2014	Justin Ritter	Wollong...	New Sout...	Australia	Oceania	FUR-CH-5...	FR	CCHR	3,709.40	-288.77
IN-2014-CR12...	10/17/2014	Craig Reiter	Brisbane	Queensland	Australia	Oceania	TEC-PH-5...	TC	PNSZ	5,175.17	919.97
ES-2014-KM1...	1/28/2014	Katherine Mu...	Berlin	Berlin	Germany	Western Europe	TEC-PH-5...	TC	PNSZ	2,892.51	-96.54
SG-2014-RH9...	11/5/2014	Rick Hansen	Dakar	Dakar	Senegal	Western Africa	TEC-CO-6...	TC	CPRS	2,832.96	311.52
IN-2014-JM15...	6/28/2014	Jim Mitchum	Sydney	New Sout...	Australia	Oceania	TEC-PH-5...	TC	PNSZ	2,862.68	763.28
IN-2012-TS21...	11/6/2012	Toby Swindell	Porirua	Wellington	New Zeala...	Oceania	FUR-CH-5...	FR	CCHR	1,822.08	564.84
IN-2013-MB1...	4/14/2013	Mick Brown	Hamilton	Waikato	New Zeala...	Oceania	FUR-TA-3...	FR	TBLA	5,244.84	996.48
CA-2014-AB10...	11/11/2014	Aaron Bergman	Oklahom...	Oklahoma	United St...	Central US	FUR-BO-5...	FR	BKCK	341.96	54.71
CA-2012-AB10...	3/6/2012	Aaron Bergman	Seattle	Washington	United St...	Western US	FUR-CH-4...	FR	CCHR	48.71	5.48
CA-2012-AB10...	3/6/2012	Aaron Bergman	Seattle	Washington	United St...	Western US	OFF-AR-5...	OS	GTRA	17.94	4.66
ID-2013-AJ10...	4/19/2013	Anthony Jacobs	Kabul	Kabul	Afghanist...	Southern Asia	FUR-TA-3...	FR	TBLA	4,626.15	647.55
SA-2012-MM7...	12/26/2012	Magdelene M...	Jizan	Jizan	Saudi Ara...	Western Asia	TEC-PH-3...	TC	PNSZ	2,616.96	1,151.40
MX-2013-VF2...	11/13/2013	Vicky Freymann	Toledo	Parana	Brazil	South America	FUR-CH-4...	FR	CCHR	2,221.80	622.02
IN-2014-PF19...	6/6/2014	Peter Fuller	Mudanji...	Heilongjia...	China	Eastern Asia	OFF-AP-4...	OS	APNC	3,701.52	1,036.08
ES-2015-BP11...	7/31/2015	Ben Peterman	Paris	Ile-de-Fra...	France	Western Europe	OFF-AP-3...	OS	APNC	1,869.59	186.95
CA-2012-AB10...	2/19/2012	Aaron Bergman	Arlington	Texas	United St...	Central US	OFF-ST-30...	OS	STTR	12.62	-2.52
ES-2015-PJ18...	9/8/2015	Patrick Jones	Prato	Tuscany	Italy	Southern Euro...	OFF-AP-4...	OS	APNC	7,958.58	3,979.08
IN-2015-JS15...	1/31/2015	Jim Sink	Townsville	Queensland	Australia	Oceania	TEC-CO-3...	TC	CPRS	2,565.59	28.40

ANALYTICS STORYTELLING

When teaching business analytics in a storytelling format, it is important to focus on a part of the dataset (the author call it a scene) where an initial question is given for students to answer. Typically, in the visual analytics, their answer contains more information than what the question asked, which likely lead them to another question(s), which will be handled in another scene. And this questioning-answering repeats until business insights are derived. Ultimately, at the end of the analytics operation, students end up producing a series of answers based on the questions they have developed along the way. In the following section, this scene-based inquiry-answering process is described using the sales challenge case.

Scene 1

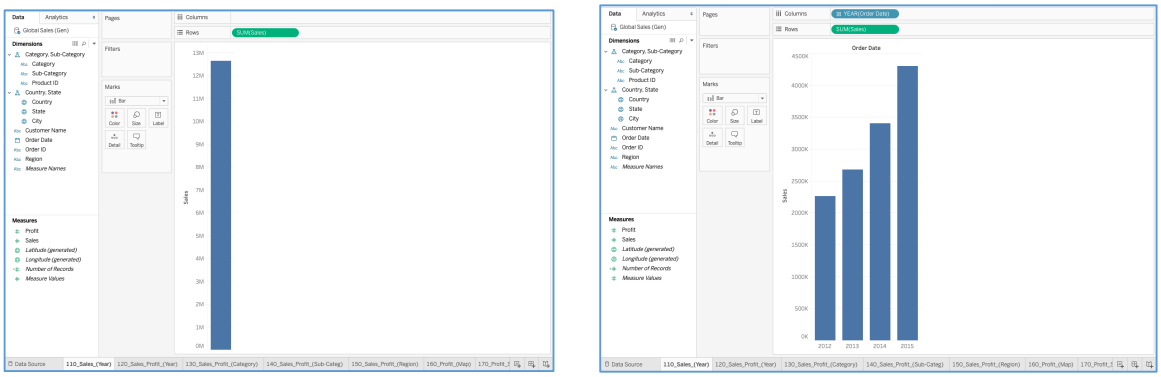
"John, I wonder how the sales has been over the years. Can you show me?"

Lesson Focus

In this scene, the focus of the instruction is that, unlike Excel, the default option for the information granularity in Tableau is an aggregate (such as sum, average, variance, and etc.). In other words, the individual values are not accessible directly. This is a major difference students have to get used to because most of time their analysis is done in an aggregate form and, depending on how much detailed information they want to have, they need to divide the aggregate value to the finer level. For example, as shown in Figure 1, Left, when students bring the "Sales" variable (numerical variable) out, they'll automatically get one aggregate vertical bar, which represents the total sales (i.e., sum of individual values in the Sales variable). This can be in turn divided to small pieces using a dimension variable. Figure 1, Right, shows when they bring the "Year" variable (dimensional variable) out, the total Sales is broken into the ones for the four different years.

In Figure 1, Right, students not only get the sales values for each year but they get the sales trend immediately since in Tableau the graphical presentation is the default option. In fact, in this particular presentation, the sales has increased rapidly over the years. This simple finding would be the starting point of the next scene.

Figure 1: Aggregate Sales (Left) and Aggregate Sales Broken into Years (Right)



Scene 2

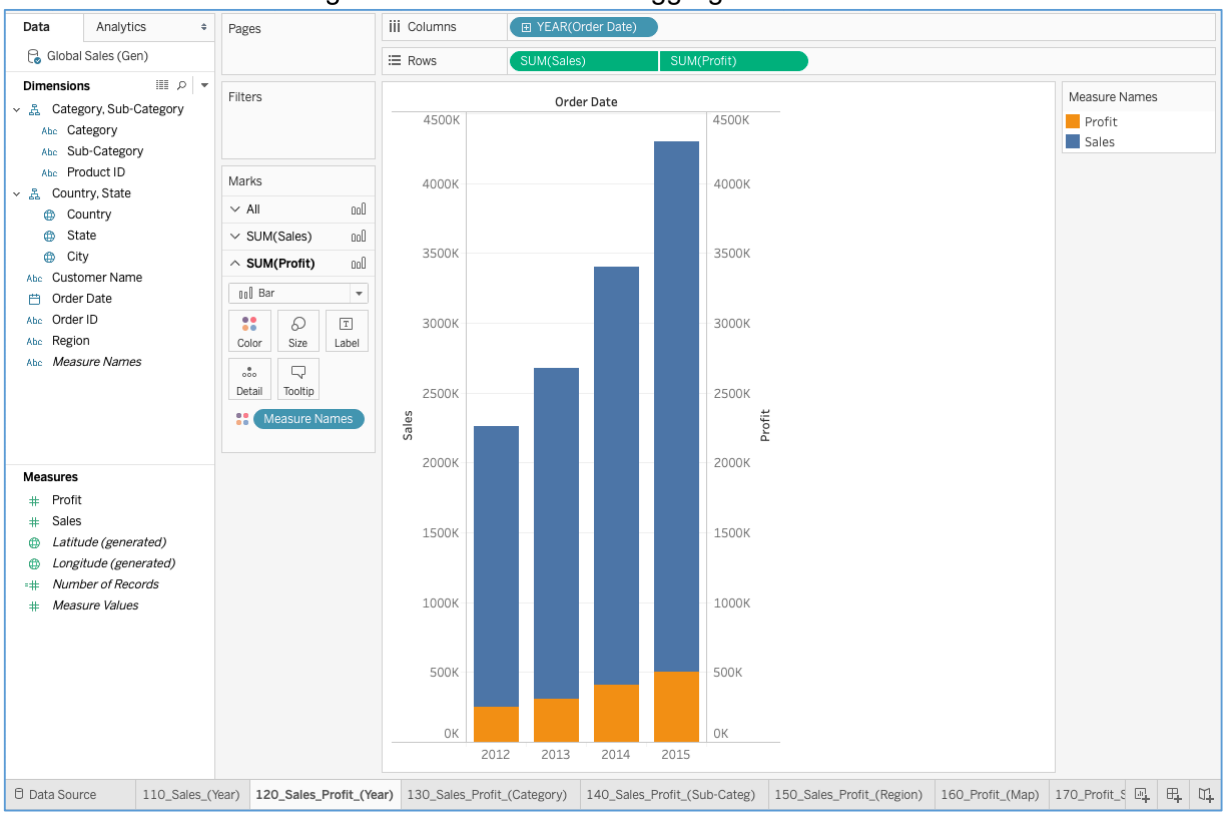
"John, sales getting better over the years, it seems… I wonder how the profit has been during the same period. Can you show me the profit together so that I can compare?"

Lesson Focus

Here, the instructor shows a way of showing another numerical variable, "Profit", to be placed on top of the existing presentation of "Sales", using the format already established so that they can be easily compared. The result is in Figure 2 where now the profit is shown in orange.

One possible reaction to this presentation is that, since the sales and profit information is in a graphical format, students may notice even if both of them performed better year after year their increase rates are not the same. In fact, the profit has performed at the same level as the sales has done. If you are a manager of the business, a natural reaction to this would be to think about "why" and this will generate another question in the next scene.

Figure 2: Sales and Profit Aggregates into Years



Scene 3

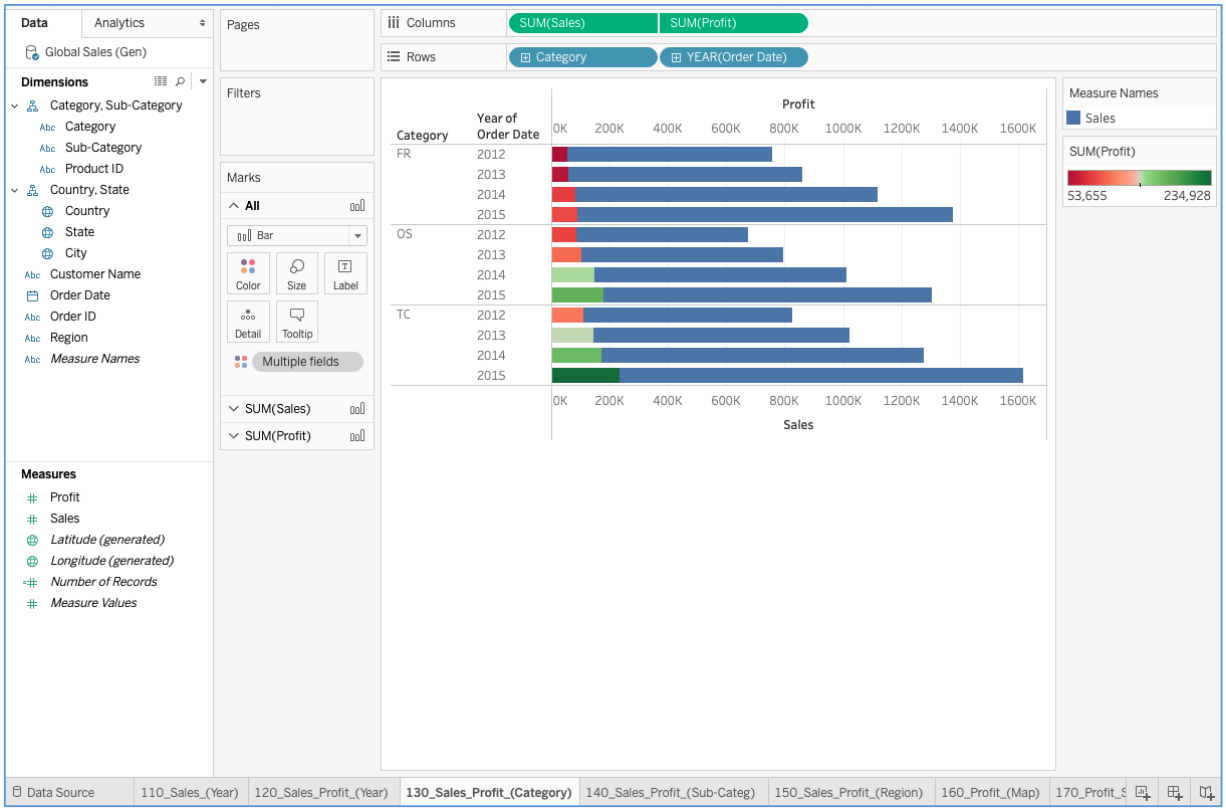
"'Hmm… Profit increase struggles. Can we break it down to…say, per "PRODUCT CATEGORY'?"

Lesson Focus

It is important at this point for the instructor to remind students the level of detail deepens from the overall sales and profit to the individual category level (using the "Category" dimension variable), which is aligned with the intention of the question that Paul wants to know what product categories are doing well and poor over the years. The result is shown in Figure 3. To do this easily, the color scheme has changed for the profit where different shades of the red and the green indicates different levels of poor and good performance of the profit, respectively.

The figure obviously grabs students' attention that the product category, 'FR', has performed bad consistently over the years. And it'll likely make them think to drill down further into the sub-category or product level, which is handled in the next scene.

Figure 3: Yearly Sales and Profit Performances Per Product Category



Scene 4

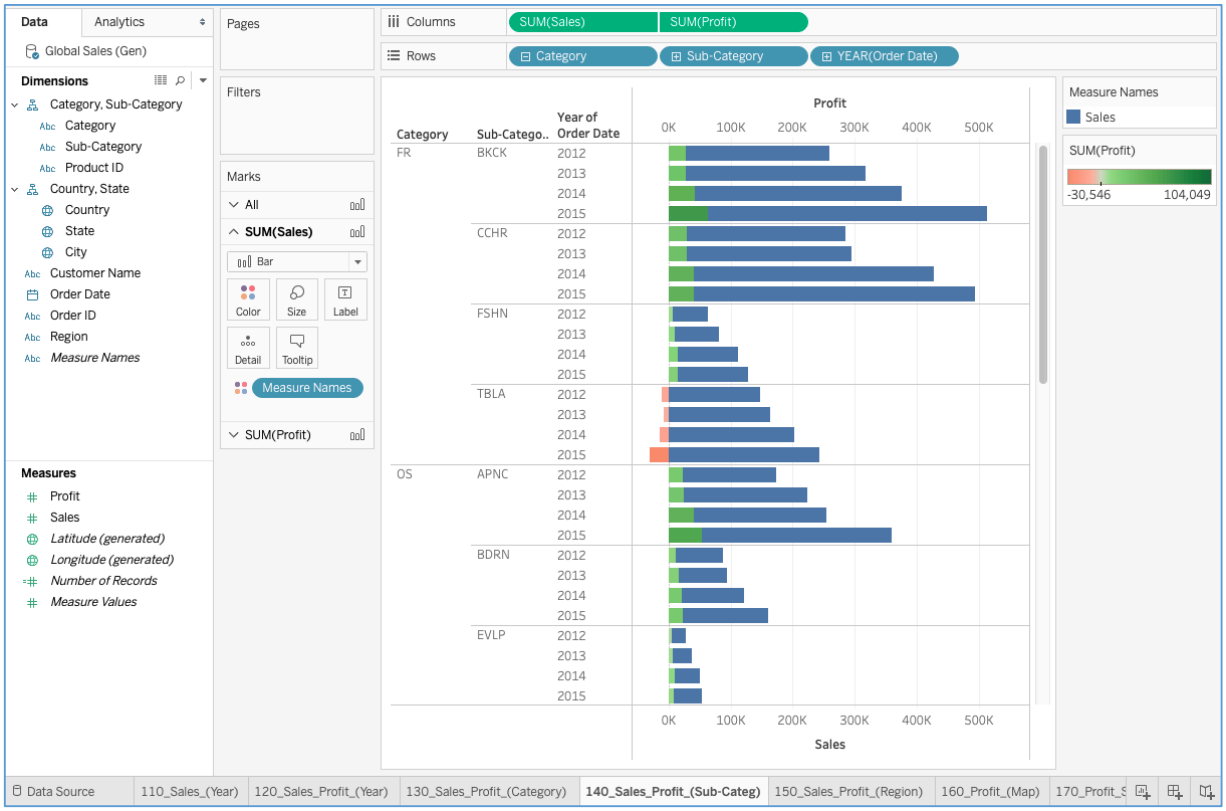
"Uh-Oh… John, category 'FR' has struggled. Can we look into more of its specific contributing products?"

Lesson Focus

As in the previous scene, the instructor needs to retain students' attention that they deepen their inquiry process to identify more specific areas, such as product sub-categories. Using another dimension variable, "Sub-Category", the information is presented in a more granularized fashion as shown in Figure 4.

From the figure, students will recognize that, even in the poorly performing product category 'FR', not all the sub-categories are performing poorly but the sub-cateogry 'TBLA'. They could stop their analysis at this point, but, they also need to think other ways of drilling down to more specific reason why the profit has been struggling. This is discussed in the next scene.

Figure 4: Yearly Sales and Profit Performances Per Product Sub-Category



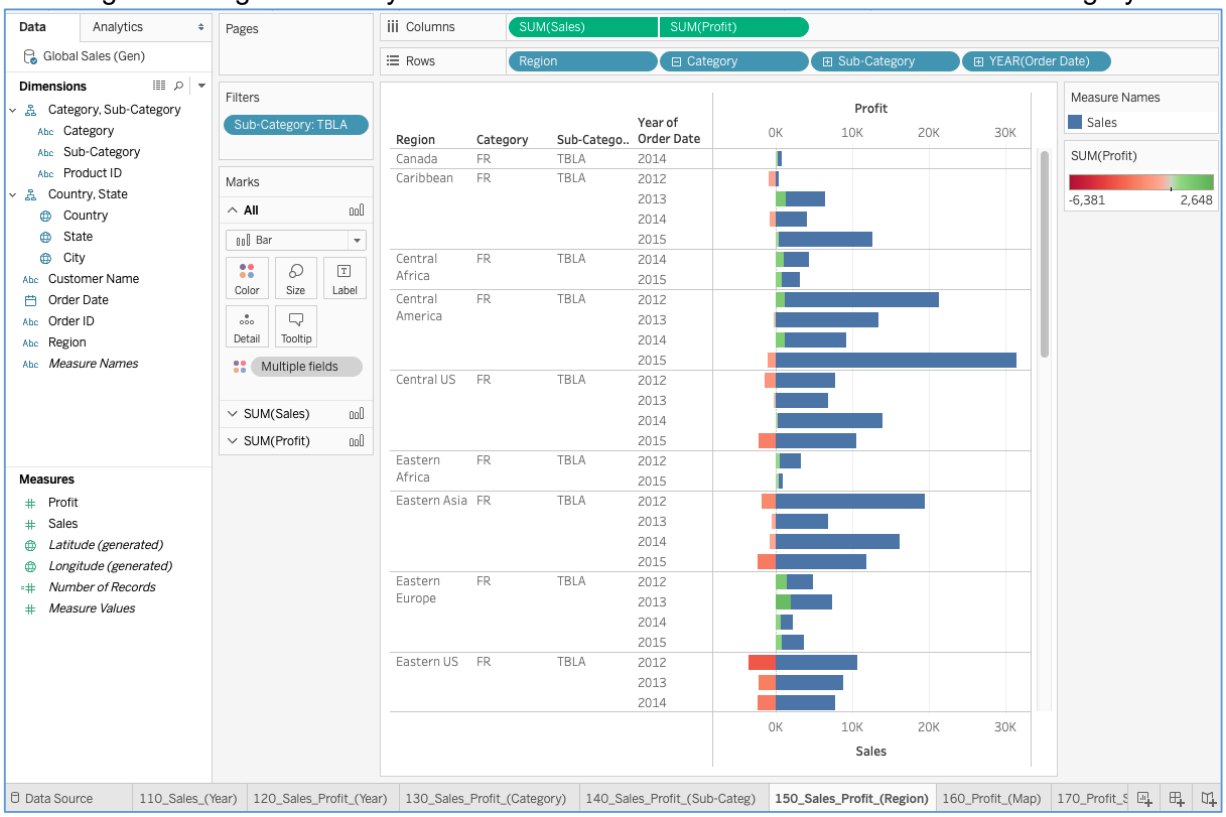
Scene 5

"Wow.. the product 'TBLA' is the culprit of the poor profit performance. Thank you, John. But we have markets internally so I assume regions are performing at various levels. Can we further look at it per REGION?"

Lesson Focus

With the new dimension variable, “Region”, the profit performance is now shown according to the regions as shown in Figure 4. There are multiple regions with mixed performance outcomes but the worst performing region is ‘Eastern US’, which is still too big for Paul to make any actionable recommendations to remedy the poor profit situation. Students at this point get used to think in a way to continue their inquiry process, which continues in the next scene.

Figure 5: Regional Yearly Sales and Profit Performances Per Product Sub-Category



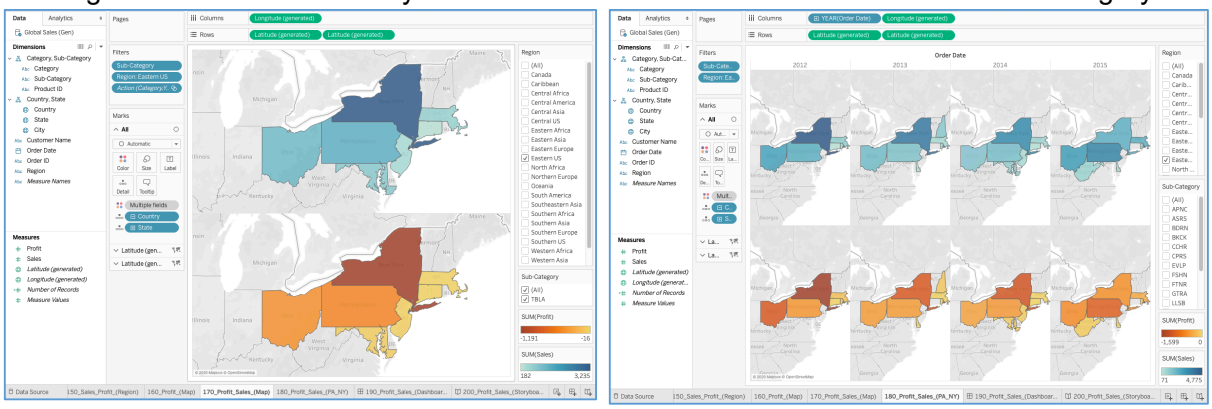
Scene 6

“John… ‘Eastern US’ got us the biggest damage. Can we zoom into it? I’d like to pinpoint a particular State to take an action.”

Lesson Focus

The instructor now shows how the bar chart-based sales and profit information presented in the previous scene can be turned into the map-based information, which is shown in Figure 6 below. It is a good technical exercise to demonstrate to students there are multiple ways of presenting the same information but with different intentions.

Figure 6: State-Level Yearly Sales and Profit Performances Per Product Sub-Category



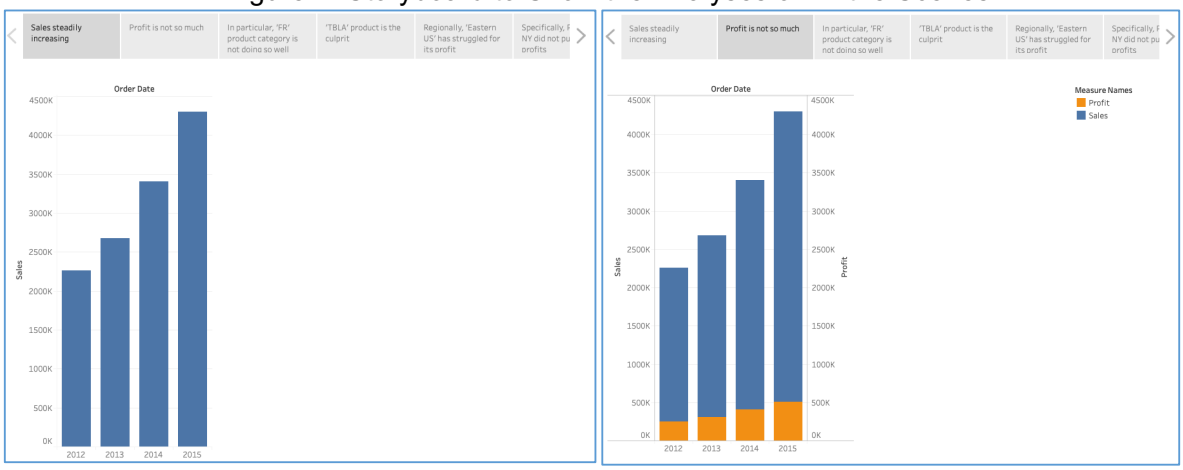
Final Scene: Storyboard

"Seems 'NY' is getting better but 'PA' has not been improving. I need to put together these findings so that I can share them with the Board."

Lesson Focus

The instructor teaches a storyboarding technique where students assemble all the individual stories from the scenes in a way to connect with the non-technical business users as shown in Figure 7. This is a critical part beyond the analysis operation and they understand they need to operate themselves not only from a technical perspective but also from a business perspective.

Figure 7: Storyboard to Show the Analyses of All the Scenes



SUMMARY

The technique described in this article is to demonstrate how both the technical and business requirements are met by developing step by step procedures of visual analytics in a relevant business context. The biggest lesson for the learners is that it is not the answer found in an analysis is important but rather how this answer leads them to the next relevant question that needs to be answered. And how the similar process can go on until the end of a series of related investigations, which they put in a story of related analyses. The proposed teaching method has been positively taken by various learners from traditional college and graduate students to industry practitioners.

Students learn the major difference between spreadsheet and visual analytics in that to tell the story you have to know how to ask questions, not just answering a question. In fact, it is this questioning capability that is only possible if they have the business understanding as well as analytics technical skillset.

DECISION SCIENCES INSTITUTE
On the extraction of Cyber risks from Structured Products

Michel Verlaine
Michel.verlaine@icn-artem.com

ABSTRACT

The aim of this paper is to develop an approach to extract information about cyber risks from structured financial products. We consider decision makers that are interested in extracting information about the uncertainty of Cyber risks. The value of information can be evaluated using recently developed entropy approaches in Finance. The underlying idea is that what we call Arrow-Debreu Cyber Risk state prices can be extracted provided the right structured products be "created". It is shown that different market based approaches can be used to get a better idea of the shape of the loss distribution facing firms. This information is potentially of interest to evaluate the risk premiums of insurance products. Comparisons between extracted market expectations can also be informative for risk evaluation, notably the distribution of unexpected losses and the eventual shortfall calculations. Finally, recent information-theoretic approaches enable us to make the link between pricing and the value of information to investors.

KEYWORDS:: Risk Management, Pricing, Cyber Risk, Information Theory.

JEL classification: G11, D82, N2.

Introduction

The market for Cyber insurance products is rapidly increasing and the European market seems lagging behind the US. The EU Cyber Insurance market generates approximately 3-4 bn$ but is expected to growth to 20bn $ by 2025. The market development, however, is currently impaired by a lack of commonality in the language to assess the risks and thus the uncertainty concerning the premiums. This common language is part of an information collection process and implies the creation of standards in information communication and questionnaires. Moreover, there does not seem to exist a common approach to evaluate cyber risks. Cyber risks, however, are becoming a major challenge due to the systemic risks as well as the security threats involved.

A recent literature in financial economics also uses information from derivative products to extract information about macro-economic catastrophic events. The impact of disasters on stock markets has been acknowledged for some time (Barro (2006), Longstaff and Piazzesi (2004), Rietz (1988) but the shape of the distribution as well as the correct pricing kernel have become a focus more recently. The extreme distributions can be modelled with a normal and a jump component, a combination which follows an extreme value distribution. With limited data however, there is

considerable uncertainty about the shape of the distribution. Backus *et al*. (2011) use equity options with different strike prices to extract the distribution of equity returns. The advantage of the approach is that those products inform how market participants evaluate the likelihood and impact of extreme events, whether they occurred in the sample or not. The link between distributions inferred from option prices and statistical distributions, however, influences the pricing kernel and risk neutral versus true distributions. Backus *et al*. (2011) draw on Alvarez and Jermann (2005) and use the entropy of a pricing kernel to compare pricing kernels. This is a useful property as entropy can be represented as the infinite sum of cumulants of the log of the pricing kernel, which is useful to compare pricing kernels across different environments and financial products. As the concept of entropy has also been suggested as an important measure of value to investors (Cabrales *et al*. (2013)), it is a fundamental building block for information extraction and pricing.

We use those theoretical foundations to build a framework and a methodology to extract information about cyber risk distributions from structured financial products. The underlying idea is that if markets are efficient, market prices correctly reflect market expectations. We assume that a Decision Maker (DM), a government or an insurance company for instance, is interested in getting information about future cyber risks. This information can then be used by a DM to evaluate insurance products or eventually sell the information to third parties. This, however, leads to the question of how to evaluate or price the information. The issue of the value of information has been addressed by Ganuza and Penalva (2010), Azrieli and Lehrer (2008) as well as Cabrales *et al*. (2013). Cabrales *et al*. (2013) suggest that the general measure of the value of information for most decision makers with different risk aversions can be evaluated by a reduction in entropy provided by information.

Entropy as a measure of uncertainty takes into account the state probabilities in different states and bears a close relationship with Arrow-Debreu state prices (Arrow (1964), Debreu (1959)). Arrow-Debreu state prices indicate the price today of 1 unit of account in different future states. They are also called risk-neutral probabilities whose dispersion can be analyzed using entropy measures. Such risk neutral probabilities can be extracted from standard derivative products for standard market based risks (Breeden and Litzenberger (1978). This approach cannot be directly replicated for cyber risk indexed financial products, but a similar approach can be developed with more complex and especially created structured financial products.

Cyber risks statistically bear a certain resemblance with credit risks and their distributions seem to be heavy-tailed and can be modelled with extreme value theory (Zhan *et al*. (2013,2015)). Backus *et al*. (2011) develop a structural model to evaluate extreme risks and disasters. Such a model is potentially adapted to model extreme cyber risks. The model consists of two components: a normally distributed component for standard risks and a jump component where the latter can be calibrated for different degrees of jump severity and intensity. This approach can then be adapted to structured financial products that could depend on cyber risks.

Cyber event swaps based on the model of credit default swaps could be used to extract the risk premium for a portfolio of underlying assets subject to cyber risks. Assuming rational expectations and efficiency, this premium is then the best possible estimate of the expected losses due to cyber risks. This approach, however, cannot take into account cyber accumulation risks (Cambridge Center for Risk Studies (2016)) as well as correlation risks between firms. More advanced financial products, called correlation products, have to be used. The most adapted structured product is the Collateralized debt obligation (CDO). As those products have different layers of insurance coverage for different levels of loss severity, the risk premiums for the tranches potentially indicate the expected losses for each tranche. By structuring the products in finer and finer tranches, a finer and finer estimate of the loss distribution can be extracted. Each loss within a certain tranche can

be viewed as a loss in a certain risk state and the Arrow-Debreu state price approach can be applied together with the entropy approach to evaluate potential information in this distribution.
We show how such an approach could be implemented using the CDO pricing approach to extract the expected losses per tranche given observed prices. If there exists a market for such structured products, observed market prices correctly reflect expected losses, provided the market is more or less efficient. The extracted expected loss distribution, however, is based on a risk neutral probability distribution and can be compared to the historical probability distribution estimated from past data. This information can be used to evaluate the pricing kernel and the value of the information through a reduction in entropy as in *Cabrales* et al. (2013).
Section 1 discusses recent methods to evaluate the value of information for investors. I then elaborate on the role of the efficient markets hypothesis in section 2. In section 3, I elaborate on the concept of Arrow-Debreu State Prices. Recent cyber risk modelling approaches are then analyzed in section 4. The following section 5 focuses on the pricing of extreme risks in standard financial products and section 6 develops the CDO pricing tools. Finally, in section 7, I explain the implementation of the approach.

1. The Value of Information for Decision Makers

From an economic viewpoint, the value of information in general depends on how new information reduces uncertainty concerning events. The value of information thus depends on who uses it and what his prior information as well as his risk aversion is. It is thus important to distinguish what economists call private versus public information. Private information is held by individual entities whereas public information is known by the market. Those notions are also used in the efficient markets literature which is used to analyze how information is revealed in financial markets. There is academic evidence that the market aggregates information better and faster than any expert (Lo(2017)). An information broker could thus set up a financial market for structured cyber products along with other information sources. Provided the market is efficient it will provide the best possible expectations of risks and prices. The Cyber risk market faces the same problems than the standard structured credits market namely the definition of events, uncertainty of losses when the event happens and the correlation between events.
If we focus more specifically on the value of information for Decision Makers, a slight problem arises. A typical information broker that would like to evaluate the value of the information for Decision Makers who are willing to buy information needs to know the risk aversion and the private information this Decision Maker has. Economics has made a lot of progress in modelling decision makers' behavioral features. Standard economics typically focuses on the relative or absolute risk aversion which can be easily estimated using elicitation techniques. A more recent literature in behavioral economics, however, points out that behavior is more complex and more complex Prospect Theory types of models would have to be elicited (Wakker (2010)).
Recent theoretical developments also enable us to define the value of information with more precision and independence of risk aversion. The fundamental and general question is when a piece of information is more valuable to a Decision Maker than any other piece of information. The answer to this question is not straightforward and depends on three considerations:

- The prior beliefs (private information) of the Decision Maker about different possible events as information provides signals about possible future events.
- The risk aversion and the wealth of the Decision Maker.
- The specific decision problem to which information is applied and which depends on the firm.

Blackwell (1953) was the first to address the problem of ranking information structures. In the Blackwell sense, an information structure consists in a set of "signals" that indicate the likelihood of events. Blackwell's Theorem states that the most valuable information structure is the one that provides the highest Expected Utility for any decision problem. More formally, an information structure α is given by a finite set of signals S_α with a set of probabilities $\alpha_k \in \triangle(S_\alpha)$ for every *k*. α_k is the probability that the signal observed by the DM is *s* when the state of nature is *k*. Such an information structure is typically represented by a stochastic matrix with as many rows as states and as many columns as signals. Row *k* then indicates the probability distribution $\left(\alpha_k(s)\right)_{s\in S_\alpha}$. Signal *s* thus has a probability of $p_\alpha(s) = \sum_k p(k)\alpha_k(s)$. The posterior probability on *K* given *s* can be derived with the Bayes theorem:

$$q_\alpha^s(k) = \frac{p(k)\alpha_k(s)}{p_\alpha(s)} \quad (1)$$

The Blackwell ordering is a partial order where the most informative structure is the one that fully reveals the state of nature *k* and the least informative the one that has no informational content. This approach, however, cannot completely rank information structures. More recently, the issue of the value of information has been addressed by Ganuza and Penalva (2010), Azrieli and Lehrer (2008) as well as Cabrales *et al*. (2013).

Cabrales *et al*. (2013) suggest that the general measure of the value of information for most decision makers with different risk aversions is provided by a reduction in entropy provided by information. Besides the technical details, the idea is simple to understand. Imagine a Decision Maker facing uncertainty concerning future events. Entropy is a measure of uncertainty in a distribution. It is 0 when there is no uncertainty and otherwise positive. More formally, Cabrales *et* al. (2013) consider the following Shannon (1984) entropy:

$$H(q) = -\sum_{k\in K} q(k)ln(q(k)) \quad (2)$$

Where $q \in \triangle(K)$ is a probability distribution over a set of states $k \in K$. If we consider a DM with beliefs *q*, the entropy of *q* indicates the uncertainty about the states of nature held by the DM. The entropy measure thus measures the degree of uncertainty in a distribution and is at its maximum when the distribution is uniform and every state has the same probability. When there is full uncertainty the value of information is the highest. The value of entropy is the lowest when one of the states is sure to occur. When a state is sure to occur, information about the occurrence of the state has no value.

The entropy can then be used to evaluate the value of information structures. Cabrales *et al*. (2013) show that the value of an information structure α can be evaluated through the expected reduction of entropy following the observation of a signal *s* in the information structure. Assume the agent's signal is *s* with probability $p_\alpha(s)$ and the posterior probability on *K* following signal *s* is q_α^s. The so-called *entropy informativeness* is evaluated by the expected reduction of entropy for a DM with prior information *p* and taking into account signal *s*.

$$I(\alpha) = H(p) - \sum_s p_\alpha(s)H(q_\alpha^s) \quad (3)$$

Note that $I(\alpha)$ depends on the information of the DM prior to the signal *p* and the information structure α. Given a prior *p*, the index $I(\alpha)$ defines a complete ordering of information structures. Information makes sense only when the future events that occur are not known with certainty. The Decision Maker thus buys information only when there is uncertainty and when the information reduces this uncertainty. The reduction in uncertainty measured through entropy is the correct way, from an economic viewpoint, to evaluate the value of information provided to a Decision Maker. Note that this approach is general and can be applied to any decision and information problem (f.i. Cyber risks and information policies) provided it is calibrated to the reference problem. As we will

discuss more in detail later, in our case the historical information held by financial institutions can be viewed as the prior *p* and the risk neutral versus historical distribution extracted from structured products can be used as a signal.

In light of the discussion about Efficient Markets, interesting approaches might be developed if we make the link between what are called Arrow-Debreu State prices (Debreu (1959, Arrow (1964)) that can be used to extract probabilities and subjective evaluations of risks expected by financial markets. Those probabilities can be used to price assets and bear a close relationship with entropy. Relatively recent applications of those techniques to structured financial products and so-called Economic catastrophe bonds (Coval *et al*. (2009)) provide potentially interesting development possibilities to develop cyber risk indexed structured markets and use information from some of those products to evaluate risk and information. Such a market does not yet seem to exist. As already alluded to earlier, cyber risks seem to bear features that are close to those in the credit risk market.

Both risks are characterized by events that occur or do not occur and uncertainty about the losses when the event occurs. The first is measured through the probability of the event (Probability of event, PE) and the second the loss given the event (Loss given Event, LGE). Both elements combined provide the expected losses (PE*LGE) used to calculate the risk premium. Different modelling methods exist, notably structural or reduced form (Schönbucher (2003). More interestingly, here, is the fact that the same kinds of structured products could be used to create Cyber structured products. For instance, there exist Credit Default Swaps that enable two parties to agree to swap credit risks. A similar approach could be used to create some kind of "Cybersecurity Incident Swap", CIS. If such a product is traded it enables to extract the market risk premium for such a product, the latter being the best possible expectation, called rational expectation in economics. In order to extract more precise information about correlation of risks more advanced structured products such as Cyber indexed CDO could be created. Such approaches exist for standard credit structured products (Colin-Dufresne *et al*. (2012)) but could be transposed with some modifications in the modelling of losses and defaults.

2. Cyber risks and the Efficient Market Hypothesis

The Efficient Markets Hypothesis states that information about asset prices is rapidly integrated into asset prices. Formally, any price at any time is given by the best possible expectation given the available information set. A recent working paper by Lin *et al*. (2018) uses an approach based on the EMH to extract information about the potential loss distribution following cyberattacks. This information is important for insurance companies to evaluate and price the cyber insurance products. As suggested by Gordon and Loeb (2002), firms should invest in information security up to where the marginal benefit is just larger than the marginal cost of the cyber incident, even though catastrophic breaches should be accounted for. Gordon *et al*. (2015) go further in analysing the systemic impact in a connected network of firms. The authors point out that the investment in information security is too low from a socially optimal perspective as firms do not take into account negative externalities on other firms. It is often argued that cyberinsurance is priced too high given limited data to come up a with a consistent pricing approach. A way to estimate the premium is to add up the accumulated costs following a cybersecurity breach. The problem is that those costs are difficult to evaluate and are many faceted and some are intangible.

A direct approach would be to model firm losses directly and to fit a distribution to the available data as for instance in Zhan *et al*. (2015). Apart from the issue of limited data, this approach has

the drawback that it can only measure direct costs but not indirect and intangible costs. Moreover, from the insurance industry viewpoint the correlation between different firms is of foremost importance. The question thus is how to evaluate the insurance premiums taking into account those correlations and the indirect and intangible costs. Lin *et al*. (2018) use an approach based on the Efficient Markets Hypothesis to extract information about total costs, direct, indirect and intangibles. If the market is informationally efficient, the stock market is going to reflect the long run impact of the cyber event immediately. The informational efficiency, however, might depend on the Market Architecture as alluded earlier (Madhaven (2000)). As documented in the finance literature (Lo (2017)), the stock market is often more efficient in aggregating information than a set of experts. The more developed the Market Architecture, the more efficient the aggregation of privately known information. This means that a so-called event study approach can be used to evaluate the loss impact of cyber security breaches.

The idea of the event study approach is that stock market reactions around events, in our case a cyber security breach, can be used to evaluate the total loss impact on the firm. The total cumulative abnormal return (CAR) around the breach announcement for a large set of firms is the most adequate source to estimate the total loss distribution to determine insurance premiums. In order to estimate the total loss distribution, the authors analyse cyber breaches between 2011 and 2016. The focus is on publically traded firms where observations with other types of announcements around the breach dates have been removed from the sample. Not surprisingly, the sample indicates that finance and insurance related firms are the most frequent targets of cyber attacks, making up around a fourth of the sample. In order to evaluate the abnormal stock returns around the breach announcements, the CAPM together with the value-weighted CRSP index are used as a benchmark to determine normal performance. The focus is on CAR five days surrounding the announcement. The stock prices fall on average by 1.44% over the five day window surrounding the breach announcement. The CAR, which in this case is negative, can then be multiplied by the respective market capitalization of the firm to get an idea of the losses due to cyber breaches. The average dollar loss is 587 million. This average, however, is heavily influenced by the presence of very big firms whereas the median loss is “only” 77 million. The sample of losses around cyber breaches can now be used to extract the loss distribution. The gamma distribution seems to fit the empirical frequency of losses well.

In order to evaluate the cyber risk premium, however, network effects and so-called Third party as well as Accumulation risks have to be taken into account. Given the types of distributions of losses, the dependency cannot be modelled with standard correlation measures that are adapted to measure dependency between normally distributed random variables. Dependency measures that are adapted in this case are so-called Copulas which can be used to model the dependency between any type of distribution. The copula is a function that combines univariate density functions into their joint distribution. It enables us to disentangle the effect of two different risk factors from their co-movements. It measures the dependence whatever the marginal distributions of the risk factors. Copulas can be either extracted from multivariate distributions or a selected copula can be combined with the marginals to generate a multivariate distribution. The main interest rests with the fact that marginal distributions and the dependency can be modelled separately. This stems from Sklar’s Theorem which states that any group of random variables can be joined into their multivariate distribution using a copula. Moreover, if marginal distributions are continuous, the copula is unique. A good overview is provided in Nelson (2006). Note that a convex combination of copulas is also a copula. Different types of copulas have been suggested in the literature. Zhin *et al*. (2018) refer to the normal and t-copulas, whereas the latter accounts for tail dependence. This is consistent with Gaussian copula approach implicit in the Vasicek CDO pricing model.

Estimates indicate a clear link between firms' spending and the probability of breaches, with zero spending by firms converging to an almost certain breach. Given the systemic nature of those risks, the link between cybersecurity spending and probability of breaches is also useful for the government in order to devise optimal communication strategies and evaluate the value of information about potential cyber attacks. The information about Cumulative Abnormal Returns is then used, together with probability of breach estimates to evaluate expected losses. The expected losses are estimated by multiplying both components and the sample indicates that for insuring a 10 billion dollar market capitalization the expected loss would be 432000 dollar with a standard deviation of 17.7 million dollar however. The expected loss evaluation, together with information about the spending of firms, can then be used by insurance companies to evaluate the risk premiums.

Our approach aims to go further by using derivative products to extract more details about the distribution. Our aim is to extract finer details about cyber risk distribution and make the link with pricing of the products and the value of the information through entropy as discussed earlier. The link stems from the pricing kernel and the link with entropy.

3. Complete Markets and State Prices

One of the fundamental problems of market economies is the coordination in the face of limited information and uncertainty. Theoretically, this problem was addressed by Arrow (1964) and Debreu (1959). They suggested that if contingent markets where assets that depend on future states of the economy could be traded, then such uncertainty could be traded making markets more efficient. A security actually is such an asset by essence. A security, however, takes different values in different states of the economy or the market and cannot isolate different states. This led to the concept of Arrow-Debreu securities that pay off 1 dollar in a particular state of the economy and 0 otherwise. Apart from all technicalities and mathematical elaborations, the idea is relatively straightforward.

The state of the economy is generally a variable of interest that reflects the best to worst states for the decision maker. In general applications, the level of the S&P500 is supposed to reflect the state of the US Economy. Now, if we can divide the S&P 500 into different levels from the lowest to the highest level by increasing by 1% steps, we can in principle create financial products that would pay 1 dollar in the lowest state for instance. This is an example of an Arrow-Debreu security and gives the investor one dollar in this worst-case scenario and nothing otherwise. It actually can be viewed as some kind of insurance that insures against the worst-case state. Now, what is the price of such an asset? First, imagine that there were hundred such assets, each one for each interval of 1%. If you would buy all those Arrow-Debreu Securities you would get 1 dollar for sure. This portfolio of Arrow-Debreu Securities is then non-risky and its value can be calculated by discounting backwards with the risk-free rate. The Arrow-Debreu security for each state is then given by the probability to get 1 dollar discounted at the risk-free rate. As the risk-free rate is known and the payoff of 1 dollar as well, if we knew market prices of such securities we could extract the implicit price to insure different states of the economy.

Example: The price of a security paying 1 dollar in a year in the worst 1% market situation is today 0.9 cents. The risk-free rate is 2%. Using a simple mathematical representation, the equation that solves for the implicit state price also called a risk neutral probability would be: $0.9 =$

$(StateProba * 1) * \frac{1}{(1+2\%)}$. *StateProba* is also called a State Price as it indicates, apart from the discounting, the price to insure a state of the economy.

A fundamental asset pricing theorem states that any asset can be priced consistently using state prices extracted from financial products. However, sometimes the number of independent assets is not enough to cover all the states. The markets are then said to be non-complete. In non-complete markets the price of some assets cannot be evaluated exactly. If, however, there are many independent assets one can eventually create a set of Arrow-Debreu securities that can be used to extract state prices. The market is then said complete. The degree of completeness of the market depends on the development of financial markets and the variable of interest of the decision maker. The US financial markets are in that sense the most complete markets.

Typically, the markets that are used to extract state prices are Options markets, notably Put and Call options. We focus here on Call Options that are generally used to extract state prices. This approach pioneered by Breeden and Litzenberger (1978) is used for Credit derivatives in Coval *et al*. (2009). A typical Call option pays off $(S_T - K)^+$, which is a compact way of indicating that it pays $S_T - K$ when it positive and otherwise 0. *K* is called the strike price, it is the price at which an asset can be bought. S_T is the price of the security at date T. So the Call option gives the right to buy a security at a predetermined price K and it is only interesting to buy if it is cheaper than the price in the market.

Interesting from a pricing and information extraction perspective, Breeden and Litzenberger (1978) have shown that the state price density can be extracted from Call option prices by calculating the second derivate of the call price function with respect to the strike price. In order to understand the intuition, consider the following example:

The S&P500 market moves between 80 and 120 and we have access to Call options with strikes from 80 to 120 and steps of 1. Imagine that we want to know the state price of the level of 85 in the market. If we buy two Calls with strikes 84 and 86 respectively and sell short two Calls with strkes 85, the payoff would be: $(S_T - 84)^+ + (S_T - 86)^+ - 2 * (S_T - 85)^+$. Such a strategy is called a Butterfly spread. Now consider three scenarios:

If $S_T \leq 84$, Payoff=0

If $S_T = 85$, Payoff =1

If $S_T \geq 86$, Payoff=0

Which in this case is an Arrow-Debreu Security in a discrete case with 1 value steps. The Butterfly spread strategy can be written in a different way:

$$[(S_T - 86)^+ - (S_T - 85)^+] - [(S_T - 85)^+ - (S_T - 84)^+] \quad (4)$$

Which is a difference of differences that in a continuous setting converges to the second derivative of the call price with respect to the strike price K. The price of an Arrow-Debreu security for state m_τ is thus

$$q(m_\tau) = \frac{\partial^2 C(K,\tau)}{\partial K^2} \quad (5)$$

Where τ is the time to maturity. The formula is quite simple when the asset prices respect the Black and Scholes (1973) conditions. In reality, however, derivative products exhibit so-called "smile" effects meaning that prices of calls and puts are not consistent with the Black and Scholes formula, especially in the tails of the distribution. The "smile" actually provides information about market expectations of volatility (rather than past estimates of volatility) and has led to the creation of an index called VIX. The VIX is based on the volatility estimates implicit in derivative products at one point of time. It should provide the latest expectation of upcoming volatility. Edwards and Preston (2017) provide an analysis of how the VIX can be used to forecast future volatility. Cyber

risks, however, might exhibit much more extreme risks than just volatility shifts and exhibit jump components, an issue analyzed in the next section.

4. Cyber risks and extreme value distributions

As pointed out in Zhan *et al.* (2015), despite its importance little progress in predicting Cyber risks has been done till recently in Zhan *et al*. (2013). The latter used the *gray-box prediction* to predict the cyber attack rate, namely the number of attacks per unit of time. *Gray-box* models take into account the statistical properties exhibited by the data, whereas *black-box* models do not take into account statistical properties. Zhan *et al*. (2013) found that Long-Range Dependence is exhibited by honeypot-collected cyber attack data. The *gray-box* model approach could apparently predict cyber attacks one-hour ahead of time with accuracy of 70 %to 80%. They note that the prediction errors can be attributed to the inability to predict large attack rates. Zhan *et al.* (2015) note that this phenomenon is due to extreme values that can be modelled with extreme value theory (EVT).
According to the authors, the prediction of cyber attacks has not been addressed in the literature due to:

- The rule of thumb that cyber risks are unpredictable
- The lack of real data
- The lack of readily usable prediction models

Zhan *et al.* (2015) point out that the grey-box methodology together with the EVT approach enables to provide a high accuracy of Cyber rate attacks. In their case, extreme values means high attack rates above a threshold line. In order to understand the behaviour of random variables above that threshold line (tail of the distribution), EVT provides a kind of Central Limit Theorem (CLT) for the tails of a distribution. It provides some kind of convergence to a distribution in the tails.
More formally, consider $M_n = max(X_1, \dots, X_n)$, the maximum value of a sum of *i.i.d.* random variables. M_n's are called Block maxima. The theory states that the only possible non-degenerate limiting distribution for normalized block maxima belongs to the generalized extreme value family (GEV). Its density function *df* is given by

$$H_\xi(x) = \begin{cases} exp\left(-(1+\xi x)^{\frac{-1}{\xi}}\right) & \xi \neq 0 \\ exp(-e^{-x}) & \xi = 0 \end{cases} \quad (6)$$

Where $1 + \xi x > 0$.
The formula can be adjusted to integrate a location μ and a scale parameter $\sigma > 0$.

$$H_{\xi,\mu,\sigma}(x) \coloneqq H_\xi\left(\left(\frac{x-\mu}{\sigma}\right)\right) \quad (7)$$

ξ is the shape parameter of the distribution and $H_\xi$defines the sub-type of distribution. GEV subsumes three types of distributions according to the values of ξ:
$\xi > 0$: Fréchet distribution which is heavy-tailed
$\xi = 0$: Gumbel distribution which is less heavy-tailed than the Fréchet distribution
$\xi < 0$: Weibull distribution which is short-tailed
The Weibull distributions have so-called finite endpoints $x_F = sup\{x \in \mathbb{R} \parallel F(x) < 1\}$
Convergence, here, implies that for some real constants d_n and c_n

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \leq \mathrm{X}\right) = \lim_{n\to\infty} F^n(a_n x + d_n) = \mathrm{H}(x) \;\; x \in \mathbb{R} \quad (8)$$

If convergence holds for some non-degenerate density function *H*, then *F* is said in the maximum domain of attraction of *H*. $\left(F \in MDA(H)\right)$. A Theorem due to Fisher-Tippet and Gnedenko states that if $F \in MDA(H)$ then *H* must be a distribution of type H_ξ and thus belong to the Generalized Extreme Value family. Tails of Fréchet distributions decay like a power function and the rate of decay is called the tail index $\alpha = \frac{1}{\xi}$. The tail index indicates whether the tail distributions exhibit infinite higher moments. For further details the reader can consult *Embrechts et al.(EKM) 1997.*
If X is a non-negative rv with $F \in MDA\left(H_\xi\right)$ for $\xi > 0$, then $E(X^k) = \infty$ for k > $\frac{1}{\xi}$. So basically if ξ is greater or equal than $\frac{1}{2}$ the variance goes to infinity. If ξ is greater or equal than $\frac{1}{4}$ the forth moment goes to infinity, but the variance is bounded.
The GEV distribution can now be "estimated" with block maxima, provided the distribution is more or less *i.i.d.* The true distribution of the n-block maximum $M_n$can be approximated for large enough n by a three-parameter GEV distribution $H_{\xi,\mu,\sigma}$. In order to "estimate" the distribution we assume that the dataset can be divided into *m* blocks of size *n*. The block maximum of the jth block is denoted $M_{nj}$and the respective dataset thus consists of $M_{n1}, \dots, M_{nm}$. The GEV distribution is typically estimated with the Maximum Likelihood Method (MLM).
If $h_{\xi,\mu,\sigma}$ is the density function of the GEV distribution, then the log-likelihood function is given by:

$$l(\xi, \mu, \sigma; M_{n1}, \dots, M_{nm}) = \sum_{i=1}^{m} ln\, h_{\xi,\mu,\sigma}(M_{ni}) =$$

$$-mln\sigma - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{m} ln\left(1 + \xi\frac{M_{ni}-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(1 + \xi\frac{M_{ni}-\mu}{\sigma}\right)^{-\frac{1}{\xi}} \quad (9)$$

Note that there is a trade-off between the size *n* and the number of block maxima. A higher *n* leads to a more accurate approximation of the block maxima distribution, while a large value of *m* provides more data for the ML estimation and thus leads to a lower variance in the parameter estimates.
Zhan *et al*. (2015) mix EVT with other time series modelling. As the data exhibits Long Range Dependence, which implies that the process is not memoryless as in the Poisson case and the autocorrelation decays slower than in the exponential decay, Zhan *et al*. (2015) use FARIMA (Fractional Integrated Moving Average) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) Models. The authors use EVT and the Time Series approach interactively, where EVT is used for long run prediction of extreme events and the time series approach for all types of events. Basically, there is supposed to be a stationary process of cyber events with extreme events on top of it.
Different models are tested:

- *M1*: The standard Generalized Pareto Distribution (GDP)
- *M2*: GDP with time-invariant shape parameter ξ but time-dependent scale parameter $\sigma(t) = exp\left(\beta_0 + \beta_1 log(t)\right)$
- *M3*: GDP with time invariant scale parameter σ but time-dependent shape parameter $\xi(t) = \gamma_0 + \gamma_1 t$
- *M4:* GDP with time-dependent parameters $\sigma(t) = exp\left(\beta_0 + \beta_1 log(t)\right)$ and $\xi(t) = \gamma_0 + \gamma_1 t$.

First the *M1* model is fitted to the data and if the model doesn't fit well other more sophisticated *M2,..,M4* models are fitted.

An important element of EVT is the quantile above which an event is evaluated as extreme. The maximum quantile is chosen to guarantee that at least 30 extreme events occur as this is a rule of thumb for reliable fitting. Then different quantiles are tested with the minimum quantile 20% less and steps of 5%. For instance, with 1000 observations the maximum quantile is $1 - \frac{30}{1000} = 97\%$ and the set of quantiles that are tested is 77%, 82%, 87%, 92% and 97%.
Zhan *et al*. (2015) conclude that FARIMA-GARCH models can accommodate the LRD and the extreme value phenomenon and can predict 1-hour ahead attack rates with substantive accuracy. This implies that tail risk measured by the shape parameter ξ has to be taken into account. As most standard financial market models such as the CAPM, however, are not taking into account those tail risks, we have to find financial products that can "extract" information about the pricing of tail risk. Such a framework has been developed for extreme risks in a standard financial markets framework.

5. Modelling extreme risks in financial markets

In line with our approach to extract information about the cyber risk distribution from derivative and structured products is the literature on the extraction of the catastrophe risks from equity index options. Those options provide evidence about how market participants value extreme events whether they happen in the dataset or not. By looking at options with different strike prices we learn more than by looking at the risk distribution alone. It is also interesting to highlight the importance of the notion of the pricing kernel, a pricing tool in finance, and the link made with entropy, an information-theoretic measure we discussed when analysing the value of information. We elaborate more in detail on this approach now.
A more structural approach to evaluate extreme risks such as disasters which is potentially useful to model cyber accumulation risk has been developed in Backus *et al* (2011). They add a jump component to a normally distributed random variable to model disasters. This approach had been developed by Barro (2006) to model macroeconomic disasters. Consider a variable g of interest that would be log-normally distributed with a jump component and would be specified the following way:

$$\log g_{t+1} = w_{t+1} + z_{t+1} \quad (10)$$

Where the components (w_t, z_t) are independent of each other and over time. The first component w_t is normally distributed and the second component is a jump component that is modelled as a Poisson mixture of normals. The first component could be seen as small attacks and the second as extreme attacks. The most important element of the latter process is *j* the number of jumps which takes nonnegative integer values with probabilities $e^{-\omega}\omega^j/j!$. The jump intensity parameter ω indicates the mean of *j*. Hence, conditional on *j*, the jump component is normal.

$$z_t|j \sim \mathcal{N}(j\theta, j\delta^2) \text{ for } j=0,1,2\ldots \quad (11)$$

For small ω, there is one jump per unit of time and it occurs with probability ω, The jump component is then well approximated by a Bernoulli mixture of normals. If ω is large however, there can be a significant probability of multiple jumps. The moment generating function (MGF) is typically used to generate moments of a distribution. The MGF for a random variable x and real variable s is:

$$h(s;x) = E(e^{sx}) \quad (12)$$

The cumulant-generating function, $k(s) = \log h(s)$, has the following power series expansion:

$$k(s;x) = logE(e^{sx}) = \sum_{j=1}^{\infty} \kappa_j(x) s^j/j! \quad (13)$$

This representation is derived using a Taylor series expansion of *k(s)* around *s=0*, where the "cumulant" κ_j is the j^{th} derivative of k at s=0. Those cumulants bear a close relationship with statistical moments. κ_1 is the mean, κ_2 is the variance and so on. Skewness γ_1 and kurtosis γ_2 can be calculated the following way:

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} \quad \gamma_2 = \frac{\kappa_4}{\kappa_2^2} \tag{14}$$

The cumulants of log g are given by the following terms:

$$\begin{aligned} \kappa_1 &= \mu + \omega\theta \\ \kappa_2 &= \sigma^2 + \omega(\theta^2 + \delta^2) \\ \kappa_3 &= \omega\theta(\theta^2 + 3\delta^2) \\ \kappa_4 &= \omega(\theta^4 + 6\theta^2\delta^2 + 3\delta^4) \end{aligned} \tag{15}$$

This use of cumulant-generating functions was suggested by Martin (2007). Those cumulants reflect complex combinations of parameters. For instance, negative skewness implies $\theta < 0$ (negative mean jump), but the magnitude depends also on other parameters such as ω (the probability of jumps) and δ, the dispersion of jumps. Apart from jumps that might occur with a different size in either direction, the disaster risk is measured by the extreme left tail of the distribution. The disaster is defined by log of the random variable being below a threshold *–b* for a large b>0. More formally, the probability of the disaster event $\mathcal{D}_b$is:

$$p(\mathcal{D}_b) = \sum_{j=0}^{\infty} p(\log g \leq -b|j) \cdot p(j) = \sum_{j=0}^{\infty} N\left(\frac{-b-\mu-j\theta}{\sqrt{\sigma^2+j\delta^2}}\right) \cdot \frac{e^{-\omega}\omega^j}{j!} \tag{16}$$

Where N(.) is the standard normal cumulative distribution function. It is interesting to note that a specific disaster probability can occur in different ways. For instance, if the probability of jumps ω is small, the probability of disaster $p(\mathcal{D}_b)$ is mainly approximated by the first two terms j=0 and j=1, and high value of the mean jump θ and the jump dispersion δ are needed for higher values of disaster probability. If ω is high, however, all the terms in the sum matter and small jump sizes can generate a high disaster probability value. More generally, the method used is "infinitely divisible" and can be for different time frames and thus to potentially model different types of cyber risks such as a big shock that does not often occur (low ω, high θ, high δ) or the accumulation in a short time frame of different small cyber shocks (high ω, low θ, low δ). The probability of jumps (cyber shocks in our case) will also depend on the interval considered and the focus of the time frame is thus important. Again, to highlight, such information can be modelled with statistical models that estimate the data generating process, but another approach is to extract the information from financial assets that are sensitive to the type of disasters of interest. We thus need to find assets that are sensitive to different types of cyber risks and/or 'complete" the market by creating financial assets that are sensitive to the cyber risks of interest.

6. Pricing Kernels and entropy

In order to extract probabilities from derivative products, we need to focus on the pricing of risks implicit in those products and this is done through the pricing kernel and the concept of risk neutral probabilities. The non-arbitrage theorem states that in any arbitrage-free environment, there exists a positive random variable *m*, called a pricing kernel, that satisfies the following relationship for gross returns r_t^j and all traded assets *j*.

$$E_t\left(m_{t+1} r_{t+1}^j\right) = 1 \tag{17}$$

This pricing kernel provides a link between statistical distributions and evaluations that depend decision makers' utility functions. In macro-economic applications, the pricing kernel is typically modelled directly with utility functions, whereas in finance it is typically extracted from derivative products. In asset markets, the pricing kernel properties can be inferred from asset returns. A notable example is the Hansen-Jagannathan (1991) bound. A similar 'entropy bound" has been developed by Alvarez and Jermann (2005). Those measures relate dispersions of pricing kernels to expected differences in returns. The entropy of a positive random variable x can be defined thus:

$$L(x) = logEx - Elogx \tag{18}$$

Alvarez and Jermann (2005) show that the entropy bound links the entropy of the pricing kernel to expected differences in log returns in the following way:

$$L(m) \geq E\left(logr^j - logr^1\right) \tag{19}$$

For any asset *j* with positive return where r^1 is the gross return on a one-period risk free bond. This relationship thus relates estimates of return premiums to entropy bounds of the pricing kernel. As entropy is a general measure of dispersion it can be used as a flexible measure of departures from normality. Backus *et al*. (2011) highlight that the entropy of a pricing kernel can be expressed in terms of cumulant-generating function and cumulants of log *m*.

$$L(m) = logE(e^{logm}) - Elogm = k(1; logm) - \kappa_1(logm) = \sum_{j=2}^{\infty} \kappa_j(logm)/j! \tag{20}$$

It can be shown that if $logm$ is normal, entropy is one half the variance, but in general there will be contributions from higher moments.

In the derivative pricing literature, the pricing kernel is generally not directly specified but known as the risk neutral probability distribution that can be extracted from derivative products. The risk neutral probability bears a close relationship with Arrow-Debreu state prices. If we consider finite number of states *x* and historical probabilities $p(x)$, the above-mentioned pricing kernel equation can be expressed thus:

$$E\left(mr^j\right) = \sum_x p(x)m(x)r^j(x) = 1 \tag{21}$$

A one period bond's return can be expressed thus:

$$q^1 = Em = \sum_x p(x)m(x) = \frac{1}{r^1} \tag{22}$$

So-called risk neutral probabilities can be rewritten using the reference bond price:

$$p^*(x) = \frac{p(x)m(x)}{Em} = \frac{p(x)m(x)}{q^1} \tag{23}$$

It is important to highlight that those $p^*(x)$ are not true state probabilities but are called risk neutral probabilities because they sum to one. They are actually pricing tools and indicate the price today of getting a unit of account in the respective states. The pricing relationship can now be rewritten with the risk neutral probability distribution.

$$q^1 \sum_x p^*(x)r^j(x) = q^1 E^* r^j = 1 \tag{24}$$

Where $E^*$is the expectation under the risk neutral probability measure. This risk neutral probability can thus be used to evaluate financial products and the asset pricing theorem states that if the market is complete and without arbitrage this measure is unique.

The pricing kernel $m(x)$ can now be written as a function of the true and risk neutral probabilities, the equivalent of the Radon-Nykodim derivative in a continuous setting.

$$m(x) = \frac{q^1 p^*(x)}{p(x)} \tag{25}$$

And as q^1 is assumed constant, the entropy of the pricing kernel is

$$L(m) = L\left(\frac{p^*}{p}\right) = logE\left(\frac{p^*}{p}\right) - \text{Elog}\left(\frac{p^*}{p}\right) = -\text{Elog}\left(\frac{p^*}{p}\right) \tag{26}$$

As $E\left(\frac{p^*}{p}\right)$=1.

$L\left(\frac{p^*}{p}\right)$ is the entropy of p^* relative to p and can also be expressed as a function of cumulants.

$$L\left(\frac{p^*}{p}\right) = k\left[1; \log\left(\frac{p^*}{p}\right)\right] - \kappa_1\left[\log\left(\frac{p^*}{p}\right)\right] = \sum_{j=2}^{\infty} \kappa_j \left[\frac{\log\left(\frac{p^*}{p}\right)}{j!}\right] = -\kappa_1\left[\log\left(\frac{p^*}{p}\right)\right] \quad (27)$$

Those links imply that any asset *j* can be priced using different approaches: the pricing kernel approach, the historical probability and the risk neutral approach together with the risk free rate. Those approaches can be used interactively depending on the problem to be solved and the information at hand. The link with entropy and the cumulants can eventually be useful to determine the value of information for investors that do not have access to detailed information from financial products such as government sponsored OTC for instance. As shown in Backus *et al*. (2011) in the case of equity index options the ratio of risk neutral versus historical probabilities can be used to extract information about priced-in extreme risks. In our case, we are interested in using structured products with a portfolio indexed on cyber risks to extract the loss distribution from different tranches of those products.

7. Towards a Market for Cyber Indexed Structured Products

As alluded to in the core text, Cyber risks bear a close relationship with credit risks, even though the uncertainty about the loss distribution is much higher as the event can have a much broader systemic impact. As the market for credit risk products is extensive and tools to extract information have been developed, the same tools could be used for cyber risks provided such a market be created. First, it is important to distinguish two types of "markets' and both can be used interactively. There exist "Over the counter" (OTC) markets where assets are traded informally between parties. It is basically a contract between two parties but it is not traded on a formally organized market. Then there are also exchange traded products that are traded on a formally organized trading platform. The products traded on the trading platform are more liquid and they enable a better price discovery but are more standardized. The advantage of the OTC market is that any condition can be specified, even though some standards exist such as through ISDA. It would be interesting to specify similar guidelines for the Cyber indexed products.

The standard products are some kind of swap that enables one counterparty to insure itself against some risk against the payment of a premium. The most used product in credit risk is the Credit Default Swap (CDS), where one counterparty receives a premium to insure the losses when the credit risk hits. The premium will be dependent on the expected losses which are estimated by evaluating the probability of default and the loss given default. Once the market exists, the observed premia can be used to evaluate the risks expected by the market. More interestingly, however, more complex products can be created to extract information about the different types of risk. Of importance in the Cyber domain might be First party coverage, Third party coverage and post incidence expenses (Enisa (2017)) as well as Accumulation Risk (Cambridge Center for Risk Studies (2016)).

First, different kind of swaps could be specified that take into account those types of risks. The development of correlation products such as first layer tranches and CDOs can be used to extract uncertainty concerning accumulation risk. A Collateralized Debt Obligation (CDO) is basically a portfolio of assets whose risks are "sold" to different counterparties who will cover different types of risks. There are different tranches with different degrees of riskiness where the first tranche covers the first risks against payment of a premium. As risks accumulate other more senior tranches will cover the risks. As the risks are first covered by the more risky tranches (Equity) the more senior tranches are less risky and bear a lower premium. Given that with risk accumulation, less

risky tranches are impacted step by step, the premiums implicit in those tranches could in principle be used to extract market expectations about Cyber accumulation risk and extract the expected loss distribution. More formally, the pricing approach of a CDO works the following way.

The principle behind CDOs is that a portfolio of risks is sold to investors. In order to make the risks acceptable for different kinds of investors, the relevant risk (in our case cyber risks) of the portfolio underlying the CDO is sold in different *tranche*s. A *tranch*e is defined by lower (K_L) and upper (K_U) attachment points that cover a fraction of the losses in the reference portfolio. An investor is thus exposed to losses that are limited to K_U-K_L percent of the initial portfolio value. In order to illustrate the functioning of CDOs, consider the initial value of the portfolio *P* of underlying credit risks. L_t denotes the percentage of losses of the portfolio at date t. At any date the total loss in value of the portfolio is thus L_tP. Now, we consider j tranches and $L_{j,t}$ denotes the loss of the holder of tranche j up to time t. The loss of tranche j up to time t can now be expressed in the following way:

$$L_{j,t} = min\left\{L_t, K_{U_j}\right\} - \left\{L_t, K_{L_j}\right\} \tag{28}$$

Losses are in general paid with a certain frequency. For instance, each quarter the tranche holder pays the losses on the tranche, realized since the last payment date. If we consider the timeframe *dt*, at *t+dt* the tranche holder will have to pay the fraction $L_{j,t+dt} - L_{j,t}$ of the CDO with notional P.

For the risks that the tranche holder bears, he gets a premium π_j that he receives with a frequency *dt*. The outstanding notional of tranche j at time t is given by $P_{j,t}$ and evolves the following way:

$$P_{j,t} = \left(K_{U_j} - K_{L_j}\right)P - L_{j,t}P \tag{29}$$

Hence at each payment date during the life of the CDO the tranche holders receive $\pi_j dt P_{j,t}$ and pay $L_{j,t} - L_{j,t-dt}$. Whereas the premium π_j is constant, the outstanding notional $P_{j,t}$ evolves with time. As soon as the percentage loss of the portfolio P_t exceeds the upper attachment point the respective *tranche* is eliminated.

Pricing of CDOs consists in finding an appropriate premium π_j for the j different tranches. CDOs are priced like swaps or CDS. Thus the net present value of cash flows received and paid is zero so that there is no upfront payment. Two legs are distinguished: fixed and floating legs. The fixed leg represents the payments that the tranche holder receives, whereas the floating leg represents the payments that they disburse. More formally, consider a CDO with different payment dates $\{t_1, \dots, t_K\}$. The contract starts at t_0. The value of the fixed leg is then given by:

$$FixL_j = \sum_{k=1}^{K} e^{-(r_{tk}-r_0)}\,\pi_j dt E\left[\left(K_{U_j} - K_{L_j} - L_{j,t_k}\right)P\right] \tag{30}$$

The value of the floating leg is given by:

$$FloatL_j = \sum_{k=1}^{K} e^{-(r_{tk}-r_0)}\,E\left[\left(L_{j,t_k} - L_{j,t_{k-1}}\right)P\right] \tag{31}$$

Since there is no upfront payment at t_0 the premium is given by

$$\pi_j = \frac{\sum_{k=1}^{K} e^{-(r_{tk}-r_0)} E\left[\left(L_{j,t_k} - L_{j,t_{k-1}}\right)P\right]}{\sum_{k=1}^{K} e^{-(r_{tk}-r_0)} dt E\left[\left(K_{U_j} - K_{L_j} - L_{j,t_k}\right)P\right]} \quad (32)$$

Given interest rates, attachment points and payment dates, computation of tranche premium rests on evaluating the expectations of tranche losses. In order to determine the distribution and expectation of $L_{j,t}$ we have to map the losses on characteristics such as individual cyber event probabilities, loss given cyber event and event correlations within the reference portfolio.

We adapt the Vasicek single factor model for cyber risks. Consider a portfolio of N firms where each firm might be hit by a cyber shock as soon as some random variable drops below some reference level (RL). In order to derive a closed form for the cyber shock probability of the portfolio, namely the fraction of firms in the portfolio hit at at time t, we transpose the Vasicek (2002) assumptions for cyber risks. In order to determine the portfolio cyber risk, we need to model the individual cyber risks as well as the correlation between firms' cyber risks. Those correlations are affected by the random variables and are typically modelled using factor models. Vasicek assumes that the correlation coefficient between any pair of random variables is the same for any two obligors, hence:

$$corr\left(dw_{n,t}, dw_{m,t}\right) = \rho_{n,m,t} = \rho_t \quad \forall n \neq m \quad (32)$$

There exists a common correlation coefficient that partly drives all the firms' cyber risks. The idea is that there exists a random systemic factor that affects the cyber risk of all firms. The random variables triggering the cyber shock can now be written:

$$dw_{n,t} = \sqrt{\rho} M_t + \sqrt{1-\rho}\varepsilon_{n,t} \quad (33)$$

This shock process can be linked with equation (10) if we are ready to model $\varepsilon_{n,t}$as a more complex not normally distributed random variable.

The cyber shock is determined by two factors: M_t which is a common or systemic factor and $\varepsilon_t$which is an idiosyncratic factor. M_t generates the cyber risk dependency as it affects all obligors in the same way. This model is equivalent to the Gaussian copula. The cyber risk probability at time t can now be evaluated by conditioning on the common factor M_t.

$$\begin{aligned} p_n(M_t) &= P\left[dw_{n,t} < K_{n,t} | M_t\right] \\ &= P\left[\varepsilon_{n,t} < \frac{K_{n,t} - \sqrt{\rho} M_t}{\sqrt{1-\rho}} | M_t\right] \\ &= \Phi\left(\frac{K_{n,t} - \sqrt{\rho} M_t}{\sqrt{1-\rho}}\right) \end{aligned} \quad (34)$$

Conditionally on M_t, the cyber risk probabilities are independent. It is assumed that we know the individual cyber risk probabilities and we can calculate $K_{n,t} = \Phi^{-1}\left(p_{n,t}\right)$. Note that as $p_{n,t}$ is a function of market spreads, we can directly infer the cyber risk probabilities from the market data. If we further assume that the cyber risk probabilities are equivalent for all firms, we get the following expression for conditional default probability:

$$P(M_t) = \Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{\rho}M_t}{\sqrt{1-\rho}}\right) \quad (35)$$

Now, consider the fraction of losses due to the cyber risks on the portfolio at time *t*, FL_t . The unconditional distribution function of a portfolio characterized by a cyber risk probability p and correlation ρ is

$$F(FL^*|p_t, \rho_t) = P(FL_t \leq FL^*) \quad (36)$$

If we assume that the number of firms tends to infinity, then due to the conditional independence assumption the fraction of losses FL_t converges to the individual default probability $p(M_t)$. The unconditional distribution function of the fraction of losses can hence be expressed thus:

$$\begin{aligned} F(FL^*|p_t, \rho_t) &= P(FL_t \leq FL^*) \\ &= P(p(M_t) \leq FL^*) \\ &= P\left[\Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{\rho}M_t}{\sqrt{1-\rho}}\right) \leq FL^*\right] \\ &= P\left[M_t \geq \left(\frac{\Phi^{-1}(p_t) - \sqrt{1-\rho}\Phi^{-1}(FL^*)}{\sqrt{\rho}}\right)\right] \\ &= 1 - \Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{1-\rho}\Phi^{-1}(FL^*)}{\sqrt{\rho}}\right) \\ &= \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(FL^*) - \Phi^{-1}(p_t)}{\sqrt{\rho}}\right) \end{aligned} \quad (37)$$

By evaluating this formula for different values of cumulative loss fractions FL^*, the formula can be used to construct the density of losses. This formula can now be used to evaluate the expected losses due to cyber risks. If we are interested in knowing the cyber risk distribution of the portfolio of underlying firms we can calibrate the average cyber risk probability and the most important variable is the correlation parameter as it is sufficient to generate different types of risk distributions. The model can now be applied to different tranche sizes to calculate the expected losses per tranche

$$E(L_{j,t}) = E\left[min\left\{L_t, K_{U_j}\right\} - min\left\{L_t, K_{L_j}\right\}\right] \quad (38)$$

As tranche premia bear a close relationship with expected losses and will indicate the expected losses per tranche (under the risk neutral measure). In the standard model it is assumed that the loss given event LGE_t is the same for each firm in the portfolio. L_t is then evaluated by multiplying the portfolio cyber event rate by the *LGE*. The prices of the loss tranches can now be evaluated by putting the evaluation of the following expectation in the formula:

$$E(L_{j,t}) = E\int_0^1 \left[min\{FL^* \cdot LGE, K_{Uj}\} - min\{FL^* \cdot LGE, K_{Lj}\}\right] dF(FL^*|p, \rho) \quad (39)$$

Where the integral is typically numerically integrated.

The premia of the tranches thus depend on the following elements:

- Attachment points $K_{L_j};\ K_{U_j}$
- The term structure of interest rates $e^{-(r_1-r_0)},\ldots,\ e^{-(r_k-r_0)}$
- The frequency of payments
- Default probabilities $p_{t_1},\ldots,p_{t_k}$
- Event correlations $\rho_{t_1},\ldots,\rho_{t_k}$
- Losses given event $LGE_{t_1},\ldots,LGE_{t_k}$
- The distribution function $F(FL^*|p,\rho)$

Those variables are then fed into the formulas to simulate the loss distribution and the impacts on tranches. Our approach consists in using different information sources and to invert the approach to extract information about risk neutral cyber loss distribution from prices of tranches.

8. A procedure to extract the Cyber risk distribution

From the foregoing developments, it is clear that the prices of tranches reflect expected losses within the different tranches. Typically, the tranches are evaluated by simulating the expected losses within the tranches, but we could also extract implicit expected losses from the tranche prices. Moreover, if we could create products with small enough tranche sizes, we could extract a finer and finer loss distribution and isolate some kind of risk neutral loss probability distribution which would indicate the Arrow-Debreu state price for different levels of losses. If such a distribution is compared to more direct modelling approaches such as those developed in Zhan *et* al. (2015) and Zin *et al*. (2018) it can provide information about risk neutral versus historical probabilities. Moreover, such information can be used to evaluate the value of the information for a decision maker having access to the historical information only and provides information about the implicit pricing of extreme cyber risks. Our approach could be viewed as the structured products equivalent to the Breeden and Litzenberger (1978) approach where the tranches sizes play the role of different strike prices to isolate state prices. Much as the volatility smile, the correlation skew can be used to diagnose information about standard model specification. The base correlation approach in credit risk models can then be applied to analyse expected losses by tranche.

The base correlation approach has been introduced by McGinty et al. (2004). Even though, the base correlation approach is highly questionable, it has become a kind of market standard, notably as a calibration tool to value bespoke CDO´s. Basically, market spreads for the respective tranches are used to calculate the expected losses for those tranches. Those tranches are then combined to form wider first loss tranches and the model then iterates to find the base correlation that generates the expected loss of this first loss tranche.

More formally, consider tranches with attachment points $K_L$and. K_U Then expected losses of a mezzanine tranche can be expressed thus:

$$E[L(K_L,K_U)] = E[L(0,K_U)] - E[L(0,K_L)] \tag{40}$$

For the $(0,K_L)$ tranche we can evaluate the implied correlation that generates the spread observed in the market. It is the base correlation for attachment point K_L. Given that base correlation we can fix the expected loss for that respective tranche. Given the spread of the (K_L,K_U)tranche, we can iterate over the correlation parameter ρ until the expected loss of the fictive tranche $(0,K_U)$ is such that the expected loss of the (K_L,K_U) tranche in the above equation is in line with that implied by the spread.

This approach can now be applied to a CDO on a portfolio of cyber risks with finer and finer tranches. Instead of calculating using the expected losses in equation (40) to price the tranches, the approach is inverted and the market prices are used to infer expected losses by tranche. We observe a set of prices of tranches and if the tranches are fine enough the prices can be viewed as the Arrow-Debreu state prices and the risk neutral distribution. The more information we have about the historical past probability distribution p, the more we can compare both distributions, the risk neutral p^* and the historical p to infer information about potential cyber shocks and the pricing as in equation (24). Moreover, the equation (27) can be used to evaluate the cross-entropy between p^* and p. As explained in Section1, If we view the distribution p as the prior information held by the "industry", entropy-based approaches can now be used to infer the value of the information for a data broker that would sell the information. Entropy informativeness in equation (3) can now be used

$$I(\alpha) = H(p) - \sum_s p_\alpha(s) H(q_\alpha^s)$$

Where *H(p)* is the entropy for prior historical information p and $\sum_s p_\alpha(s) H(q_\alpha^s)$ would be given by different expected values of the entropy of the risk neutral distribution extracted from different structured products. As discussed in an earlier section, this entropy measure can also be linked to the pricing kernel. The excess return should be directly linked to the entropy of the pricing kernel. The pricing kernel $m(x)$ can be reconstructed using equation (25). The entropy of the pricing kernel is given in eq. (26). and is expressed as a function of cumulants in (27). The cumulants in (14) and (15) provide a flexible parametrization to characterize the relevant distribution.

Conclusion

This paper is a first theoretical attempt to apply theoretical finance techniques to the domain of cyber risk. We formalize an information extraction and valuation problem and provide theoretical guidance on the creation and the techniques that would permit the extraction of expected cyber risks. Such information could be useful for insurance companies to develop and price products. Public authorities also might use such a market to monitor potential threats of the type that have never occurred. The approach is consistent with a recently emerging literature that elicits information from different types of derivative markets. The links between pricing kernel, distributions and entropy enable us to formalize the information theoretic approach with the pricing problem. Future research will focus on implementations of the methods.

References

Abramowitz, M. and Stegun, I.A. (1965) *Handbook of Mathematical Functions.*
Acharya, V., Pederson, L., Philippon, T. and Richardson, M. (2017) "Measuring Systemic Risk", *The Review of Financial Studies,* Volume 30, Issue 1, Pages 2–47.
Alvarez, F. and Jermann, U. (2005) "Using asset prices to measure the persistence of the marginal utility of wealth", *Econometrica,* 73, 1977-2016.
Arrow, Kenneth J. (1964). "The Role of Securities in the Optimal Allocation of Risk Bearing." *Review of Economic Studies,* 31(2): 91–96.
Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999) "Coherent measures of risk" *Math Fin.* 9(3), 203-228.
Azrieli, Y. and Lehrer, E. (2008) "The Value of Stochastic Information Structure", *Games and Economic Behavior* 63 (2): 679-693.
Backus, D., Chernov, M. and Martin I. (2011) "Disasters implied by Equity Index Options", *The Journal of Finance,* Vol. LXVI, No 6.
Barro, R. J. (2006) "Rare disasters an d asset markets in the twentieth century", *Quarterly Journal of Economics* 121, 823-867.
Bergemann, D. and Morris, S. (2019) "Information Design: A Unified Perspective", *Journal of Economic Literature,* 57(1), 44-95.
Black, F. and Scholes, (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy,* 81(3), 637-654.
Blackwell, D. (1953) "Equivalent Comparisons of Experiments", *Annals of Mathematical Statistics,* 24; 265-72.
Breeden, D. and Litzenberger, R. (1978) "Prices of State Contingent Claims Implicit in Option Prices", *Jouranl of Business,* 51(4), 621-651.
Cabrales, A., Gossner, O. and Serrano, R. (2013) "Entropy and the Value of Information for Investors", *American Economic Review,* 103(1), 360-377.
Cambridge Center for Risk Studies (2016) "Managing Cyber Insurance Accumulation Risk"
Collin-Dufresne, O., Goldstein, R. and Yang, F. (2012) "On the Relative Pricing of Long-Maturity Index Options and Collateralized Debt Obligations" *The Journal of Finance,* Vol LXBII, No 6, 1983-2013.
Constantinides, G., Jackwerth, J. and Savo,v, A. (2013) "The puzzle of index option returns", *Review of Asset Pricing Studies* 3, 229-257.
Coval, J., Jurek, J and Stafford, E. (2009) "Economic Catastrophe Bonds", *American Economic Review* 99:3, 628–666
Debreu, G. (1959) *The Theory of Value: An Axiomatic Analysis of Economic Equilibrium,* Yale University Press.
Edwards, Tim and Preston, H. (2017) "Reading VIX: Does VIX predict future volatility", S&P Global Working Document.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance,* Springer.
EIOPA (2018) "Understanding Cyber Insurance – A Structured Dialogue with Insurance Companies", Publications office of the EU.
ENISA (2017) "Commonality of risk assessment language in cyber insurance: Recommendations on Cyber Insurance".
Froot, K. A. (2008) "The Intermediation of Financial Risks: Evolution in the Catastrophe Reinsurance Market", *Risk Managementt and Insurance Review,* Vol. 11, 281-294.

Ganuza, J. and Penalva, J. (2010) "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions", *Econometrica,* 78(3), 1007-1030.
Gordon, L.A. and Loeb, M. P. (2002) "The economics of information security investment", *ACM Trans. Inf. Syst. Secur.* 5(4), p. 438-457.
Gordon, L.A., Loeb, M.P., Lucyshyn, W. and Zhou, L. (2015) "Externalities and the Magnitude of Cyber Security Underinvestment by Private Sector Firms: A Modification of the Gordon-Loeb Model", *Journal of Information Security* 6, p. 24-30.
Hansen, L.P. and Jagannathan, R. (1991) "Implications of security market data for models of dynamic economies", *Journal of Political Economy,* 99, 225-262.
Harrison, J. and Kreps, J. (1979) "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, Vol. 11, p. 418-443.
IAS (2016) "Issues Paper on Cyber Risk to the Insurance Sector", IAIS publication
Jurek, J. and Stafford, E. (2015) "The Cost of Capital for Alternative Investments", *The Journal of Finance,* Vol. 70; Issue 5, page 2185-2226.
Karatzas, I. and Shreve, S. (1988) "Brownian Motion and Stochastic Calculus", *Springer Verlag*.
Kojen, R. and Yogo, M. (2015) "The Cost of Financial Frictions for Life Insurance", *American Economic Review,* 445-475.
Lin, Z., Parsa, R. , Rees, J.U. and Sapp, T.R. (2018) "Pricing Cyber Security Insurance: A Copula Model Using an Objective, Verifiable, Loss Measure", *SSRN Working Paper Series*.
Longstaff, F, and Piazzesi, (2004) "Corporate earnings and the equity premium", *Journal of Financial Economics,* 74, 401-421.
Marsh (2016) "Continental European Cyber Risk Survey: 2016 Report", Marsh LLC, October.
Martin, I. (2007) "Disasters and asset pricing: Evidence from option markets", Working Paper, Harvard University.
McGinty, L. and Ahluwalia R. (2004) "A Model for base correlation calculation", Credit Derivatives Strategy JP Morgan
Merton, R. (1987) "A simple model of capital market equilibrium with incomplete information", *Journal of Finance* 42, 483-510.
Nelson, R. B. (2006) *An Introduction to Copulas,* Second Edition, Springer Series in Statistics.
O'Hara, M. (2015) "High Frequency Market Microstructure", *Journal of Financial Economics 116,* 257-270.
Parsa, R. and Klugman, S. (2011) "Copula Regression", *Variance,* Vol 5. Issue 1.
Rietz, Th. (1988) "The equity risk premium: A solution", *Journal of Monetary Economics,* 22, 117-131.
Schönbucher, Ph. (2003) *Credit Derivatives Pricing Models: Models, Pricing and Implementation,* Wiley Finance.
Shannon, C. (1948) "A Mathematical Theory of Communication", *Bell System Technical Journal* 27: 379-423, 623-56.
Wakker, P. (2010) *Prospect Theory for Risk and Ambiguity*, Cambridge University Press.
Zhan, Z., Xu, M. and Xu, Sh. (2015) "Predicting Cyber Attack Rates with Extreme Values", *IEEE Transactions on Information Forensics and Security,* Volume 10 , Issue 8.
Zhan, Z., Xu, M. and Xu, Sh. (2013) "Characterizing honeypot-captured cyber attacks: Statistical framework and case study", *IEEE Transactions on Information Forensics and Security,* Volume 8 , Issue 11.

Verlaine On the extraction of Cyber Risks from Structured Products

DECISION SCIENCES INSTITUTE

Profit-Based Credit Models with Risk-Neutral, Risk-Averse and Loss-Averse Lending

Edwin Baidoo
Tennessee Technological University
Email: ebaidoo@tntech.edu

Ramachandran Natarajan
Tennessee Technological University
Email: RNAT@tntech.edu

ABSTRACT

In recent years, emphasis is on credit models that consider the profit motive of the lender. The objective of this paper is to consider credit scoring models that address the profit maximization objective, risk aversion and loss aversion on part of the lender. The results of the models discussed in this paper show that for risk-neutral lending the profitability approach yields better results compared to the model that considers only the default risk. However, loss-averse lending leads to reduction in profits because disproportionately more weight is put on avoiding a possible loss than on a potential gain of equal magnitude.

KEYWORDS: Credit Scoring Models, Default-based Credit Scoring, Profit-based Credit Scoring, Risk Aversion, Expected Utility Maximization, Loss aversion

INTRODUCTION

A review of the literature on credit scoring models suggests that the focus of most of the studies has been on estimating/predicting the credit default risk and on actions to mitigate the potential loss due to default. This emphasis has shifted in recent years with credit models that consider the profit motive of the lender. Although traditional credit scoring methods have largely focused on default risk, it represents only one aspect of the entire credit granting process. The main objective – referred to as profit scoring in the literature – is the lender's goal to maximize profit given predicted default probability conditioned on borrower characteristics, Thomas (2000). In this regard, default probabilities alone may not be a sufficient indicator for profit, although there may be a strong correlation between higher levels of default probabilities and lower profit levels and vice versa. According to Hand and Henley (1997), there is a gradual change from default-based credit scoring models to profitability approach in credit modeling. This is largely driven by the observation that the statistical loss function to predict default risk - for instance, maximum likelihood estimator - does not effectively capture the lender's objective. Finlay (2008, 2010) put it better when he describes existing default-based models as being, at best, a crude approximation to the real objective of the lender, which is to identify customers who will contribute to some profitability metric.

LITERATURE REVIEW

The history of default risk modeling in consumer lending is rich and continues to grow. Often, the goal is to use statistical techniques to predict the probability of default in order to better understand the population under study. In this tradition, Cyert, Davidson, and Thompson (1962) used Markov Chains in their analyses of doubtful accounts. They binned lender's account

receivables by age and then modeled the loss expectancy rate within each bin. Combining the two stages, a lender would now be able to set aside an allowance in anticipation to potential losses given a default event.

Hand and Kelly (2002) introduced the concept of super-scorecards, a classification ensemble of individual credit scoring models that yielded superior results when compared to its components. An attractive feature of this approach is that an arbitrary number of standalone models could be used in the construction of the super-scorecard. Also, the proposed model is more flexible but retains the ease of interpretability of the standard linear scorecard.

Machine Learning models have been applied more recently with significant positive results. Using a Classification and Regression Tree (CART) model on combined consumer banking transaction and credit bureau data, Khandani, Kim, and Lo (2010) were able to detect non-linear relationships that a traditional logistic or discriminant model would not be able to find. The model's accuracy allowed them to predict default events three to twelve months in advance. Because machine learning models in the credit space are vast, Lessmann et al. (2015) and Baesens, Roesch, and Scheule (2016) provide an excellent overview.

Owing to the fact that borrowers and lenders do not often share the same objective, Keeney and Oliver (2005) approached the credit issue by using Cooperation Negotiation Analysis and Efficient Frontier Curves to develop a model that identifies and integrates both borrower and lender preferences. For example, a borrower's preference for a lower loan price may be matched with a lender's preference for profit or market share. The outcome is a win-win product that has the potential to significantly decrease the probability of default.

The choice of a profitability metric plays an important role in profit scoring. While there is no universally accepted profitability metric, those that are used are often robust and well established in the literature. For example, Serrano-Cinca and Gutierrez-Nieto (2016) used the Internal Rate of Return (IRR), Barrios, Andreeva, and Ansell (2014) used Customer Lifetime Value (CLV), Lieli and White (2010) used the Net Present Value (NPV) and Maldonado et al. (2017) used the Return on Investment (ROI). In this paper, we use the NPV as well.

MODELS

Credit Model with Risk-neutral Lender

We follow the structure and notations presented in Lieli and White (2010) and define the following:

- Let $\lambda > 0$ represent the loan principal
- Let $r \in (0,1)$ be represent the interest rate on the loan
- Let t be represent the duration on the loan

Therefore, a loan is characterized by a vector $\ddot{X} = (\lambda, r, t)$. Assume further that the lender only issues a conventional loan that is payable in equal monthly installment and that in the event of default, the lender stands to recover a fraction of the principal. Define this recovery rate as $q \in [0,1]$.

Let $\tilde{X}$ denote features from the sample data that is sufficient to predict borrower default risk. Then, we define the binary default risk variable as follows:

$$Y = \begin{cases} +1 & \text{for a good borrower} \\ -1 & for \text{ bad borrower} \end{cases} \quad (1.1)$$

where a good borrower is defined to be those who did not default on the loan, while a bad borrower is defined as one who defaulted. It is important to realize that Y, the outcome variable is only observed at loan maturity. In other words, the lender does not definitively know Y at the time of the loan's origination.

Consider the following definitions:

- Let *D* = {*A, R*} represent the lender's decision to accept or reject a loan application
- Let $\pi_{\{d \in D,\ y \in Y\}}$ represent the lender's profitability metric as a consequence of their decision to accept or reject a loan application and their classification of the borrower at the end of loan duration as good or bad.

Then, Table 1.1 gives an overview of the framework of the lender's profitability. Here, the lender is only profitable when they accept a good borrower and stands to lose money when they accept a bad borrower. In the case where a borrower is rejected, the profit is defined to be zero. In other words, profit is calculated only for those applicants who are accepted.

Table 1.1: Lender's Profitability Scenarios

	No Default (Y=1)	Default (Y=-1)
Accept (A)	$\pi_{A,1}\left(\tilde{X} = \tilde{x}, \ddot{X} = \ddot{x}\right) \geq 0$	$\pi_{A,-1}\left(\tilde{X} = \tilde{x}, \ddot{X} = \ddot{x}\right) < 0$
Reject (R)	$\pi_{R,1}\left(\tilde{X} = \tilde{x}, \ddot{X} = \ddot{x}\right) = 0$	$\pi_{R,-1}\left(\tilde{X} = \tilde{x}, \ddot{X} = \ddot{x}\right) = 0$

Although loan profitability is conditioned on the lender's decision and the loan outcome, it is also a function of borrower's features in the sample data as well as the characteristics of the loan contract, such as loan rate, amount, and duration.

For a conventional loan with equal payments, we choose the Net Present Value (NPV) as the lender's profitability metric. For such a loan, let CF_k represent the cash flow or the equal installment payable at time *k*, where *k* ⊆ *t*. Then, the NPV is defined as

$$\pi = \sum_{i=1}^{i=t} \frac{CF_k}{(1+r)^i} - \lambda \quad (1.2)$$

Rearranging Equation 1.2 gives the following

$$\pi = CF_k \sum_{i=1}^{i=t} (1 + r)^{-i} - \lambda \quad (1.3)$$

because the lender is only profitable when they accept a good borrower, the second quadrant in Table 1.1 is calculated as follows

$$\pi_{1}\left(\tilde{X}=\tilde{x},\ddot{X}=\ddot{x}\right)=CF_{k}\sum_{i=1}^{i=t}(1+r)^{-i}-\lambda>0 \tag{1.4}$$

That is, the lender always stands to make a profit when they accept a good borrower. However, in the event of a default, the assumption is that the lender can recover only a fraction of the loan. In this case, the lender can invest the recovered amount in a risk-free government note of the same maturity as the loan. If we denote the return on the government note as r_g then it should be noted that $r_g < r$.

The loss associated with the default event is calculated as follows

$$\pi_{-1}\left(\tilde{X}=\tilde{x},\ddot{X}=\ddot{x}\right)=q*\lambda(1+r_{g})^{-t}-\lambda<0 \tag{1.5}$$

Rearranging Equation 1.5 gives the following

$$\pi_{-1}\left(\tilde{X}=\tilde{x},\ddot{X}=\ddot{x}\right)=\lambda\left(q*(1+r_{g})^{-t}-1\right)<0 \tag{1.6}$$

Because the lender cannot observe Equation 1.1 at the time of the loan application, their best approach is to decide based on expected profit or loss. Let $\alpha = P(Y = 1|\tilde{X} = x, \ddot{X} = \ddot{x})$ denotes the probability of not defaulting on a loan. Then, the lender needs to make a decision such that expected profit is maximized. That is,

$$\underset{d\in\{A,R\}}{max}\ E\left(\underset{D,Y}{\pi}|\tilde{X}=\tilde{x},\ddot{X}=\ddot{x}\right)=\underset{d\in\{A,R\}}{max}\left\{\alpha\,\underset{A,1}{\pi}-(1-\alpha)\underset{A,-1}{\pi}\right\} \tag{1.7}$$

Therefore, a lender will accept a borrower's loan application if there is an expected economic gain. This is stated as

$$\alpha\,\underset{A,1}{\pi}-(1-\alpha)\underset{A,-1}{\pi}>0 \tag{1.8}$$

Here, α, known as the cutoff value, serves the purpose of regulating who the lender will accept or deny. Solving for it by equating the expected gains with the expected losses gives the following equation for the critical/threshold cutoff value

$$\alpha=\frac{\pi_{A,-1}}{\pi_{A,-1}+\pi_{A,1}}\overset{\text{def}}{=}c(\ddot{x})\in(0,1) \tag{1.9}$$

Simplifying that gives Equation 1.10

$$\alpha = \left(1 + \frac{\pi_{A,1}}{\pi_{A,-1}}\right)^{-1} \stackrel{\text{def}}{=} c(\ddot{x}) \in (0,1) \tag{1.10}$$

Traditionally, the cutoff value in the industry is determined by past business experiences and it is often expressed as a constant number. However, as a direct consequence of using Equation 1.3 as the profit function, Equation 1.9 and Equation 1.10 presents a variable cutoff that is a function of the loan characteristics. As a result, two important observations emerge

1. The cutoff function is directly tied to the lender's profitability objective. Therefore, the lenders control whom to extend credit based on their profit function.
2. The cutoff function uses borrower loan characteristics. This means that borrowers with different loan contracts will also have different cutoff values.

In addition to the benefits highlighted above, Equation 1.9 carries an intuitive economic interpretation: the variable cutoff per applicant is a ratio between potential losses when they default and potential gains when they do not default.

Expected Profit Optimization

Let $\tau(\tilde{X}, \ddot{X}; \theta)$ be some parametric model having $\mathrm{T}(\tilde{X}, \ddot{X}; \theta) \in [0,1]$ as the CDF. Although the exact nature of $\mathrm{T}(\tilde{X}, \ddot{X}; \theta)$ is not known, we will interpret its values to be a conditional probability (conditioned on $X^{\sim}$). Then, the lender's profitability objective can be framed as the following optimization problem:

$$\underbrace{max}_{d \in \{A,R\}} E\left(\pi_{\{D,Y\}} \middle| X^{\sim} = x^{\sim}, \ddot{X} = \ddot{x}\right)$$

$$\text{subject to D} = \begin{matrix} R & \text{if } T(x^{\sim}, x^{\cdot\cdot}; \theta^{*}) \leq c(x^{\cdot\cdot}) \\ A & \text{if } T(x^{\sim}, x^{\cdot\cdot}; \theta^{*}) > c(x^{\cdot\cdot}) \end{matrix} \tag{1.11}$$

We choose $T(x^{\sim}, x^{\cdot\cdot}; \theta)$ to be the logistic distribution as per the industry standard.

Here, the lender seeks to find an optimum parameter θ^{*} (or family of parameters) that solves Equation 1.11. According to Elliott and Lieli (2013), the optimization problem can be written as

$$\underset{\theta^{*} \in \Theta}{max}\, E\left\{b(Y + 1 - 2c(\ddot{x})) * sign(T(\tilde{x}, \ddot{x}; \theta^{*}) - c(\ddot{x}))\right\} \tag{1.12}$$

Where, sign (v) = 1; for v > 0; sign (v) = -1 for v < 0 and Y takes the same form as Equation 1.1 and $b = \pi_{A,-1} + \pi_{A,1}$ corresponds to the denominator of Equation 1.9.

Credit Model with Risk Aversion

To model risk-averse lending using standard utility theory, consider a decision maker i.e., the lender, who will make a choice among risky lotteries, i.e., borrowers, by ranking their expected utilities. Therefore, the objective function of the lender must be generalized to maximizing the expected utility.

The decision maker's attitude towards risk is captured by the property of the underlying utility function. Risk aversion is modeled by a concave Von Neumann-Morgenstern utility function $U(.)$

which is strictly increasing and twice differentiable. We will use the following result from expected utility theory, Myerson and Zombrano (2019)**.**

For a lender with initial wealth W_i, we have the following definitions:

- Let R be a lottery with non-zero mean
- Let CE be the certainty equivalent - the sure amount of wealth that has the same effect on lender's welfare in utility terms as bearing the risk of the lottery (see Figure 1)
- Let E be the sure amount of increase in initial wealth

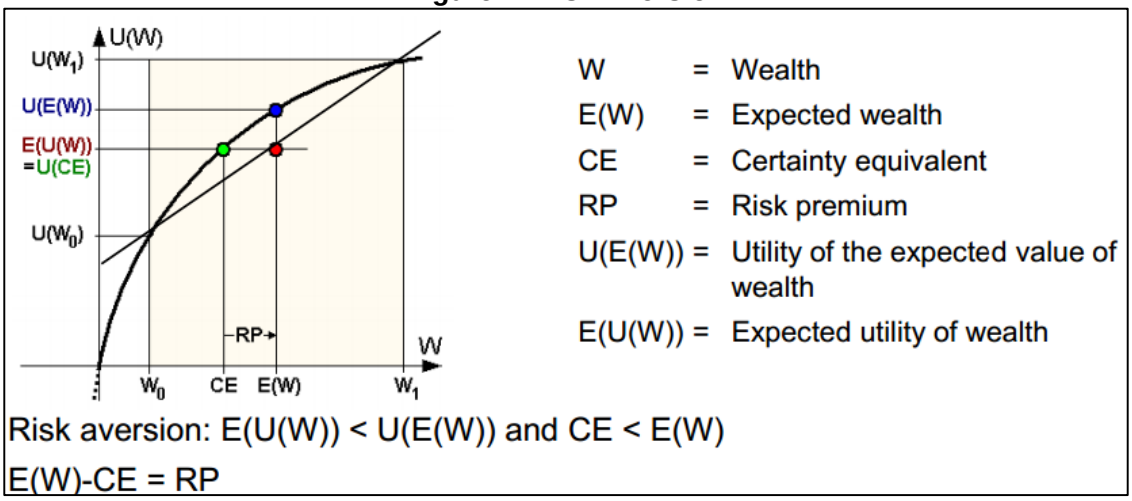
Then, the expected utility of the lender is given as:

$$E(U(W_i + R)) = U(CE) = U(W_i + E) \quad (2.1)$$

That is, the relationship between the initial wealth, W_i, and certainty equivalent is given by $CE = W_i + E$. The implication is that a lender is indifferent between the random payoff of lottery R (associated with a risky borrower) and E, the sure amount of increase in initial wealth. Note that for a given *U*(.), in general, *E* will depend on the initial wealth W_i, except for constant absolute risk aversion (CARA) utility functions, Myerson and Zombrano (2019). In case of risk aversion, Risk Premium *RP* will be positive but its value will, in general, depend on initial wealth level.

The usefulness of this result lies in the fact that an expected utility maximizing lender will choose the lottery **(**associated with risky borrowers**)** which has the highest certainty equivalent. Additionally, the result does not depend on a functional specification of $U(.)$. It is better to think in terms of certainty equivalent instead of the associated utility function because of its convenient property i.e., it is measured in the same units as wealth.

Figure 1: Risk Aversion



We have the following proposition for the risk-averse lender, with its proof given in the appendix.

Proposition 1: The cutoff value for a risk-averse lender will be greater than the cutoff value for a risk-neutral lender.

This makes intuitive sense: a risk-averse lender under utility maximizing constraints will want a higher probability of not defaulting and apply a more stringent criteria than a risk-neutral lender.

Risk-averse case poses difficulties in terms of computational illustration because there are a variety of risk-averse (concave) utility functions to choose from. Not only assumptions have to be made about the utility function, but also about initial wealth levels. That is the reason why, in this paper, computational illustrations have not been provided.

Credit Model with Loss Aversion

Loss aversion refers to the phenomenon, validated by empirical studies, that in general people put more weight on the losses than on the corresponding gains. This was first formalized by Kahneman and Tversky (1979) in their pioneering work on prospect theory.
Prospect theory has been invoked to explain a number of phenomena that otherwise cannot be explained within an expected utility framework. Loss aversion has been one of the tenets of behavioral economics and behavioral finance. Thaler (2005) has considered many applications of prospect theory in finance**.**

Research shows that loss aversion impacts how people make financial decisions. For instance, in their study of home sellers' behavior Genesove and Mayer (2001) conclude that home sellers exhibit strong aversion to selling their homes at a loss. In an experiment by Haigh and List (2005) traders from Chicago Board of Trade showed stronger aversion to losses than non-professional subjects.

Analytically, the modeling of loss aversion is based on the S-shaped value function proposed by Kahneman and Tversky (1979) with the following properties:

1. Value is a function of the changes in the asset or wealth position rather than their final states. This implies a reference level of the assets or wealth in relation to which the deviation in terms of gains and losses are measured.
2. Value function is generally concave in gains and generally convex in losses.
3. It is steeper for losses than for gains.
4. Figure ***2*** illustrates all the three properties of the value function.

Figure 2: Value Function for Loss Aversion

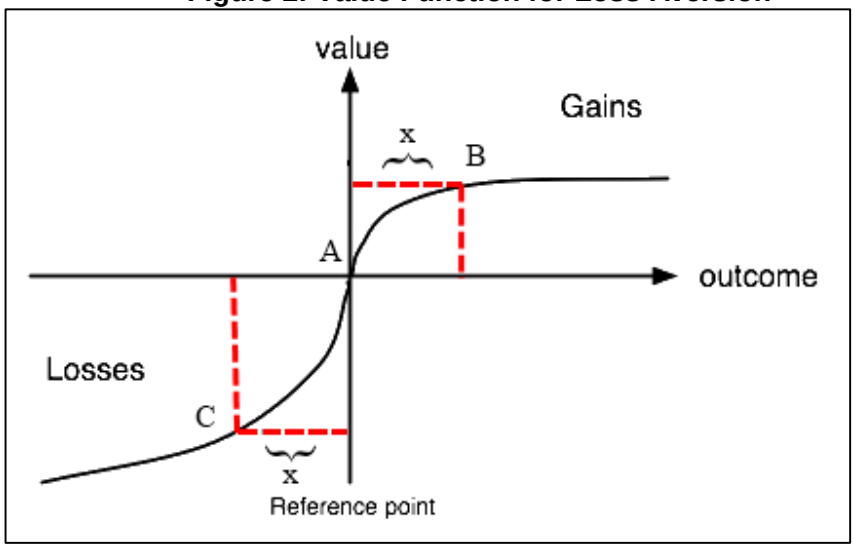


Figure *2* illustrates loss aversion. It shows the graph of perceived value of gains or losses against numerical value of gain or loss. For example, a loss of $x is perceived as a much greater loss in value than a gain of $x. Gains and losses are asymmetric in terms of their perceived value. The value function is kinked at the reference point, i.e. has different slopes at the origin**.**

Tversky and Kahneman (1992) state that the value function of prospect theory has the following functional form

$$V(x) = \begin{cases} x^{\psi} & for\ x \geq 0 \\ -\lambda(-x)^{\beta} & for\ x \leq 0 \end{cases} \quad (3.1)$$

Where $\psi \geq 0, \beta \geq 0\ and\ \lambda \geq 0$

Their empirical work in resulted in the following median values $\psi, \beta = 0.88$ and $\lambda = 2.25$, Tversky and Kahneman (1992).

Using the above specification for *V*(.), the expectation of the value function and subsequent cutoff can now be expressed as follows:

$$E(V(\pi)) = \alpha^{**}(\pi_{+1})^{0.88} - (1-\alpha^{**})(2.25)(\pi_{-1})^{0.88} \geq 0 \quad (3.2)$$

$$\alpha^{**} = \frac{2.25(\pi_{-1})^{0.88}}{(\pi_{+1})^{0.88} + 2.25(\pi_{-1})^{0.88}} \quad (3.3)$$

Similar to Proposition 1, it can be shown that for the particular functional form and the parameter values of the value function given above, α^{**} is greater than α of equation 1.10. This leads to:

Proposition 2: The cutoff value for a loss-averse lender will be greater than the cutoff value for a risk-neutral lender. The proof is given in the Appendix

METHODS

The data used to validate the model is proprietary in nature and was provided exclusively for the purpose of research on the condition of anonymity. It consists of elements of traditional credit data. More specifically, it relates to the applications for subprime automobile loans. There is a total of 23,981 applications and 823 variables or features. As part of the preprocessing stage, the data was stripped of any personally identifiable information such as names, date of birth, social security number, address and telephone numbers. Additionally, all variables whose meaning could not be found in the data dictionary were excluded from the data. The 823 variables were screened and filtered for use in the logistic regression (see Table 1.5 and Appendix).

Loan-specific values and loan-specific information such as the amount, rate, and maturity were assigned to each borrower, in absence of observed data, based on relevant literature and plausibility of the underlying application.

According to (Adams, Einav and Levin, 2009) credit scores and loan sizes are negatively correlated. That is, if the lender determines that the borrower has a low default risk, as measured by a high credit score, then they can be confident to give the borrower a higher loan amount. However, the loan amount should also take into consideration the borrower's income and other debt(s) as well. This is done by introducing the debt-to-income (DTI) ratio.

According to Experian, a DTI ratio should be at or below 40% of the borrower's income for auto loans, Experian (2020). Therefore, the maximum allowable loan amount assigned to an applicant was 40% of their income. Loan amount was assigned based on relative default risk classification, as estimated using the credit score. We constructed this ranking, as shown in Table 1.2, by dividing the credit score into quartiles, with low risk individuals in the first quartile, followed by medium, etc.

Table 1.2: Credit Risk Classification

Credit Quartile	Credit Risk	Loan Amount
First	High	(10%) * Income
Second	Upper medium	(20%) * Income
Third	Lower medium	(30%) * Income
Fourth	Low	(40%) * Income

Auto loans have fixed durations of 36 months, 48 months, 60 months and 72 months. Unlike loan sizes, borrower default risk, as measured by a credit score, and loan maturity are positively correlated (Kuvíková, 2015). That is, the greater a borrower's credit score, the more time they will be assigned to repay the loan because the lender trusts them to hold unto the loan for a longer period.

Like the loan sizes, we create borrower default risk by dividing their credit score into quartiles, with low risk individuals in the first quartile, followed by medium, etc. Loan maturity was assigned according to the default risk of the borrower. For example, borrowers with low default risk have 72 months to repay the loan, etc. This is shown in Table 1.3.

Table 1.3: Default Risk Classification

Credit Quartile	**Income Risk**	**Loan Maturity**
First	High	36 months
Second	Upper medium	48 months
Third	Lower medium	60 months
Fourth	Low	72 months

Following Lieli and White (2010), the loan rate, was calculated as being 10% more than the monthly risk-free rate on the US treasury yield curve, U.S. Treasury Yield Curve (2020).

Following Lieli and White (2010), we speculate that borrowed amount and recovery rate are also negatively correlated. That is, if the amount borrowed is small, the lender can expect to recover a great deal of it in the event of default. Table 1.4 below shows the recovery rate for loan amount $10{,}000\lambda$. This is also reminiscent of Lieli and White (2010) where the recovery rate was capped at 50%.

Table 1.4: Recovery Rates

Loan Amount	Recovery Rate
$\lambda \leq 0.4112$	0.95
$0.4112 < \lambda \leq 0.8223$	0.90
$0.8223 < \lambda \leq 1.2335$	0.85
$1.2335 < \lambda \leq 1.6446$	0.80
$1.6446 < \lambda \leq 2.0558$	0.75
$2.0558 < \lambda \leq 2.4670$	0.70
$2.4670 < \lambda \leq 2.8781$	0.65
$2.8781 < \lambda \leq 3.2893$	0.60
$3.2893 < \lambda \leq 3.7004$	0.55
otherwise	0.50

RESULTS

The results of the following models were computed and compared for: a) the risk-neutral lender (Table 1.6); and b) the loss -averse lender (Table 1.7)

Model I - a logistic regression model with a constant cutoff value.
The logistic regression model (Table 1.5) was fitted to the data using maximum likelihood estimation. The characteristics of the applicant and the logistic regression are used to derive the default probability. This default probability is compared against the *constant* cut off value to make the accept/reject decision.

Model II - a logistic regression model with a variable cutoff.
The logistic regression model (Table 1.5) was fitted to the data using maximum likelihood estimation. The characteristics of the applicant and the logistic regression are used to derive the default probability. This default probability is compared against the *variable* cut off value (Equation 1.10) to make the accept/reject decision

Model III - an expected profit-based optimized logistic regression model.
The logistic regression model (Table 1.5) was fitted to the data using the optimization model. It should be noted that the parameters of this logistic regression model are *not* estimated via the method of maximum likelihood. The characteristics of the applicant and the *optimized* logistic regression are used to derive the default probability. This default probability is compared against the *variable* cut off value (Equation 1.10) to make the accept/reject decision. Because of the non-smooth nature of Equations 1.11 or 1.12, the Simulated Annealing algorithm was used to solve the optimization problem.

The results of the logistic regression model are presented in Tables 1.5. The variables used in the logistic regression are described in the Appendix.

Table 1.5: Logistic Regression Results

Variables	Estimates	Std. Error	Pr > Chi.Sq
Intercept	-1.199	0.045	<.0001
UT01	0.154	0.028	<0.0001
IV06	0.058	0.015	<0.001
NI01	-0.074	0.007	<0.0001
AT01	0.018	0.016	0.2656
PA06	0.036	0.011	0.0014
VI01	0.019	0.003	<0.0001
CLA01	-0.006	0.001	<0.0001
AO01	-0.002	0.000	<0.0001
AP01	-0.000	0.000	<0.0001
NBK01	-0.000	0.000	0.0017
CN01	0.041	0.006	<0.0001

Next, we present the results of the profit-based approach outlined earlier for the risk-neutral (Table 1.6) and loss-averse (Table 1.7) cases. For each model, we present the results for the following metrics:

1. Accept Ratio (AR) - Proportion of applications that were accepted
2. Reject Ratio (RR) - Proportion of applications that were rejected
3. P(G|A) - Proportion of good borrowers that were accepted
4. P(B|R) - Proportion of bad borrowers that were rejected

5. Profit – Optimal expected profit per application (based on all the 23,981applications)
6. Total Expected Profit - It is the total expected profit that can actually be realized. It equals optimal expected profit per application in 5 above x acceptance ratio (AR) x the total number of applications. The total number of applications (23,981) is the same for all the three cases and hence, omitted from the calculations in the last column of Tables 1.6 and 1.7 without affecting the comparison of the two cases with respect to this metric.

As in Lieli and White (2010), because we have no way of knowing the lender's explicit cutoff value, for comparison purposes, the constant cutoff value was chosen such that the acceptance ratio under Model I and Model II will be roughly similar.

Table 1.6: Risk-neutral Lender

Model	Cutoff	AR	RR	P(G|A)	P(B|R)	Expected Profit	Total Expected Profit
Logistic Reg	Constant	0.925	0.075	0.229	0.891	$181.48	$167.87
Logistic Reg	Variable	0.939	0.060	0.231	0.889	$187.73	$176.28
Expected profit-based	Variable	0.836	0.194	0.243	0.884	$213.01	$178.08

Table 1.7: Loss-averse Lender

Model	Cutoff	AR	RR	P(G|A)	P(B|R)	Expected Profit	Total Expected Profit
Logistic Reg	Constant	0.141	0.859	0.327	0.788	$36.19	$5.10
Logistic Reg	Variable	0.087	0.913	0.309	0.792	$42.39	$3.69
Value-based	Variable	0.972	0.028	0.217	0.593	$134.22	$130.46

CONCLUSIONS

1. The profit-based approach (using NPV) for risk-neutral lending yields better results than the loss -averse scenario. Loss aversion lead to sacrificing some of the total expected profits that can be realized.
2. The question of whether this is generalizable to other profit functions (e.g., ROI, CLV) prompts further research.
3. In the loss aversion scenario, the expectation of the value function is derived by using probabilities of defaulting and not defaulting. Future research will consider the weighing function of the probabilities as suggested by Prospect Theory.

APPENDIX

Variable	Meaning
VI01	Number of vehicle inquiries within 24 months
NI01	Number of inquiries within 30 days
UT01	Number of utility (e.g., electricity, water) inquiries within 12 months
AT01	Number of financial accounts opened within 12 months
CN01	Number of non-medical accounts in collections
AP01	Total payments
PA06	Number of financial accounts rated 60 days past due
IV06	Number of vehicle inquiries within 6 months
AO01	Months since oldest of financial accounts opened
CLA01	Months since most recent collection assigned
NBK01	Non-bank revolving balance (e.g., credit card) with limit

Proof of Proposition 1

$$E[U(W_i + R)] = U(CE)$$

Where

$$R = \begin{cases} \pi_{+1} \, with \, probability \, \alpha^* \\ \pi_{-1} \, with \, probability \, 1 - \alpha^* \end{cases}$$

W_i is the initial wealth

$$\alpha^* U(W_i + \pi_{+1}) + (1 - \alpha^*) U(w_i - \pi_{-1}) \\ = \alpha^* U(W_i + E) + (1 - \alpha^*)(W_i + E)$$

Rearranging the terms gives the following:

$$\alpha^*[U(W_i + \pi_{+1}) - U(W_i + E)] + (1 - \alpha^*)[U(W_i - \pi_{-1}) - U(W_i + E)] = 0$$

Considering the first term in Taylor's series approximation, we have the following:

$$\alpha^* U'(W_i + E)[W_i + \pi_+ - W_i - E] + (1 - \alpha^*) U'(W_i + E)[W_i - \pi_{-1} - W_i - E] = 0$$

Or

$$U'(W_i + E)[\alpha^*(\pi_{+1} - E) + (1 - \alpha^*)(-\pi_{-1} - E)] = 0$$

However,

$$U'(W_i + E) > 0$$

Which implies,

$$\alpha^*(\pi_{+1} - E) + (1 - \alpha^*)(-\pi_{-1} - E) = 0$$
$$\alpha^*(\pi_{+1} + \pi_{-1}) = \pi_{-1} + E$$

Now, let α^* denote the cutoff value for the risk-averse lender. Then, like Equation 1.9, we can express the cutoff probability as

$$\alpha^* = \frac{\pi_- + E}{\pi_{+1} + \pi_1}$$

Note that *E* = 0 implies no lending because the welfare (utility) of the lender is not increased by lending and *E* = π_{+1} implies there is no risk in lending! Hence, in practice, we have $0 < E < \pi_{+1}$ Now, from Equation 1.9 we have

$$\alpha = \frac{\pi_-}{\pi_{+1} + \pi_1}$$

Therefore, for the risk-averse lender it is clear that $\alpha^* > \alpha$. Thus, a risk-averse lender raises the threshold for the probability of not defaulting on the loan.

Proof of Proposition 2

From Equation 1.10, it is shown that

$$\alpha = \left(1 + \frac{\pi_+}{\pi_{-1}}\right)^{-1}$$

Similarly, $\alpha^{**} = \left(1 + \frac{\pi_+^{0.88}}{2.25\pi_-^{0.88}}\right)^{-1}$

$$\alpha^{**} = \left(1 + \frac{1}{2.25}\left(\frac{\pi_+}{\pi_-}\right)^{0.88}\right)^{-1}$$

Because $\pi_+ > \pi_-$, it implies that $\left(\frac{\pi_+}{\pi_-}\right)^{0.88} < \left(\frac{\pi_+}{\pi_-}\right)$

Therefore, $\left(\frac{1}{2.25}\right)\left(\frac{\pi_+}{\pi_-}\right)^{0.88} < \left(\frac{\pi_+}{\pi_-}\right)$

Comparing the above expressions for α and α^{**}, leads to $\alpha^{**} > \alpha$.

REFERENCES

Baesens, Bart; Roesch, Daniel; and Scheule, Harald. Credit Risk Analytics: Measurement Techniques, Applications, and Example in SAS. 2016. ISBN: 9781119278344.

Barrios, Luis Javier Sanchez; Andreeva, Galina; and Ansell, Jake. "Monetary and relative scorecards to assess profits in consumer revolving credit." In: *Journal of the Operational Research Society*. 2014. ISBN: DOI:10.1057/jors.2013.66. DOI: 10.1057/jors.2013.66.

Cyert, Richard M; Davidson, H Justin; and Thompson, Gerald L. "Estimation of the allowance for doubtful accounts by Markov Chains." In: *Management Science* (1962). ISSN: 0025-1909. DOI: 10.1287/ mnsc.8.3.287.

Elliott, Graham and Lieli, Robert P. "Predicting binary outcomes." In: *Journal of Econometrics* (2013). ISSN: 03044076. DOI: 10.1016/j.jeconom.2013.01.003.

Experian (2020) Retrieved on September 26, 2020

Finlay, S. M. "Towards profitability: A utility approach to the credit scoring problem." In*: Journal of the Operational Research Society* (2008). ISSN: 01605682. DOI: 10.1057/palgrave.jors.2602394.

Finlay, Steven. "Credit scoring for profitability objectives." In*: European Journal of Operational Research* (2010). ISSN: 03772217. DOI: 10.1016/j.ejor.2009. 05.025.

Genesove, D. and Mayer, C. (2001). "Loss Aversion And Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics* 116 (4), 1233-1260.

Haigh, Michael S. and List, John A. (2005). "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *The Journal of Finance* 60 (1), 523-534

Hand, David J. and Kelly, Mark G. "Super scorecards". In: *IMA Journal of Management Mathematics* (2002). ISSN: 1471678X. DOI: 10.1093/imaman/ 13.4.273

Kahneman, Daniel and Tversky, Amos. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* Vol 47 (2), pp. 263-292.

R. L. Keeney and R. M. Oliver. "Designing win-win financial loan products for consumers and businesses." In: *Journal of the Operational Research Society* (2005). ISSN: 01605682. DOI: 10.1057/palgrave.jors.2601992.

Khandani, Amir E.; Kim, Adlar J.; and Lo, Andrew W. "Consumer credit-risk models via machine-learning algorithms." In: *Journal of Banking and Finance* (2010). ISSN: 03784266. DOI: 10.1016/j.jbankfin.2010.06.001.

Lessmann, Stefan et al. "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research." In: *European Journal of Operational Research* (2015). ISSN: 03772217. DOI: 10.1016/j.ejor.2015.05.030.

Lieli, Robert P. and White, Halbert. “The construction of empirical credit scoring rules based on maximization principles.” In: *Journal of Econometrics*. 2010. ISBN: 0304-4076. DOI: 10.1016/j.jeconom.2009.10.028.

Maldonado, Sebastian et al. “Integrated framework for profit-based feature selection and SVM classification in credit scoring.” 2017. DOI: 10.1016/j.dss.2017. 10.007

Serrano-Cinca, Carlos and Gutierrez-Nieto, Begona. “The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending.” in: *Decision Support Systems* (2016). ISSN: 01679236. DOI: 10.1016/j.dss.2016. 06.014.

Thomas, Lyn C. “A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers.” In: *International Journal of Forecasting* (2000). ISSN: 01692070. DOI: 10.1016/S0169-2070(00)00034-0.

Thomas, Lyn C. Consumer credit models: Pricing, profit and portfolios. 2009, p. 6. ISBN: 9780191715914. DOI: 0.1093/acprof:oso/9780199232130.001.1.

Myerson, Roger B. and Zambrano, Eduardo. Probability Models for Economic Decisions, Chapter 3, Second Edition, 2019 (MIT Press, Cambridge).

Thaler, R., Advances in Behavioral Finance, Vol. II 2005, (Princeton University Press, Princeton).

Tversky, Amos and Kahneman, Daniel, “Advances in prospect theory: Cumulative representation of uncertainty,” *Journal of Risk and Uncertainty*, 5, 297-323. 1992,

U.S. Treasury Yield Curve (2020) Retrieved on September 26, 2020

DECISION SCIENCES INSTITUTE

A framework for Effective Communication and Delivering Online Courses during the COVID-19 Pandemic

Khaled A. Alshare
Qatar University
kalshare@qu.edu.qa

ABSTRACT

The COVID-19 pandemic has enforced higher education institutions across the globe to move from traditional face-to-face learning to online education. It has been proven that higher education institutions that were well prepared with respect to digital transformation in general, and e-learning in particular, had minimal difficulty in adopting online learning during the COVID-19 pandemic. Communication and clear expectations are extremely important in the online education, especially during a stressful time like the COVID-19 pandemic period. This paper will provide instructors with some guidelines for effective delivering online courses. The proposed framework for effective delivering online courses addresses four major dimensions including effective communication, class management tactics, course design and content, and alternative assessments for online exams.

KEYWORDS: E-Learning, Effective communication, Student engagement, Class management, Online course design, COVID-19

INTRODUCTION

Distance learning has been with us for a long time in different forms. It started with the correspondence courses in the eighteenth century. With advancement of IT, e-learning becomes more popular and attractive to many students and institutions. However, many faculty in different institutions are still not ready and they do not feel comfortable in delivering online courses for many reasons. For example, instructors believe that online courses require more effort and they believe that online courses is not effective comparing to the traditional face-to-face classrooms (Kwun, Alshare, Grandon, 2005). Moreover, instructors have concern regarding the assessment integrity of online courses. Therefore, there is a need to provide faculty with the proper training programs in order to be ready to deliver online courses (Oyarzun et al. 2020). The COVID-19 pandemic left no choices for universities, but to go online to continue the spring 2020. It has been reported that there were 10.3 million enrollments in online courses on Coursera in April and May of 2020, which represents an increase of 644% from the same period last year (DeVaney, et al., 2020).Thus, both faculty and students have to espouse e-learning in a short time. Despite the effort by the universities in providing on demand training programs for both faculty and students; yet, not all of them were ready and comfortable in adopting such learning/teaching environment (DeVaney et al., 2020).

BACKGROUND PERSPECTIVES

Distance Education Challenges

E-learning environment consists of integrated components. These components include course content and design, information technology (learning management system platforms), students and faculty readiness for such environment. In addition to these primary components, there are supporting components like technical and counseling supports. It has been reported that these factors are necessarily for smoothing the e-learning process (Sadeghi, 2019; Fleming, Becker, and Newton, 2017; Ngampornchai and Adams, 2016). The supportive services are most needed and they have a strong impact during a stressful time like the situation of the COVID-19 Pandemic. There are certain challenges facing students and faculty in the e-learning environment. For example, students' feeling of isolation, self-discipline, motivation, effort requirement, IT literacy, self-efficacy, adaptability effort, and time management skills (Dixson, 2015; Driscoll et al. 2012; Huei-Chuan et al. 2020; Hung et al., 2010; Ng, 2019). On the other hand, instructors face other types of challenges such as quality and integrity of student's assessment, technology adoption (e-readiness), and student engagement (Gay, 2016). Stenger (2020) stated "Educators who don't pay special care to ensure that all of their students—no matter where they're located—are equally involved in the classroom run the risk of students becoming disengaged and disappointed." Martin and Bolliger (2018) found that among the three modes of interaction (engagement) student-to-student, student-to-instructor, and student-to-content, student-to-instructor engagement was the most appreciated by the students. Kuwn et al (2005) reported that instructors in both US and South Korea felt that in online learning, cheating would be easier compared to face-to-face traditional learning environment. In the same study, the authors reported that students in both countries felt that online courses, compared to the traditional face-to-face education, would require more effort. It should be noted that creating a positive attitude will help in overcoming these challenges. For example, Pappas (2015) discussed many benefits of creating positive attitudes for students such as increasing students' motivation, helping them in remembering and comprehending the course topics, emotional engagement, and achieving their goals.

E-learning success depends mainly on both students and faculty. In addition, a strong supportive environment is an important factor too. However, the student remains the corner stone for the online education success. Other factors contribute to the success of the e-learning include the quality of the content, course design (using modules), self-motivation and management (Oyarzun et al. 2020). Other factors like ease of use of online learning systems, confidence and comfort in using the online learning environment, the quality of information provided by instructors in form of feedback, instruction, and interaction (Bichsel, 2013). Along with this line, instructors should avoid the common mistakes as reported by Opie (2020). These mistakes include:

- Not having a lesson plan (you should have a teaching plan and practice it before you go live).
- Not having a dynamic class (have some course material posted in advance, videos, current materials)
- Not having breaks (have a break every 20 minutes for 5 minutes)
- Not seeking feedback from students during the semester (ask for feedback maybe every week especially from your excellent students).
- Not checking on your students readiness (make sure that they have what they need to complete successfully the course)
- Not considering students' emotions

CONCEPTUAL FRAMEWORK

The following paragraphs describe a framework for effective delivering online courses. The unique feature about the proposed framework is that it is based on best practices reported by instructors from many countries. The proposed framework will help instructors in managing effectively their online courses. The framework consists of four main components as shown in Table 1. The first component is related to effective communication means. The second component addresses class management techniques. The third component describes the course design and content. The last factor provides alternative assessments to online exams.

Table 1: A framework For Effective Delivering Online Courses*

Effective Communication - Section A
Establish ground rules up front to provide effective e-learning environment (see Appendix 1)
Delineate clear expectations
Summarize the learning objectives
Clarify and answer any remaining questions
Announce the next topic and any upcoming assignments or examinations
Send more emails and announcements
Personalize in a professional manner your students' inquires
Provide students with tips on how they can be self-disciplined and manage their times effectively
Make yourself available for student support on a regular basis (online office hours)
Class Management Tactics – Section B
Check your Attitude before you go online (creating a positive attitude)
Ensure student engagement and if it is allowed use video and keep their cameras on throughout the class
Place the students into virtual breakout groups to enhance their participation
Give five minutes for socializing in breakout sessions
Provide FAQs document (See appendix 2)
Encourage students participation since it is the most important part of a class session
Establish e-learning participation etiquette • Be present and eliminate distractions • Open the mic only when you are speaking • Use chat and virtual hand-raising when you want to speak • Act and behave like if you are in the classroom
Course Design and Content – Section C
Convert traditional F2F courses into online environment by deciding on: • What the topics/learning activities to discuss during the live sessions (group work, perspectives, peer learning, … • What topics to leave for students to read on their own (having an ideas about concepts before the class, students need time to reflect, employed flipped class strategy) • What topics you may eliminate
Consider using varied methods of delivering instruction such as slides, videos, polls, lectures, reflection activities, among others
Develop modules for the course content and prepare activities for each module
Share the agenda with students for each session
Create a detailed lesson plan for each session
Create course calendar by scheduling all class activities
Run breakout exercises in which each group develops written answers and formal presentations (practice virtual collaboration)
Divide the lecture into short intervals (15- to 20-minute each) followed by breaks to keep students engaged (e.g., ask questions or taking a quick poll)
Take 5-10-minute break every hour
Use flipped classroom model if you feel it is appropriate for your course (each video should be less than 15 minutes)

Provide students with the chance to reflect on the class topics during and after the live session by: • Asking polls/questions/quizzes. This would give you an idea if the students indeed are learning • Asking what is the topic that you are still confused about it?. The answers will help you in modifying your next class agenda. • Asking what is the most interested topic you learned today? The answers will provide you with information on how students are connected to particular topics.
Have plan A and plan B for your lessons (do not trust technology)
Alternative Assessments for Online Exams – Section D
Game, puzzle, simulation, or contest among teams (best design, best story, best solution)
Writing a report on arguments for/against (debate) certain topics
Sharing learning experience and reflecting on "what went right and what went wrong"
Having a set of problems (could be cases) and randomly assign a problem to each student or group and ask them to solve it based on what they have learned in the course
Designing posters
Providing different scenarios and performing "what if" analysis
Asking each student or group to explain a concept that was not covered in the class by preparing a presentation (it could be written (slides) or live presentations)
Interpreting results/outputs of a solution of a problem (e.g., asking so what?)
Asking students to find mistakes (errors) for a case study (for example having a diagram, which has some errors and ask students to find the errors and fix them).
Asking students to reflect on their experience during completing a project.
Preparing a literature review
Creating an executive summary or writing a short abstract about a topic
Creating a conceptual diagram
Applying course concepts to a case study (assigning a case study and then ask students what concepts from the course they can apply to analyze it)
Be flexible, for example, provide 5 questions and ask students to answer three and one of them could be mandatory

*. Main Source: (Harvard Business Publishing Education- www.hbsp.harvard.edu).

The first step in developing and creating a successful online course is to establish a good communication plan. For example, setting rules for the methods of communication (e.g., emails or through learning management system (e.g., Blackboard). The time and frequency that students should expect a response from their instructors. Appendix 1 provides an example of such rules. Instructors may have online office hours; this could be either by email or by being available through live sessions. Employing variety of methods of communication. For example, send messages by email and posting same messages on the announcement section on e-learning management system. Establishing clear expectations was reported to be the most helpful strategy for students and faculty in online education (Oyarzun et al. 2020).

An essential part of an effective communication plan between the instructor and the students is establishing class management rules. These rules should cover most of the questions that students might have in their mind. The best way to address this issue is to create a FAQs document that has the answers for the most questions that student might have about the course. For examples, "Will there be scheduled class meetings?", "Will there be scheduled office hours?", "Where do I find the course materials on Blackboard?" These are examples of such important questions that their answers will make students have a peace of mind. For a list of FAQs, see Appendix 2. Equally important is having a positive attitude by both faculty and students. Pappas (2015) suggested many tips for creating a positive attitude in the e-learning environment. These tips include empowering students (learner centered approach), explaining the e-learning course value to students, reducing negative attitude and fear and stress (learning from mistakes, asking good questions, utilizing online discussion board), providing positive feedback, and considering students' emotions (engagement and interest).

The third component in the framework addresses the online course structure, which includes the design and the content of the course. It should be noted that teaching online courses is not just converting the content of the course in the traditional face-to-face classroom, it requires to redesign the course so that it serves the e-learning environment better. According to Levy, (2020) when you design the online course, you should decide on what topics (content) that you will deliver during synchronous mode; the topics that you will leave them to students to read on their time (asynchronous mode); and the topics that you may omit. The main theme in teaching online courses is to keep students engaged in the class activities (Tai et al. 2019). It has been reported that student engagement in online courses improves student satisfaction and performance, and decreases the feeling of isolation (Martin and Bolliger, 2018; Dixson, 2015). To achieve this goal, instructors need to divide the lesson into small chunks (15-20 Minutes) and create for each portion activities to keep students' attention and involvement. These small chunks should be followed by breakout sessions so that students groups discuss among themselves the topics at hand and then report to the entire class their reflections. Additionally, students should be informed in advance about the course schedule and content by creating course calendar which layouts all activities for the course. Moreover, the lesson agenda should be communicated in every session. To increase students' participation and engagement, polls and questions could be used during each chunk.

It is very important as an instructor to decide on what is your main goal from students' assessment. Is it to test their knowledge or applying their knowledge or both? Additionally, the focus of any assessment should be more on critical thinking and problem solving skills, especially in the e-learning environment. Moreover, the assessment requirement and instruction must be clear to students. One way to limit the chance of cheating during the online assessment is to limit the time for the assessment (e.g., 2 hours). As shown in Table 1-D section, there are many alternative assessments to the traditional exams that could be used in the e-learning environment. For example, instructors based on their course type, may employ games, puzzles, or contests among teams (e.g., providing best design, best story, and best solution). Another assessment could be writing a report on arguments for/against (debate) on topic. It should be noted that the course subject and design should determine the assessment types and methods (diagnostic assessment (pretest and students surveys), formative assessment (quizzes, assignments), and summative assessment (course project, final exam) (Orlando, 2011). Instructors should employ different assessment methods with clear evaluation criteria considering the technology requirement, student-learning styles, and learning outcomes (Mimirinis, 2019; Gaytan, 2004).

CONCLUSION

There is no one-size-fits-all. Each course has its own features and each learner has different needs, and energy for participation in the online classes. Therefore, understanding students' needs and capabilities will make the transition from the traditional learning environment to online learning settings a very positive experience. For example, checking on your students' expectations (some are aiming for A and B+ but others aiming on just passing the course) and their perceptions about the course will help in setting the stage for the class and enable instructors to create constructive learning environment. Instructors, with little effort, can help students to be active participants in the online learning environment by encouraging them to ask questions, submit assignments before due dates, prepare for the class discussion, and create good image about themselves. Educators should be aware of the challenges during a stressful time and remind their students with all support services available to them. Embracing a positive mindset and being a supportive instructor will create the sense of community for students, which will help in releasing the stress.

APPENDIX

Appendix 1: Rules and Policies

Rules and Policies
Class XXXX

1. *You should read carefully the syllabus.*
2. *Be on time in attending the live sessions. According to XX University policy, missing more than 25% of the class meetings will results in an F grade in the course.*
3. *Be prepared to answer the pre-assigned readings.*
4. *If no answer or response received from the student, then the student has to come up with a statement of wisdom or a joke!*
5. *Class Joke Contest—every month we will have a joke contest in which students submit a jock related to course topics. The winners will get extra points.*
6. *You are encouraged to participate in the discussion in the class and to post extra materials related to the course topics on the discussion board.*
7. *No make up for missed quizzes.*
8. *In order to see how you are doing in the class, at least every 2 weeks or so, send me an e-mail.*
9. *Assignments will be graded and return back to the students in one week, otherwise, a full grade will be given to every un-graded assignment. This is not applicable to the exams.*
10. *No Late Assignments will be accepted.*
11. *You should observe all rules and policies that are listed in the course syllabus and any rules the might be announced in the class.*

I wish you an enjoyable and productive semester.

Appendix 2: Frequently Asked Questions for XXX Class

FAQ: Will there be scheduled class meetings?
Answer: Yes, we will meet as scheduled in the courses schedule posted by the university. We will have live sessions every Sunday, Tuesday, and Thursday from 10:00 – 11:50 AM. Blackboard Collaborate Ultra will be used for all live sessions.

FAQ: Do I need to attend the live sessions for the class?
Answer: Yes, the attendance will be taken in every live session through Blackboard attendance feature.

FAQ: Will there be scheduled office hours?
Answer: Yes, every Wednesday from 10 to 11 AM

FAQ: Where do I find the course materials on Blackboard (Bb)?
Answer: All course material will be posted on Blackboard. The three main sections that will be used on Bb are:

- Course Content: all chapters related documents (PowerPoint files, handouts, tutorials,…)
- Assignment Section: All assignment will be listed in the assignment section.
- Discussion Board: Two forums will be available on Bb:
  - Questions Forum: you may post questions related to the course.
  - Discussion Forum: This forum gives you the chance to post any recent material related to topics discussed in the class. The Discussion Forum will allow you to initiate some "conversations" and contribute to the class with your own personal/professional experience. This interaction is a great opportunity to learn from others. Participation in the Discussion Board (if there is any) will count toward your final grade as a bonus (maximum 2 points). Therefore, I may post few questions for each chapter for discussion during the week for that chapter.

FAQ: Do we work at our own pace?
Answer: No.

- There is a schedule for each course activity. The online courses fall within the structure of semester-based course offerings. In order to complete the course in a timely manner, it is best that the work be paced over the entire time available (8 weeks).
- To that end, you will find that all assignments and exams have a very specific due date and time. These times will be rigidly enforced. Assignments received after this time will not be graded. Of course, you can always "work ahead." Doing this can give

you some of the freedom of "working at your own pace." Even though you will have sufficient time to do them, I encourage you to start as early as you can to meet the deadlines.

FAQ: I needed information to complete the assignment. Will it be counted as late?
Answer: Yes. It will be considered late.

- I will make every effort to respond to your inquiries according to the schedule that I have laid out. But, the fact that you have e-mailed, and are waiting on a response, is NOT a proper reason for waiting to turn in your assignment.

FAQ: Does it matter what web browser I use?
Answer: Yes! It does!

- Blackboard is quite sensitive to differences in the operating characteristics of the different web browsers, according to the ITS department, you should use (Recommended) browser is Chrome. You can generally tell the version of your browser by clicking on "About" under "Help" on the Browser menu.

FAQ: Are there any special instructions for Internet Service Provider (ISP) users?
Answer: No! There is no special instruction for ISP all the same when it comes to interfere with Blackboard.

FAQ: Are there any special e-mail rules?
Answer: Yes, there are!

- All e-mail related to this course MUST have "MIST-201" at the start of the subject line. Any e-mail that does not conform to this rule will not be answered. Additionally, you must use your QU email account.
- Students should NOT e-mail partially (or fully) completed assignments for me to "look over" prior to submitting them for grading (You are asking me to grade it twice.) On occasion, I may request that you do so, but only then would it be appropriate. However, you are encouraged to ask questions for clarifications about the assignments.

FAQ: E-mail Schedule? Is there an e-mail schedule for the course? When can I expect a response to my question?
Answer: Because of the volume of e-mail that may be generated by an on-line course, I may have to take special measures:

- E-mail that arrives in volume is best dealt with in volume. Many students have the same question, so I can handle several inquiries with one, consistent response. In the Discussion Board Section you will find a Forum “Questions Forum” for asking question where you can post your questions so that all students will benefit from the answer. Therefore, use the discussion board (Questions Forum) instead of e-mails to ask questions.
- E-mail, which arrives privately may be dealt with in public. Questions that arrive by e-mail may be transferred to the discussion board section. By doing this, all in the class can see the answer given; even if they did not realize that they had a similar question.
- Some questions are not readily answered. If I do not respond immediately, the answer may require some thought, or research. I have probably read the question, but I have not figured out how to state the answer. Therefore, do not expect me to respond promptly to your inquiries!

FAQ: Any special e-mail settings?
Answer: Yes. The mail that I send from Blackboard may look to your ISP like it is Junk Mail or Bulk Mail. You will have to check the Junk Mail/Bulk Mail settings to allow the course e-mail to come through. (Normally students received emails from Blackboard from do-not-reply@XXX.edu email account, they need to make sure it is not in the junk list.

FAQ: What if I miss an e-mail message?
Answer: When I send out e-mail to the entire class, it may not go through for some students. While I will make an effort to resend, I may eventually have to give up. As a secondary path of information, I will post ALL bulk messages to the announcements page (You can see the recently sent messages there at this time). As you log on each time, you should check for recent messages that you might have missed. If you notice that there are messages that you are not receiving, you should investigate further to clear up any e-mail problems. You may contact the ITS Service Desk helpdesk@XXXXX or by calling XXX-XXXX

FAQ: How do we submit our work?
Answer: All work will be submitted using the special link under each assignment. You click on the Assignment Section, then click on a particular assignment link. After clicking on the link, you may type comments to me in the "Comments" box and attach your files in the link provided and click on Submit.

- Once you competed your assignment and you want to submit it, make sure you click on the SUBMIT Button. The SAVE Button will not send your work. Also, you need to know that once you submitted your work, you can not resubmit it again. Blackboard will not allow you to submit the same assignment twice. In case you did submit your work and you realized that you did not do all questions in the assignment and the due date is still not passed, then you can send me an e-mail, so I can delete your early submission then at this moment you will be able to resubmit your work again (THIS SHOULD BE DONE BEFORE THE DUE DATE IS OVER). (Remember that ONLY the SUBMIT Button will get your work to me).
- Do not forget to type your name and your Student ID in any document you submit.

FAQ: What about Assessment Activities and Grading?
Answer:

- The assigned grades for all assessment activities are list in the syllabus. However, there will be 7 categories of different assessments:
  - 2 Exams 50%
  - Practical Excel and Access Exams 30%
  - E-commerce Project 5%
  - Ethics Case 2.5%
  - Data Analysis project 5%

- VB Exercise 5%
- Project Management 2.5%

- Some of the work will be approached from an objective point of view (either right or wrong). For this work, there are a certain number of points and points will be deducted for identified flaws in the work. On the other hand, some of the work in this class will be graded subjectively (project and case). For such assignment, there will be no right or wrong answers; instead, there will be poor, good, better, or the best answers. This is what I meant by subjectively (relatively) to the best answer. There is no one correct answer; instead, there is an efficient versus inefficient solution, or a feasible versus infeasible solution.
- Grade for assignment will be posted on Blackboard once it has been graded. You should receive a notification from Blackboard.
- I will try my best to grade your work as soon as possible. However, depending on the number of students in the class, I may need more time to grade. You will know whether your assignments are graded if you see a grade in the grade book. Otherwise, you will see an exclamation mark (!) that means that you submitted the assignment but I have not graded it yet.

FAQ: How about Group project Responsibility (Rules)?
Answer: Yes, there are Group Projects policy (Unfair Work Load Allocation on Group Project(S)

- On occasion, students will find themselves working in a group wherein a group member has the tendency to act as a "free loader". This poses a problem for the group as well as the individual member who will one day realize that this attitude won't cut it in the real world. Particularly in this course, where one aspect of project management is working with, overseeing, and motivating the project team members, this problem will not be tolerated. When this becomes a problem, the following steps should be taken:

1. Groups must directly and openly confront instances of such behavior as early as possible. Ideally, once confronted, the behavior will cease and no further steps are required.
2. If early group efforts to confront the problem fail, the team members must let their instructor know that a problem exists. Their instructor will schedule a meeting with all group members to discuss the problem, its causes, and the solution. Hopefully, the behavior will then cease and no further steps are required.
3. If the problem continues even after steps 1 and 2, the instructor reserves the right to appropriately downgrade the grade for the individual member(s) whose behavior has caused the problem.

- Please note that the failure of one or more team members to contribute, will NOT (generally) be considered an appropriate reason for the failure of the group to complete assigned tasks. When one member of the group fails to perform, that work should be picked up by the rest of the group. Peer evaluations will be part of the project grade.

FAQ: How will we receive feedback on our work?
Answer: you feedback will be a challenge, and it will come in a number of forms:

- The ultimate feedback is the grade, which you will see it on Blackboard.
- I will give feedback on some of the assignments (e.g., projects and cases) by writing in the "Comments" box of "Assignments" link.
- In some assignments, it might be possible to return a file with markings that indicate the deficiencies. I will mark up the assignments with any errors, or comments that seem to be appropriate. In the case of Word documents, I will use "track changes," which will make it easy to see what I have written. If I have comments, I will return the work to you as an attachment.
- If the time permits, I will also provide the feedback during the live sessions.

FAQ: What do I need to know/have?
Answer: to complete this course successfully, you need to:

- Have a reliable computer with the following software/or access to them:
  - PowerPoint
  - Excel
  - Access
  - VB
  - MS Project
- Have reliable access to Internet
- Look and behave like you are really/physically in a classroom
- Do your part, and I will do my part
- Do not expect me to keep talking for 90 minutes without any engagement from your side!
- Learn how to interact online with your group members such as (Bb Groups feature), WhatsApp, or Google Docs.

REFERENCES

Bichsel, J. (2013). The state of E-Learning in Higher Education: An Eye toward Growth and Increased Access (Research Report), Louisville, CO: EDUCAUSE Center for Analysis and Research, June 2013, available from http://www.educause.edu/ecar.

DeVaney, J., Shimshon, G., Rascoff, M., and Maggioncalda, J. (2020), Higher Ed Needs a Long-Term Plan for Virtual Learning, retrieved from https://hbsp.harvard.edu/inspiring-minds/higher-ed-needs-a-long-term-plan-for-virtual-learning?itemFindingMethod=Editorial, on May 30.

Dixson, M. (2015). Measuring Student Engagement in the Online Course:

The Online Student Engagement Scale (OSE), *Online Learning*, 19 (4), 1-15.

Driscoll, A., Jicha, K., Hunt, A., Tichavsky, L., and Thompson, G. (2012). Can Online Courses Deliver In-class Results?: A Comparison of Student Performance and Satisfaction in an Online versus a Face-to-face Introductory Sociology Course, *Teaching Sociology*, 40 (4), 312-331.

Fleming, J., Becker, K., and Newton, C. (2017). Factors for successful e-learning: does age matter? *Education and Training* 59(1):76-89.

Gaytan, J. (2004). Effective Assessment Techniques for Online Instruction, *Information Technology, Learning, and Performance Journal*, 23 (1), 25-33.

Gay, G. (2016). An assessment of online instructor e-learning readiness before, during, and after course delivery. *Journal of Computing in Higher Education*. 28(2):199-220.

Huei-Chuan Wei & Chien Chou (2020) Online learning performance and satisfaction: do perceptions and readiness matter? *Distance Education*, 41 (1), 48-69

Hung, M., Chou, C., Chen, C., Own, Z. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computing & Education*, 55, 1080–1090.

Kwun, O., Alshare, K., and Grandon, E. (2005). Instructor and Student Perceptions of the Online Teaching/Learning Environment: A Cross-Cultural Study, *Academy of Educational Leadership Journal*, 9 (3), 105-130.

Levy, D. (2020). The Synchronous vs. Asynchronous Balancing Act When and How Pre-Work Can Make Your Live Sessions Stronger, Harvard Business Publishing Education, Retrieved from https://hbsp.harvard.edu/inspiring-minds/the-synchronous-vs-asynchronous-balancing-act, on September 8, 2020.

Martin, F. and Bolliger, D.U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning* 22(1), 205-222. doi:10.24059/olj.v22i1.1092

Mimirinis, M. (2019). Qualitative differences in academics' conceptions of e-assessment, Journal of Assessment and Evaluation in Higher Education, 44 (2), 233-248.

Ng, C. (2019) Shifting the focus from motivated learners to motivating distributed environments: a review of 40 years of published motivation research in Distance Education, *Distance Education*, 40 (4), 469-496.

Ngampornchai, A., and Adams, J. (2016). Students' acceptance and readiness for E-learning in Northeastern Thailand. *International Journal of Educational Technology in Higher Education,* 13, 34.

Opie, F. (2020). **"**Online Teaching Survival Guide: A 7-Part Audio Series: Prepare to Succeed in the Digital Classroom, retrieved from https://hbsp.harvard.edu/inspiring-minds/online-teaching-survival-guide-a-7-part-audio-series?itemFindingMethod=Editorial on June 12, 2020.

Orlando, J. (2011). How to Effectively Assess Online Learning, A Magna Publications White Paper, retrieved from http://www.suagm.edu/uagmcv/docs/academia/assessment/assess-online-learning.pdf on June 24, 2020.

Oyarzun, B., Martin, F., & Moore, R. (2020). Time management matters: online faculty perceptions of helpfulness of time management strategies, *Distance Education*, 41 (1), 106-127.

Pappas, C. (2015). 11 Tips to Encourage Positive Attitude In eLearning, retrieve from https://elearningindustry.com/11-tips-to-encourage-positive-attitude-in-elearning on June 19, 2020.

Ringel, R., Tarallo, B., and Lauren Green, L. (2020). 5 Steps to Stay Focused When Teaching Online, Try Balancing High Intensity Activity with Periods of Recovery, Retrieved from https://hbsp.harvard.edu/inspiring-minds/5-steps-to-stay-focused-when-teaching-online?itemFindingMethod=Editorial on May 31, 2020.

Sadeghi. M. (2019). A shift from Classroom to Distance Learning: Advantages and Limitations. *International Journal of Research in English Education*, 4 (1), 80-88.

Stenger, S. (2020). In Hybrid Classes, Some Students Are Likely to Feel Left Out
5 Tips to Ensure All Students Feel Included, Retrieve from https://hbsp.harvard.edu/inspiring-minds/in-hybrid-classes-some-students-are-likely-to-feel-left-out#:~:text=%E2%80%9CEducators%20who%20don't%20pay,students%20becoming%20disengaged%20and%20disappointed.%E2%80%9D on June 10, 2020.

Tai, J, Bellingham, R., Josephine Lang, J. and Dawson, P. (2019) Student perspectives of engagement in learning in contemporary and digital contexts, *Higher Education Research & Development*, 38 (5), 1075-1089, DOI: 10.1080/07294360.2019.1598338

DECISION SCIENCES INSTITUTE

Does the Flipped Classroom Mitigate the Impact of COVID-19 (and Other Disruptions)?

William Swart
East Carolina University
Email: swartw@ecu.edu

Kenneth MacLeod
East Carolina University
Email: macleodk@ecu.edu

Enping (Shirley) Mai
East Carolina University
Email: maie@ecu.edu

ABSTRACT

The pandemic-forced migration of face-to-face classes to an on-line setting, thereby protecting the health of our students, impacted them in many, but mostly unmeasured ways. Our schools took steps to lessen those impacts, and the focus is now turning to what can be done to prepare for future disruptions. We examine whether those classes that were taught using "flipped" approach, which requires students to make use of web-based resources to prepare for class, suffered less during the COVID-19 semester than traditionally taught lecture classes. We use data collected from Fall 2019 and Spring 2020 to examine this question.

KEYWORDS: Flipped Classroom, COVID-19

INTRODUCTION

Consider this quick timeline leading up to declaration of the pandemic:

December 2019: Fall semester is ending, and we receive reminders from our department chairs to submit syllabi for the Spring. We ignore the news about a new virus in China, figuring it is half a world away.

January 2020: We begin teaching our Spring courses, both on-line and face-to-face, much as we have done for the last several years. News from China is more alarming, but still has nothing to do with us or our students.

March 2020: As we leave for Spring Break, escalating news from everywhere has us wondering what we will come back to. It turns out we don't come back and move all our courses on-line.

For some of us, that move was easy. We had taught on-line before, even better, we were currently teaching sections of the same course on-line. We had all the materials ready for the students to use. For others, those of us who had never taught on-line and didn't want to, the end of the Spring semester was a nightmare – scrambling to understand technology we didn't want to use and trying to find a way to give our students the same experience they had been having in a face-to-face class.

Our students confronted the same dichotomy. Our on-line students didn't notice much difference. It might be a little harder to get access to some applications, but they had signed up for an on-line course and they were getting one. Our face-to-face students, though, faced the same difficulties as the face-to-face teachers: unfamiliar applications to replace lectures, the lack of structure provided by face-to-face meetings, and increased responsibility to complete work on their own.

Spring 2020 is now behind us, and we are all trying to prepare for whatever the Fall 2020 semester holds. This time, however, we are not in the desperate, time-constrained melee that we faced in the Spring. We have an opportunity to examine what lessons we have learned and plan for their introduction so that we may be better prepared for future disruption.

Beyond the technical aspects of what we learned (applications, learning management systems, collaborative software) are the issues of how well various approaches to teaching worked. We are focusing specifically on the teaching style referred to as "flipped" teaching. According to DeLozier & Rhodes (2017), the basic idea of flipped learning is that class time is used for a learning activity while lecture material is learned outside of the classroom. This is a blended format of teaching, using both web resources and face-to-face interactions. Swart & Wuensch (2016) found that flipping a class increased student satisfaction without negatively affecting learning.

We have been making use of the flipped paradigm for several years, and have, as part of a continual improvement process, collected data from each semester's students. Using Moore's Transactional Distance Theory (1980), we surveyed the students concerning what aspects of the class the students formed barriers to learning and their overall satisfaction with the class and the format. Thus, by happenstance we had data for the fall of 2019 (pre COVID) and spring of 2020 (post COVID), giving us a unique opportunity to measure the extent to which the well-publicized massive disruptions students and faculty experienced during the pandemic impacted student engagement or satisfaction with their learning experience.

LITERATURE REVIEW

Transactional Distance (Moore, 1980, Zhang, 2003) measures barriers to learning on four constructs: Student to Student, Student to Content, Student to Technology, and Student to Teacher. Swart, MacLeod, Paul, Zhang, & Gagulic (2014) used Zhang's Scale of Transactional Distance to gather data on how well the course delivery compared to an ideal course (a disconfirmation study). Each element of the Scale was answered twice, once for what the student would have considered an ideal class and once for the delivery of the actual class. They called this Relative Proximity of Transactional Distance, or RPTD. The original Scale of Transactional Distance has been altered to bring it up to date (Paul, Swart, Zhang, & MacLeod, 2015; Weidlich & Bastiaens, 2018). The validated revision of the Scale, which we dubbed RSTD-20 uses Weidlich and Bastiaens's definition of Student Satisfaction, defined by six elements, with four measures of barriers to learning: transactional distance (TD) between student and student (ΔTDSS), between student and teacher (ΔTDST), between student and content (ΔTDSC) and between student and instructional technology ΔTDSTECH.

Flipped classes grew out of the blending of traditional face-to-face classes with the technological resources that have become available with the growth of the internet (Lage & Platt, 2000). Posting lecture material so it can be studied asynchronously leaves class time for a variety of in-class activities which can foster reflection, questioning, evaluation, and forming connections between

ideas (Hockings, Cooke, Yamashita, McGinty, & Bowl, 2008). The flipped paradigm has moved from the K-12 arena to higher education (Ash, 2012). As noted by Swart (2017), there is no single unique way to flip a class. In this study, we flipped the class as detailed by Swart & MacLeod (2020).

The impact of the COVID-19 virus on higher education is frantically being evaluated. Right from the start, the Chronicle of Higher Education provided a running commentary on the impact, as shown in this partial list of articles:

3/5/20	Coronavirus Hits Campus
3/17/20	Virtual Events: Combating Coronavirus
3/25/20	How Is Covid-19 Changing Prospective Students' Plans? Here's an Early Look
3/30/20	Covid-19: The Crisis That Launched 1,000 Student Surveys
4/7/20	Campus Zero
4/8/20	4 Colleges, 2 Weeks, One Choice: How Covid-19 Scattered U.S. Higher Education
4/13/20	How Should Colleges Prepare for a Post-Pandemic World?
4/21/20	2 Campuses Give Early Answers to Higher Ed's Biggest Question: What Happens This Fall?
4/23/20	Could Coronavirus Antibody Tests Really Help Colleges Reopen in the Fall?
4/29/20	Colleges Are Urged to Reassess Admissions Policies Because of 'Extraordinary Hardships' Covid-19 Poses
4/29/20	What's Behind Colleges' Bullish Statements on Their Fall Plans?
5/15/20	The Hard Truth About the Fall

Inside Higher Ed, a website (https://insidehighered.com/blogs/learning-innovation), presented three themes, posting the third on May 28th, that they found from their studies of the on-going pandemic:

- Equity: students were not affected equally by the shift to online learning
- Place: having students, staff and faculty together is not that important
- Learning: learning has returned to the central focus of higher education

We studied data collected in a face-to-face undergraduate required business analytics class during the fall 2019 (pre-COVID) and again during the spring of 2020 (COVID). The class was conducted as a flipped class and our objective was to determine if there was any change in transactional distance and/or student satisfaction between the pre-COVID semester and the COVID semester. We were also interested in determining if there was a gender difference in how such changes manifested themselves. Our results (Hypotheses 1 - 3) indicated no reason to believe that there was any difference in the flipped class from one semester to the next.

$H1_0$: There is no difference in RPTD & Satisfaction between Fall 2019 and Spring 2020
$H1_a$: There is a difference in RPTD & Satisfaction between Fall 2019 and Spring 2020

$H2_0$: There is no difference in female RPTD & Satisfaction between Fall 2019 and Spring 2020
$H2_a$: There is a difference in female RPTD & Satisfaction between Fall 2019 and Spring 2020

$H3_0$: There is no difference in male RPTD & Satisfaction between Fall 2019 and Spring 2020
$H3_a$: There is a difference in male RPTD & Satisfaction between Fall 2019 and Spring 2020

EXPERIMENT

The revised Scale of Transactional Distance was given to two undergraduate face-to-face classes. The Fall 2019 class had 40 students, 26 female and 14 male. The Spring 2020 class had 29 students, 21 female and 8 male.

Research Hypothesis 1: Student Satisfaction before and after COVID-19

We ran an Independent Samples t-test on the data. The results are shown in Table 1:

	Levine's test Equality of Variance		Independent Samples Test				
Factor	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
ΔTDSC			-1.585	36.699	0.122	-0.1652	0.1043
ΔTDSS			-1.017	40.550	0.315	-0.1100	0.1082
ΔTDST	0.520	0.473	-0.906	67	0.368	-0.1672	0.1845
ΔTDSTECH	0.004	0.950	0.866	67	0.390	0.1595	0.1842
ΔSATISFACTION	0.055	0.815	-0.152	67	0.880	-0.0240	0.1580
Conclusion: Fail to reject H_0							

Table 1: t-Test Results for Undergraduates

In looking at the students as a whole, there was no significant difference between the two semesters. Next, we considered just female students:

Research Hypothesis 2: Female Student Satisfaction before and after COVID-19

	Levine's test Equality of Variance		Independent Samples Test				
Factor	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
ΔTDSC	10.004	0.003	-1.602	25.634	0.121	-0.2063	0.1288
ΔTDSS	11.441	0.001	-1.536	22.994	0.138	-0.2037	0.1326
ΔTDST	6.977	0.011	-1.903	32.939	0.066	-0.3632	0.1909
ΔTDSTECH	0.074	0.786	0.622	45	0.537	0.1398	0.2249
ΔSATISFACTION	1.506	0.226	-0.694	33.858	0.493	-0.1218	0.1756
Conclusion: Fail to reject H_0							

Table 2: t-Test Results for Female Undergraduates

Again, no significant difference was found, so we ran the same analysis for just the male students:

Research Hypothesis 3: Male Student Satisfaction before and after COVID-19

	Levine's test Equality of Variance		Independent Samples Test				
Factor	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
ΔTDSC	1.494	0.236	-0.589	20	0.563	-0.0893	0.1517
ΔTDSS	0.154	0.669	0.457	20	0.653	0.0786	0.1719
ΔTDST	0.697	0.414	0.524	20	0.606	0.2262	0.4313
ΔTDSTECH	0.435	0.517	0.395	20	0.697	0.1250	0.3166
ΔSATISFACTION	0.967	0.337	0.427	20	0.674	0.1488	0.3485
Conclusion: Fail to reject H_0							

Table 3: t-Test Results for Male Undergraduates

Since we found no significant difference from any of the t-tests, we became interested in the following research questions:

RQ1: What, if any, are unique and significant predictors of Satisfaction pre-COVID

RQ2: What, if any, are unique and significant predictors of Satisfaction COVID

We wanted to determine what factors were unique and significant predictors of satisfaction in the pre-COVID period (Fall 2019). To that end, we performed a stepwise multiple regression using ΔSatisfaction as the dependent variable and the TD's as the independent variables with Gender as an indicator variable. We performed the same calculation on data collected during the COVID period (Spring 2020). We used the data collected with the RSTD-20 instrument. The results of both runs, exhibited in Table 4, contain the same variables as unique and significant

2019 Data Only	
	R^2 =0.778
(Constant)	-0.067
ΔTDST	0.609
ΔTDSTECH	0.382

2020 Data Only	
	R^2 = 0.76
(Constant)	-0.016
ΔTDSTECH	0.454
ΔTDST	0.434

Table 4: Stepwise Multiple Regression Results for the Business Analytics Class

predictors of student satisfaction. However, while ΔTDST was the most important predictor prior to the COVID, ΔTDSTECH became the most important predictor during the COVID period. Thus, it appears as though the physical absence of the instructor shifted the determinant of satisfaction to the instructional technology.

To contrast our findings with those of students not learning in a flipped classroom, we studied the data from a required face-to-face marketing class. Unfortunately, as is the case for most research to determine the impact of COVID, there was no prior to COVID data (e.g. for Fall 2019). Instead, we administered a questionnaire to assess student perception of their class engagement before and after the onset of COVID (pre-March 2020 and post-March 2020). Following Francescucci & Foster (2013), multiple questions based on the National Survey on Student Engagement (NSSE) were used to measure student engagement. In this study, student engagement refers to student perception of their engagement on the following items:

- attending class,
- participating in class,

- interest in courses,
- paying attention in class,
- staying up to date on academic workload, and
- instructor interaction outside the class. University students of an International

The marketing class students were asked to complete the RPTD-20 questionnaire and a student engagement survey at the end of spring 2020 semester. Students were offered two bonus points for survey completion as an incentive. The sample size consists of 65 students. The results, shown in Table 5, indicate that there was a significant difference in student perception of their engagement with the class during the first half of the semester (pre-COVID-19) and the second half of the semester (post-COVID-19).

Differences	*df*	*t*-Value	*p*
$Engagement_B$-$Engagement_A$	64	-6.98	<.0001

$Engagement_B$ indicates student engagement before COVID-19.
$Engagement_A$ indicates student engagement after COVID-19.

Table 5: Results of paired *t*-test for pre-post COVID engagement

We also wanted to determine what factors were unique and significant predictors of satisfaction during the spring of 2020 (COVID) for the marketing class. To that end, we conducted a stepwise multiple regression with the same dependent variable (∆Satisfaction) and independent variables (TD's and Gender) as used for the flipped class. The results are shown in Table 6. As

	R^2 = 0.565
(Constant)	0.052
∆TDST	1.006
∆TDSC	0.246
∆TDSTECH	0.163

Table 6: Stepwise Multiple Regression Results for Marketing Class

with the flipped class, ∆TDST and ∆TDSTECH were unique and significant predictors, but for the non-flipped class, ∆TDSC, the transactional content between student and course content, also became a unique and significant predictor of student satisfaction.

DISCUSSION

There is no denying that higher education was severely disrupted by the occurrence of the pandemic and the assumption is that all of higher education was impacted equally. Yet, higher education is not homogenous. It consists of online and face-to-face courses, blended and lecture learning, and interactive group learning in the form of the flipped classroom. In this paper, we have shown that the COVID disruption impacted different modes of instruction differently.

Arguably, the sector that was impacted the most was undergraduate face-to-face instruction. Both students and teachers were abruptly told to pack up, go home, quarantine yourself, and now resume the teaching/learning process as best you could.

In this paper, we have collected data and analyzed how student barriers to their engagement with their learning environment and resulting satisfaction changed before and during COVIDS. We analyzed actual data collected in a flipped classroom during the fall of 2019 and spring 2020. There were no significant differences between those two periods in any transactional distance, nor in student satisfaction, whether for the entire student population studied, of by gender. This is consistent with the findings of Swart & MacLeod (2020), and it causes us to postulate that the flipped classroom is more resilient to disruption than a traditional classroom. We attribute this to the flipped classroom being "learning centric" in that students are provided with asynchronous learning materials, including vide-lectures containing materials they must learn at their convenience. The information they have learned is then reinforced through interactive group learning, much of which is facilitated through instructional technologies. We contend that mode of learning is more "portable" when classes are banished online, as they were during COVID.

While the data we collected in the traditional face-to-face class was different than what we collected in the flipped class, it was intended to reveal if there was a loss in perceived student engagement before the outbreak of COVID (prior to March 2020) and after, March 2020. The results showed a significant perceived decrease in student engagement after the outbreak. We attribute this to a traditional class being "teaching centric" relying on the instructor to convey information to students captive in a classroom at a time and place dictated by the course schedule through lectures which they then must internalize on their own. Once a face-to-face class is "banished" online, the instructor will strive to replicate their face-to-face teaching online to students who now must exhibit the discipline to be present for a live lecture broadcast at the pleasure of the instructor. Furthermore, traditional students have not developed the same networking facility with peers and find it more difficult to engage in interactive group learning.

We can only speculate that ΔTDSC emerged as a unique and significant predictor of student satisfaction in the traditional classroom during the COVID period because the more students were able to relate to the course material, the more they were satisfied with having to learn the material. However, the most significant predictor of satisfaction in the traditional classroom was the ΔTDST, the transactional distance between the student and teacher. Interestingly, in the flipped classroom, before COVID, ΔTDST was the single most important predictor of satisfaction. However, after COVID, ΔTDSTECH became the strongest predictor of satisfaction. We postulate that when the flipped classroom was "banished" online during COVID, students had an increased reliance on technology, and less so on the instructor since they were more used to interactive group learning.

Although by necessity due to data availability, the methodologies to study the impact of COVID are different, each on its own tells a story. Flipped classes are better able to cope with the impacts of disruption (whether COVIDS or other disasters) because they are learner centric and students are more accustomed to independent learning and place greater reliance on instructional technology for their satisfaction than face-to-face students.

The implication of this research is that the COVID experience has opened a door of opportunity to institutions of higher learning. Instead of returning to doing "business as usual" and replicating the past by re-instituting the fragile face-to-face instruction as the norm, they should seize the moment to institute a more robust system of learning, flipped learning, to be better prepared to cope with future disruptions.

REFERENCES

Ash, K. (2012). Educators view flipped model with a more critical eye. *Education Week*, S6–S7.

Chronicle of Higher Education, various, retrieved from https://www.chronicle.com/ on May 15, 2020.

DeLozier, S. J. and M. G. Rhodes, (21017), Flipped Classrooms: A Review of Key Ideas and Recommendations for Practice, *Educational Psychological Review*, Vol 29, 141-151.

Francescucci, A., and Foster, M. (2013). The VIRI (Virtual, Interactive, Real-Time, Instructor-Led) Classroom: The Impact of Blended Synchronous Online Courses on Student Performance, Engagement, and Satisfaction. *Canadian Journal of Higher Education*, 43(3), 78-91.

Hockings, C., Cooke, S., Yamashita, H., McGinty, S., & Bowl, M. (2008). Switched off? A study of disengagement among computing students at two universities. *Research Papers in Education*, 23(2), 191–201. http://dx.doi.org/10.1080/02671520802048729.

Inside Higher Ed, retrieved from https://insidehighered.com/blogs/learning-innovation), May 28, 2020

Lage, M. J., & Platt, G. J., (2000). The Internet and the Inverted Classroom, *The Journal of Economic Education,* 31(1), p. 11

Moore, M. G., 1980. Independent study. In *Redefining the discipline of adult education*, ed. R. Boyd and J. Apps, 16–31. San Francisco: Jossey-Bass.

Paul, R., Swart, W., Zhang, A., & MacLeod, K. (2015). Revisiting Zhang's scale of transactional distance: refinement and validation using structural equation modeling, *Distance Education*, DOI: 10.1080/01587919.2015.1081741.

Swart, W. (2017), *Extending the Principles of Flipped Learning to Achieve Measurable Results: Emerging Research and Opportunities.* Hershey, PA., IGI Global.

Swart, W. and MacLeod, K. (2020). Flipping Online Analytics Classes: Achieving Parity with their Face-To-Face Counterparts, *Decision Sciences Journal of Innovative Education,* 18(1), 119-137.

Swart, W., MacLeod, K., Paul, R., Zhang, A. and Gagulic, M. (2014). Relative proximity theory: measuring the gap between actual and ideal online course delivery, *American Journal of Distance Education*, 28(4), 222-240, DOI: 10.1080/08923647.2014.924721.

Swart, W. & Wuensch, K. (2016). Flipping Quantitative Classes: A Triple Win, *Decision Sciences Journal of Innovative Education, 14*(1), 67- 89.

Weidlich, J. and Bastiaens, T. (2018) Technology matters – The impact of transactional distance on satisfaction in online distance learning. *International Review of Research in Open and Distributed Learning* 19(3), pp 222-242.

Zhang, A. (2003). Transactional distance in web-based college learning environments: Toward measurement and theory construction. Ph.D. diss., Virginia Commonwealth University, Richmond, VA. *Dissertation Abstracts International*. Available online at http://wwwlib.umi.com/dissertations/fullcit/3082019

DECISION SCIENCES INSTITUTE
Meaningfulness as a source for optimal decision making

Jyrki Suomala
Laurea University of Applied Sciences
jyrki.suomala@laurea.fi

ABSTRACT

The article considers human decision making from a situational perspective. The situational decision-making model assumes that an individual's prior belief is the starting point of decision making. Prior mental models activate the Because-Strategy (BS), because an individual knows what are meaningful commodities of a situation. However, if something unexpected happens, an individual changes her/his strategy from BS to the What-is-out-there-Strategy (WS). The concept of meaningfulness is analyzed from the perspectives of human decision making and Bayesian reasoning. Then the situational decision-making model will be presented. In the conclusion chapter, the future research directions will be presented.

KEYWORDS: Decision making, Situation, Mental model, Meaningfulness, Bayesian reasoning

INTRODUCTION

This paper considers how meaningfulness affects human decision making. When a human maintains and searches for meaning throughout life, s/he uses two main strategies when operating in his/her environment. Most of the time s/he behaves based on the Because-Strategy (BS). If BS does not work, a human very quickly tries to understand the situation based on the What-is-out- there-Strategy (WS). When we assume that there are two strategies as the starting point and main drivers of a decision maker (DM), most of the behavioral violations from standard economic models become understood. Both standard economic (Samuelson, 1938; Von Neumann & Morgenstern, 2007) and behavioral economic (Kahneman & Tversky, 1979; Thaler, 2016) theories assume that the aim of decision making is the recovery of the real world's options (the consideration sets of commodities) from the marketing environment. However, each DM faces the problem that objects in real life are highly ambiguous and include more information than one's can perfectly access. Thus, it follows that decision-making problems must arise from a strategy that does not rely on the commodities' real objective attributes. In a wholly meaningfulness conception of decision making, a DM's subjective experiences are determined by ordering pattern-of-information qualities of commodities according to the meaningfulness and personal goals principles. In short, the role of the mental system of DM in decision situations is not to reveal the physical world, but to promote useful behaviors. In this scheme, the situations are simply the platform in which the valuation of commodities and decision making, pertinent to coping with life, is tested (See Purves et al., 2015). Decision making reflects meaningfulness based on past experience rather than objective features of the environment. The goal of this article is to clarify this idea more specifically on a conceptual level.

The article is organized as follows: The literature review introduces and analyzes the concept of meaningfulness and Bayesian reasoning from a decision science point of view. Then the

situational decision-making model will be described. The article ends with section Conclusion and future directions.

LITERATURE REVIEW

The Concept of Meaningfulness

Standard economic theories conceptualize individuals' decision making as a simple matter of choosing, in a rational and composed way, the best of the available options in question (Kőszegi, 2010). The psychological and social dimensions of human behavior have been stripped away from these theories. On the contrary, the goal of this chapter is to introduce the concept of meaningfulness as the main phenomena in human decision making. Before concentrating on this concept more specifically, two concepts closely related to the concept of meaningfulness – utility and subjective goals – will be considered.

The original concept of utility, as Bentham (1789, according to Loewenstein, 1999) used it, refers to an individual's psychological states, like pleasures of skill, self-recommendation, a good name, power, piety, benevolence, and malevolence (Kahneman et al., 1997; Loewenstein, 1999). Until the 20th century, economists and moral philosophers devoted considerable discussion to psychological utility and its characteristics; but currently, it has been characterized without any references to psychology. Probably the clearest difference between original utility and today's economic utility is the role of consumption in these phenomena. Consumption is the most essential component of economic utility, whereas it does not have an important role in the original concept (Loewenstein, 1999). Today, behavioral economists and neuroeconomists have enriched the concept of utility by giving back its original psychological richness and complexity (Kahneman et al., 1997; Leonard, 2008; Tymula & Glimcher, 2016; Tymula & Plassmann, 2016).

The third interpretation of utility is relating to evolutional theory, and it is called biological utility (Purves et al., 2015). There is a common puzzle both in neuropsychology and behavioral economics that human representations are routinely at odds with physical measurements of real world properties and can arise from neural and mental responses that nonetheless lead to effective behaviors and decisions (McKenzie et al., 2018; Purves et al., 2015). According to the biological utility approach, the task of human representation is not to recover properties of the world but to cope with different situations by promoting useful behaviors in life. Then DMs' representation of information reflects biological utility based on past experience and the most important patterns of the current environment rather than objective features of the environment (Barkow et al., 1995; Geary, 2005; Martens, 2019; Purves et al., 2015).

These three interpretations of utility may seem contradictory and mutually exclusive. However, this may not be the case. When modeling human behavior with conceptual, computational, and mathematical tools, it is important to analyze the behavior in the context. In some situations, the DM faces one or two dimensional problem spaces, when it is suitable to apply "simple" economic utility functions (Jaynes, 2003). However, when one's faced with – as most of the everyday problems – the problem, which includes many dimensions, like emotions, attitudes, prior experience history, goals, and entities in Bentham's utility sense, then a fully adequate description of a human state and functions of mind would be represented by a vector in a space of large numbers of dimensions (Jaynes, 2003; Suomala & Suomala, 2014). Thus, the human mind is adaptive and can apply many utility strategies in different situations. This flexibility idea

is included in Botond Kőszegi's (2010) personal equilibrium –model and Agnieszka Tymula's and Paul Glimcher's (2016) Expected Subjective Value Theory. Both approaches see the traditional expected utility theory as specific cases of a wider perspective of human adaptive behavior. Therefore, it is important to keep all interpretations of utility involved when describing and explaining human decision making in different contexts.

The second essential component of meaningfulness is the DM's subjective goals. The mainstream economic models of human decision making do not include personal goals in their models and there is limited research (Abeler et al., 2011; Camerer et al., 1997; Heath et al., 1999; Kőszegi & Rabin, 2006; Suomala et al., 2017) which have shown that personal goals have an essential role in people's decision making.

However, the notion of personal goal is an essential component in the evolution theory. It posits that humans evolved specific mental abilities that were selected over time to respond to the adaptive challenges repeatedly faced by our ancestors (Barkow et al., 1995; Geary, 2005; McDermott et al., 2008). The human mind has the ability to orchestrate external and internal signal flows in accordance with internal goals in different time scales. Thus, the mind as a whole supports an individual to organize and control her or his own life in ways that will enhance the expected standard of living. A unique aspect of this evolved ability is that humans can formulate an Autonoetic Mental Representation (AMR) of potential future states and are able to manipulate these representations in ways that enable the anticipation of control-related behavioral strategies (Geary, 2005). Decision making is one of the most important parts of these control-related behavioral strategies. The essential component of AMR is the "perfect world". It is one in which the individual can organize and control her or his life in ways that will enhance the expected standard of living. The ability to set goals has evolved when people began to compare the current situation to AMR of a "perfect world". This ability to "cast oneself as a player in scenarios emerging from various choices available at any given moment" facilitates the self-regulation of behavior necessary for the achievement of personally relevant goals (Geary, 2005, p. 200). Thus, goal setting is an essential aspect of human life from an evolutional perspective.

Standard economic theories describe human decision making purely from a revealed preference perspective (Samuelson, 1938; Samuelson, 1948) and behavioral economics from a status quo perspective (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). However, there is growing evidence (Abeler et al., 2011; Camerer et al., 1997; Heath et al., 1999; Kőszegi & Rabin, 2006; Suomala et al., 2017) that goals have an important role in human decision making, and it is still desirable to build a bridge between standard economic theories and the evolution theory in order to create a more coherent scientific perspective on human decision making.

Despite AMR referring to DM's mental representation – especially representation of a "perfect world" - its content is formed by socially and culturally important factors during an individual's development. Thus, AMR includes both personal and social components; however, it is difficult to separate these components perfectly. It is essential though that a human has the ability to adapt and behave effectively in different environments and learn during his/her lifetime (Wilson et al., 2018). Today, it is customary to describe the DM as a multi-level system consisting of neurophysiological, individual, and social components and their interacting mechanisms (Thagard, 2014; Thagard & Wood, 2015). For example, the most pervasive kind of social group in all human cultures are families (Brown, 1991); thus, an individual learns to expect his/her "perfect world" according to principles of his/her social environment and culture.

Above, I have analyzed the different interpretations of concept utility and the concept of subjective goals relating to human decision making. There are some differences between these varieties of abstract knowledge, but it is useful to have a single term that includes them all. Thus, I use the concept of meaningfulness as a general term, which describes a human's general tendency to behave and find personally important content and functions in her/his everyday life. Below, I describe the concept of meaningfulness more specifically.

In his famous and evocative book Man's Search for Meaning, Frankl (2006) argues that meaning is the main driving force in man. In the same vein, Loewenstein (1999) reviewed mountaineering literature in order to describe the role of meaningfulness in the activities of mountaineers. Furthermore, Thagard (2012) argues that love, work, and play make life meaningful for most people. Therefore, it is safe to argue, that humans have a general tendency to maintain and search for the meaningfulness in life and its events. In addition, there are also more specific findings about meaningfulness in scientific literature on theoretical and empirical levels.

In the gestalt school tradition, the role of external information in determining a subject's inner mental representation was conceived in terms of gestalt laws and other heuristics (Köhler, 1992). At the core of gestalt theory is the idea that any perception would be determined by the simplest amount of the information in the event. The essential question is, what is the simplest possible amount of information. The answer depends on the properties of the context and the personal goals of the DM. For example, in the multimedia learning context, it is important to design multimedia so that it supports meaningful learning (Mayer, 2009). Then, the simplest possible amount of the information is determined by the learners' visual and verbal cognitive capacity and the essential knowledge relating to learning content. In this context, meaningfulness is related to the most essential features of the educational context. Therefore, the rational behavior is related to the scarcity of information, and the most important aspect in this process is to discard extraneous information and concentrate only on the most meaningful aspect of the context. The educational context is an illustrative example of an individual's behavior, because the goals of learning have been determined in the curriculum, which has been formed by educational policy makers, not individual learners. However, if an individual learner does not feel that the learning material is meaningful, her/his learning process will be disturbed. In the same vein, the policy makers and business people in general should not concentrate only on the offerings, but also on the decision makers' sense of meaningfulness.

In addition, the notion of meaningfulness also has specific meaning in the models of human choice. Previous research, in consumer research and psychology, highlighted that a choice becomes meaningful when the reason for the choice is to fit important personal goals (Csikszentmihalyi, 2000), which can shift across life stages (Bhattacharjee & Mogilner, 2014), consistent with the previously described notion of a "perfect world".

It has been argued that meaningfulness has two components, the purpose and recognition (Ariely et al., 2008). Recognition means that some other person appreciates an individual's behavior, not necessarily to be linked to any financial incentives. Purpose means that individuals understand how their behavior might be linked to some objectives. Ariely and colleagues (2008) propose that these twin factors are two of the hidden motivational foundations of meaning for life.

Furthermore, meaningfulness is defined as the set of contrasts a consumer is able to make with respect to the distinctions between options, for example, between brands on each attribute

dimension (Ratneshwar et al., 1987). Thus, the meaningfulness is the results of the subject's efforts to interpret the properties of context in which the choice will be made. Ratneswar and colleagues (1987) suggest that the meaning of stimulus objects is a function the DM's ability to differentiate particular stimuli from one another on a given set of attributes. In order to do that, humans need the capacity to concentrate on the most important (or meaningful) features of the environment in order to navigate in his/her environments.

An illustrative empirical example is the study in which Ariely and colleagues (2008) studied the effect of meaning to the participants' choice. In the Meaningful and Sisyphus conditions, participants received payments for assembling Bionicle Lego models according to a declining unit wage schedule. The difference between the two conditions was that, in the Meaningful conditions, the participant would place the prepared Bionicle on the desk in front of her/him; whereas, in the Sisyphus conditions, the experimenter would disassemble the prepared Binocle into pieces. Hence, as the session progressed, the completed Bionicles would accumulate on the desk in Meaningful conditions, whereas in the Sisyphus conditions, the Bionicles could not accumulate; after the second Bionicle, the subject was always rebuilding previously assembled pieces that had been taken apart by the experimenter (Ariely et al., 2008). The results show that the participants in the Meaningful condition built significantly more Bionicles than those in the Sisyphus condition (Ariely et al., 2008). Therefore, the experience of meaningfulness – i.e. recognition and purpose – influences the DM's productivity more strongly. Therefore, this article's central argument is that meaningfulness is an important driver of human behavior. When maintaining and looking for the meaning, a human can constrain a reference point in the uninterrupted commotion.

The energy-saving principle and meaningfulness are consistent with the neuroscientific understanding of the human brain. The brain processes information through action potential. and this process is metabolically very costly (Lennie, 2003). This means that our brain tissue is, on average, roughly 7 times as expensive as the average tissue in our bodies (Tymula & Glimcher, 2016). Given a fixed neural activity budget, an efficient neural representation of different objects should aim to increase discriminability between the most meaningful inputs (Woodford, 2012).

As a conclusion, human behavior is biased to socially and culturally transmitted values and attitudes. Thus, each person anticipates future situations according to meaningfulness, and this leads to the situation-specific, decision-making strategies. Typically, the DM applies BS in the situations that are familiar and repetitive. When this meaningfulness – i.e. typical expectations for the specific situation - is broken, the DM tries to understand the situation using WS and updates her/his previous mental model relating to the situation. Furthermore, it is a human's predisposition to learn to divide the world up into objects, to understand the interaction of those objects, and to apply a variety of attitudes and values to the representation of these objects. Our mind helps us find the most meaningful things in the environment (Baum, 2006). The next paragraph focuses on the mechanisms that the mind uses when trying to find meaning in the different contexts.

Bayesian Reasoning

Human behavior relies heavily on anticipating future states according to the meaningfulness of the environment under uncertainty and maintaining appropriate actions to achieve personally meaningful goals. In an everyday context, the degree of uncertainty about the possible outcomes increases drastically as the products and services around us increases constantly.

According to the Bayesian approach, a human has a prior internal model of the environment on the basis of which he acts in that environment. Based on this model, the DM anticipates and interprets the structure and functions of the situations. These internal models include both innate and adaptive components; however, it is difficult to separate these components perfectly, because the innate components are also highly adaptive and the adaptive component includes mechanisms that are highly innate (Wilson et al., 2018). When going into a situation, the DM infers how much the information in a situation corresponds to her/his anticipation. According to Bayesian terminology, the DM "counts" the likelihood (i.e. probability) of the data given a hypothesis. If the information (data) of the situation does not correspond to the DM's anticipation (hypothesis), the mental disequilibrium arises, and an individual needs to update his/her internal model. In the Bayesian terminology, this updating leads to the posterior model of the situation (Baum, 2006; Jaynes, 2003; Kording, 2014). Within Bayesian statistics, a previously acquired mental model is called prior, while the discrepancy between new information and the prior model is called likelihood. In principle, if the discrepancy between prior model and likelihood is zero, an individual gets along with BS in a situation. In these cases, the likelihood is high, because an individual's prior mental model "explains" all the information in this situation. In Bayesian terminology, the prior model explains all data in the situation. However, if the discrepancy between the prior mental model and new information increases, the DM needs to change his/her strategy from BS to WS, and update her/his prior model. The essential empirical future question is, how low the likelihood must be for an individual to change from BS to WS. However, it is important to keep in mind that standard economic theories, as well as behavioral economic theories, assume that the DM comes into a decision situation with WS.

It is essential that a human has the ability to adapt and to decide effectively in different environments and learn during his/her lifetime. The Bayesian approach gives a realistic and dynamic tool in understanding human behavior and learning mechanisms in different situations. The Bayesian approach specifies how the DM should update her/his belief upon observing information; in this way, the DM could actively reformat incoming data to better serve her/his personal goals. According to the Bayesian approach, the DM's mind gives possibilities to various outcomes in the world. Thus, at any given time, we have beliefs about the world that gives us some expectations of the states and dynamics of the context.

The world around us is complex and noisy, and it includes many uncertainties. Imagine an everyday situation where a person goes to a department store to shop. A typical department store has over 100,000 products. However, only a small amount of the products affects the DM, because s/he has a prior model of this shopping environment and knows what s/he needs. If the DM knows that her child has a birthday party, s/he has "a child birthday" shopping BS. If instead, the DM knows that a typical weekend is coming, s/he has "a weekend food" BS. The DM behaves in this noisy and complex environment relatively effectively, because s/he has the prior model of the environment, and based on this model, s/he goes to the department store because "they are there" (birthday party accessories and caterings or weekend foods). However, if something unexpected happens, for example, that something new relating to child birthday party is available in the store, s/he changes from BS to WS, and her/his typical prior model will be updated to buy new products. Being able to combine new unexpected products to the prior mental model is the very essence to constrain the complexity of the environment.

The sole essence of the Bayesian models of decision making is that this model can predict that behavior in a complex environment is close to the optimal solution to a problem encountered by the DM. For example, some studies construct Bayesian models of how subjects estimate the structure of the world (Griffiths & Tenenbaum, 2006; Perfors et al., 2011) and social context

(Khalvati et al., 2019). However, the question of how optimal the human's mind is, is still an important and open question. Understanding the problem that the mind needs to solve is defined by the environment and the basic principles of the Bayesian approach, which gives tools in this direction.

Whereas most of the studies in behavioral and neuroeconomics experiments have focused on human behavior in the very constrained environment, in which the effect of two or three variables have been studied, there is the need to more ecologically relevant models of human behavior in general, and human decision making especially. However, when information in the context increases, and a human has only the finite representational capacity to process this information, we should also know something about the structure and functions of the choice sets in the different situations. Let's consider the set of all possible outcomes, which in the real world includes all from candy bars to cars (X) (See Tymula & Glimcher, 2016). Then, assume that we consider a subset of X, the subset that defines all possible 4-element choice sets X_4. What should be intuitively clear, however, is that it is very unrealistic that X_4 contains choice sets like: {Snickers, Lamborghini, Volkswagen, Hershey} as often as it contains sets like {Snickers, Hershey, Bounty, Twix} and {Lamborghini, Volkswagen, Porsche, Toyota}. Real choice sets have structure, and a human is very skillful in building prior beliefs based on these structures. In the same vein, when the DM's goal is to arrange a birthday party for his/her child, the "prior belief" structure is different than if his/her goal is to arrange an ordinary weekend for the family. This is intuitively true, the only problem is that we do not understand enough how the DM constructs these beliefs, and what the structure of these different choice sets is. As Tymula and Glimcher (2016) argue, every intuition we have is that the set of all observable choice sets must have a high degree of internal correlational structure, a fact to which economists have paid little attention in the past. The Bayesian approach is a promising model to describe and explain human behavior in real-world choice contexts. When the different choice sets have internal consistency (i.e. birthday choice sets have more internal coherence compared to the items in the weekend choice set). In this way, the prior model constrains the "hypothetical" choice sets, and not all potential choices inside the department store are equally likely. In this way, previously learned priors can radically simplify subsequent choice sets enabled by Bayesian ideas (Kording, 2014).

Bayesian inference provides a formal way to describe the DM's behavior in complex and noisy environments. It enables one to model the uncertainty of the world by combining prior beliefs/knowledge with observational evidence to infer the most probable interpretation of the environment (Ernst & Bülthoff, 2004).

However, if we concentrate on how the DM builds his/her prior mental models based on the structure and functions of environments, this makes the set of permissible prior models (and consideration sets) rather small (Kording, 2014). Thus, constraining the prior models based on the Bayesian approach, e.g. by measuring the statistical and other regular properties of the real world, seems crucial to strengthen their predictive power. In the next chapter, I will present the contextual model of human choice based on the ideas of the meaningfulness and the Bayesian approach.

SITUATIONAL DECISION MAKING MODEL

The basic assumption of the situational decision model is that the DM does not act without considering the likely consequences of his/her actions. At the same time DM does not waste

time and effort planning for future contingencies that are very unlikely to happen (Jaynes, 2003). The DM's anticipation of essentials in different contexts is based on previous experiences and meaningfulness. The DM uses these previous experiences, which we call "priors" in the Bayesian language. Figure 1 presents the situational decision-making model.

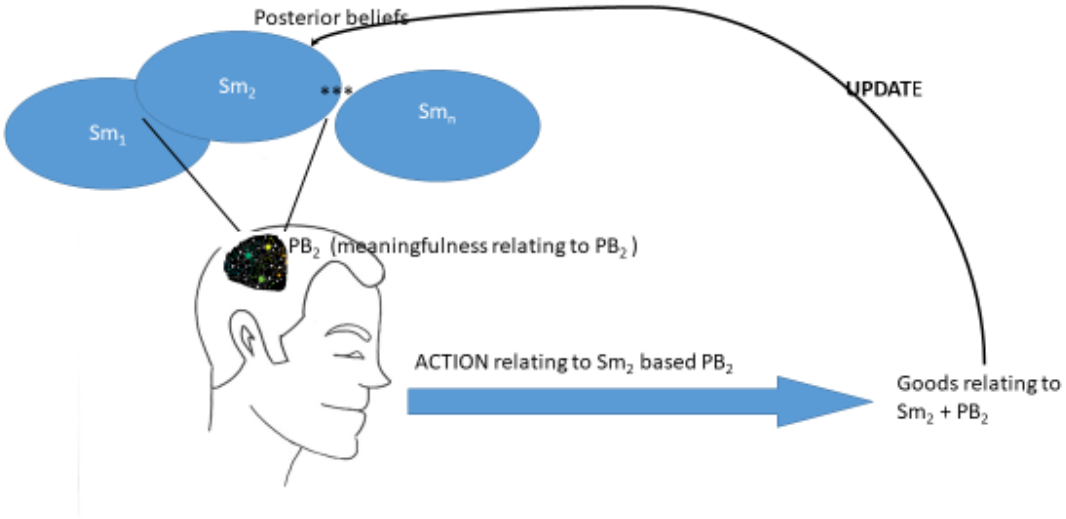


Figure 1. Situational decision-making model. Human behaves and make decisions based on situational models of different contexts. Sm_1, Sm_2 …. Sm_n mean a situational $model_1$, a situational $model_2$ and a situational $model_n$. Let's assume that the DM (Decision Maker) in the figure remembers that his daughter's birthday is coming soon. This memory activates the situational model "My daughter's birthday party", and this is described as Sm_2, which again activates prior beliefs PB_2 and the mental construct of meaningfulness about a child's birthday party. The PB2 activates the DM to make a shopping visit to the department store, where the person buys a gift/gifts and other supplies for the daughter's birthdays. If the PB_2 corresponds to the anticipation of commodities (consideration set relating to the daughter's birthday party), the update process leads to the posterior belief, which is the same as prior belief. However, if something in the department store is better than the expectations, posterior belief will be updated, and it is finally different than prior beliefs (For example, if some new ice-cream product for children's birthday party is available, the person might begin to think that it is good idea to provide also ice cream at the birthday party). Of course, updating can go also in the negative direction.

The current situational models typically assume that the decision making begins based on stimuli, like any goods in Figure 1 above (DuBrow et al., 2017; Gershman & Niv, 2013; Tymula & Plassmann, 2016), and then these stimuli activate prior beliefs about the things which cause the stimuli. For example, in DuBrow and colleagues' (2017) example, the updating process begins when an individual unexpectedly recognized, in a planned work meeting, a birthday cake. This unexpected stimuli triggers prediction error, and then an individual updates the mental model of the situation, which is now a birthday party and not work meetings. On the

contrary, it is essential, according to the situational model presented above in Figure 1, that the situational models are the starting point for the DM's behavior and decision making, and unexpectedly presented stimuli and prediction errors are not very likely triggers for the behavior/decision-making in real-life situations. On the contrary, the general mental construct "meaningfulness" is the main trigger of the behavior in this model and typically leads to BS.

Despite that the situational model presented in this article is consistent at a fundamental level with Jaynes' (2003) and Baum's (2006) models of logical thinking and rational reasoning as well as evolutional theory (Barkow et al., 1995; Geary, 2005; McDermott et al., 2008; Wilson et al., 2018), it is still only the conceptual and hypothetical model. However, it gives suitable direction to the further experimental, computational, and theoretical works. Thus, I will present the most promising direction to future scientific works in the final section.

CONCLUSION AND FUTURE DIRECTIONS

The situational decision-making model (Fig 1) presents a general approach to the human behavior in real life situations. There are three benefits of the model relating to previous models. First, the different situational models (Sm_1, Sm_2,…Sm_n in Figure 1) limit the hypothesis spaces (i.e. priors) and expected information relating these situations. Thus, it takes into account the limitations of the DM's mental capacity. When someone goes to the department store, typically s/he applies BS and goes there because s/he knows that s/he can get suitable goods for a child's birthday party or for the family's weekend dinners. The specific situational model activates the most meaningful things about situations; based on this representation, an individual constructs the consideration sets that are most suitable for each situation and applies BS.

Second, when standard economic models assume that a consumer's preferences are invariant and should follow principles of transitivity and other axioms presented on rational choice theory (Luce & Raiffa, 1989; Samuelson, 1948), the situational decision-making model takes into account the flexibility of the human's mind. This means, that a specific good can have different values inside of the option sets, depending on the situation for which the person is purchasing the equipment. Then, in the child's birthday party consideration set, a consumer may prefer potato chips over carrots and apples; whereas for the family weekend dinner party context, the same consumer will prefer apples over carrots and potatoes. This preference reveals that it is not a problem anymore according to situational model, because it is simply rational to take into account the meaningfulness in the different contexts.

Finally, the neuroeconomic models cannot explain how different situational structures affect human decision making (Tymula & Plassmann, 2016). Hypothetically, the situational decision-making model can explain this problem at least on the conceptual level. Because of each individual growth in the specific context, s/he learns to behave in meaningful ways in different situations; in the same time, s/he learns the value hierarchy of goods and other things in different situations, i.e. meaningfulness in these situations. Despite that information flow is constant, and an individual receives more information than s/he can process, each situation constrains the possible behavioral habits. Each of us then learns what the most important and meaningful issues in specific situations are; for example, in a child's birthday party, we know which things are nice and finally which things are not necessary in the situations. All of these cultural habits affect our situational models, and these models help us to find fixed points in

constant commotion. Thus, cultural habits have strong correlations for the structure of consideration sets when we behave in commercial and other situations.

The situational model is a promising way to connect real-life decision making to the more formal behavioral economics (Suomala, 2020) and standard economic models. For instance, Figure 1 depicts how one specific situational model (daughter's birthday party) affects an individual's decision making. Despite that the model concentrates only on one situational model, it is reasonable to posit that multiple situational models can co-exist at the same time, and these are not mutually exclusive. One promising direction for future research is to empirically test, how does the different situational model effects choices in different context and how individuals change their preferences in these different contexts. It is possible to test this idea both on behavioral and neuroeconomic levels. In addition, it is possible to determinate the parameters for the change from BS to WS. In other words, how much likelihood needs to be decreased for an individual to change his strategy from BS to WS. These are important questions that can be answered empirically by testing the descriptive and explanatory power of the model in terms of human decision making.

REFERENCES

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference Points and Effort Provision. *The American Economic Review*, *101*(2, April), 470–492.

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization*, *67*(3–4), 671–677. https://doi.org/10.1016/j.jebo.2008.01.004

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1995). *The adapted mind: evolutionary psychology and the generation of culture*. Oxford Univ. Press.

Baum, E. B. (2006). *What is thought?*

Beeler, J. A., & Mourra, D. (2018). To Do or Not to Do: Dopamine, Affordability and the Economics of Opportunity. *Frontiers in Integrative Neuroscience*, *12*. https://doi.org/10.3389/fnint.2018.00006

Bhattacharjee, A., & Mogilner, C. (2014). Happiness from Ordinary and Extraordinary Experiences. *Journal of Consumer Research*, *41*(1), 1–17. https://doi.org/10.1086/674724

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brown, D. E. (1991). *Human universals*. McGraw-Hill.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics*, *112*(2), 407–441. https://doi.org/10.1162/003355397555244

Csikszentmihalyi, M. (2000). The Costs and Benefits of Consuming. *Journal of Consumer Research*, *27*(2), 267–272. https://doi.org/10.1086/314324

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141–146. https://doi.org/10.1016/j.cobeha.2017.08.003

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Findlay, S. D., & Thagard, P. (2012). How parts make up wholes. *Frontiers in Physiology*, *3*. https://doi.org/10.3389/fphys.2012.00455

Frankl, V. E. (2006). *Man's search for meaning* (Mini book ed.). Beacon Press.

Geary, D., C. (2005). *Origin of Mind. Evolution of brain, cognition, and general intelligence* (Vol. 2005). American PSychological Association.

Gershman, S. J., & Niv, Y. (2013). Perceptual estimation obeys Occam's razor. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00623

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, *17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as Reference Points. *Cognitive Psychology*, *38*(1), 79–109. https://doi.org/10.1006/cogp.1998.0708

Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge University Press.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235

Kahneman, Daniel, & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes*, *103*(2), 238–255. https://doi.org/10.1016/j.obhdp.2007.02.002

Khalvati, K., Park, S. A., Mirbagheri, S., Philippe, R., Sestito, M., Dreher, J.-C., & Rao, R. P. N. (2019). Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances*, *5*(11), eaax8783. https://doi.org/10.1126/sciadv.aax8783

Köhler, W. (1992). *Gestalt psychology: an introduction to new concepts in modern psychology*. Liveright.

Kording, K. P. (2014). Bayesian statistics: relevant for the brain? *Current Opinion in Neurobiology*, *25*, 130–133. https://doi.org/10.1016/j.conb.2014.01.003

Kőszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory*, *44*(3), 415–444. https://doi.org/10.1007/s00199-009-0465-x

Kőszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165. https://doi.org/10.1093/qje/121.4.1133

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. https://doi.org/10.1017/S0140525X16001837

Lennie, P. (2003). The Cost of Cortical Computation. *Current Biology*, *13*(6), 493–497. https://doi.org/10.1016/S0960-9822(03)00135-0

Leonard, T. C. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness: Yale University Press, New Haven, CT, 2008, 293 pp, $26.00. *Constitutional Political Economy*, *19*(4), 356–360. https://doi.org/10.1007/s10602-008-9056-2

Loewenstein, G. (1999). Because It Is There: The Challenge of Mountaineering? for Utility Theory. *Kyklos*, *52*(3), 315–343. https://doi.org/10.1111/j.1467-6435.1999.tb00221.x

Luce, R. D., & Raiffa, H. (1989). *Games and decisions: introduction and critical survey*. Dover Publications.

Martens, J. (2019). Inclusive Fitness as a Measure of Biological Utility. *Philosophy of Science*, *86*(1), 1–22. https://doi.org/10.1086/701036

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology*, *77*(1), 11–37. https://doi.org/10.1348/096317904322915892

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). Cambridge University Press.

McDermott, R., Fowler, J. H., & Smirnov, O. (2008). On the Evolutionary Origin of Prospect Theory Preferences. *The Journal of Politics*, *70*(2), 335–350. https://doi.org/10.1017/S0022381608080341

McKenzie, C. M. R., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed Preferences, Rationality, and Choice Architecture. *Review of Behavioral Economics*, *5*(3–4), 337–370. https://doi.org/10.1561/105.00000091

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Purves, D., Morgenstern, Y., & Wojtach, W. T. (2015). Perception and Reality: Why a Wholly Empirical Paradigm is Needed to Understand Vision. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00156

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward Understanding the Attraction Effect: The Implications of Product Stimulus Meaningfulness and Familiarity. *Journal of Consumer Research*, *13*(4), 520. https://doi.org/10.1086/209085

Samuelson, P. A. (1938). A Note on the Pure Theory of Consumer's Behaviour. *Economica*, *5*(17), 61. https://doi.org/10.2307/2548836

Samuelson, P. A. (1948). Consumption Theory in Terms of Revealed Preference. *Economica*, *15*(60), 243. https://doi.org/10.2307/2549561

Suomala, J. (2020). When the easiest choice is the best one: A rational consumer follows the principle of the Occam's razor. Manuscript under review.

Suomala, J., Hlushchuk, Y., Kauttonen, J., Heinonen, J., Palokangas, L., & Numminen, J. (2017). Distributed brain networks reflect salary offer in accordance with the prospect theory's value function. *Journal of Neuroscience, Psychology, and Economics*, *10*(4), 167–180. https://doi.org/10.1037/npe0000083

Suomala, J., & Suomala, V. (2014). Modified Reinforcement Learning Infrastructure. *2nd International Conference on Applied Social Science Research (ICASSR 2014)*, *2*, 95–97.

Thagard, P. (2012). *The brain and the meaning of life*. Princeton University Press.

Thagard, P. (2014). The self as a system of multilevel interacting mechanisms. *Philosophical Psychology*, *27*(2), 145–163. https://doi.org/10.1080/09515089.2012.725715

Thagard, P., & Wood, J. V. (2015). Eighty phenomena about the self: representation, evaluation, regulation, and change. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00334

Thaler, R. H. (2016). *Misbehaving: the making of behavioural economics*. W. W. Norton & Company.

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Tymula, A. A., & Glimcher, P. W. (2016). Expected Subjective Value Theory (ESVT): A Representation of Decision Under Risk and Certainty. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2783638

Tymula, A. A., & Plassmann, H. (2016). Context-dependency in valuation. *Current Opinion in Neurobiology*, *40*, 59–65. https://doi.org/10.1016/j.conb.2016.06.015

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior* (60th anniversary ed). Princeton University Press.

Wilson, D. S., Hayes, S. C., & Biglan, A. (2018). *Evolution & contextual behavioral science: an integrated framework for understanding, predicting, & influencing human behavior*.

Woodford, M. (2012). Prospect Theory as Efficient Perceptual Distortion. *American Economic Review*, *102*(3), 41–46. https://doi.org/10.1257/aer.102.3.41

DECISION SCIENCES INSTITUTE
Training Synergy

Andre Vieira Caputo
Embry Riddle Aeronautical University
vieiraca@my.erau.edu
Everton Amieiroe
Embry Riddle Aeronautical University
amieiroe@my.erau.edu
Lucio de Araujo Alves
Embry Riddle Aeronautical University
dearaujl@my.erau.edu
Renan Dapena Roveran
Embry Riddle Aeronautical University
dapenarr@my.erau.edu
Leila Halawi
Embry Riddle Aeronautical University
halawil@erau.edu

ABSTRACT

Brazil is facing a growth in the airline industry where all the major airlines are currently hiring pilots, flight attendants, and mechanics. One of the challenges of those airlines is to provide training for all these professionals. Many times they do not have instructors or classrooms available. It is crucial for the airlines to promote faster and better training, to reduce cost, save time and resources to prepare a better professional. In this paper, we are showing the possibility of increasing the cost performance of airline training by providing alternatives to airlines. The idea is to promote synergy between the airlines to optimize and share training between companies. If airlines combine their training, they can have more classrooms, seats, and flexibility together with lower costs and saving time

KEYWORDS: Airlines training, Training synergy, Brazil Airlines

INTRODUCTION

Many companies in the world require training as part of their culture and operation requirements. For airlines, it is extended to a second step due to the mandatory needed training per the related country aeronautical authority necessary to keep a high safety environment that is mandatory for this industry. Due to the resources involved like trainers, students, room, infrastructure, time, preparations, etc., training has a significant impact on operations and budgets for airlines. All airlines in Brazil must have a training program, but most parts of these programs are identical for all companies due to ANAC (Brazilian Aviation Authority) standard requirements. The basic training is given for all employees in functions like mechanics, pilots, flight attendants, airport ramp personnel, check-in staff, and are similar between airlines. This study analyzes the load factor of the Brazilian airline training classrooms. It creates a process to provide seats that can be used by increasing the availability of vacancies and dates for training. Our solution will generate greater flexibility, thus resulting in a reduction in the hiring bottleneck and an increase in the load factor in the classroom. With this, airlines will have more flexibility to

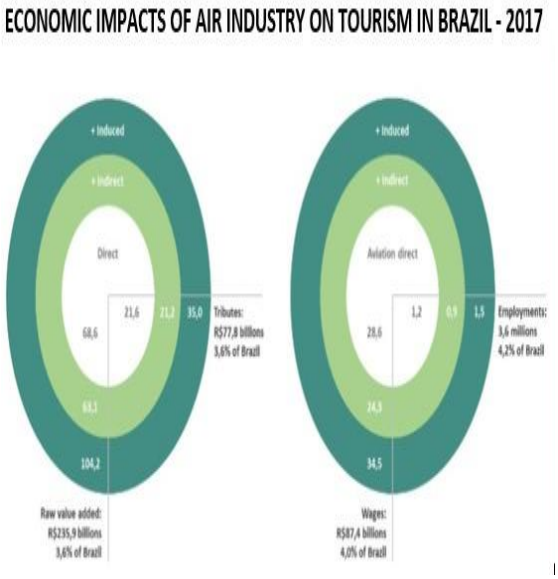
train their employees, besides being able to reduce the amount of headcount needed for the training staff. The operational groups within an airline are composed of several areas. Pilots, mechanics, and flight attendants are groups of employees highly specialized and trained that have a lot of training that can be shared between the companies. These groups will be used through this study, but the process presented herein can be expanded for other areas. It will also provide a general overview related to strategic partnerships to clarify the current scenarios of the industry. The purpose of this study is to show an alternative for better use of time, money, and resources when offering basic training. It will improve the cost performance of the airline's training by increasing the availability of dates and seats through the creation of synergy between companies. This synergy will reduce the cost of operations, which gives airlines the flexibility to work under unpredictable schedules changes. Also, it will lead to more significant interaction between professionals from different companies that will result in enriching the knowledge and exchange of experience during the training.

Strategic Partnerships and the Generation of Results

The airline industry in Brazil is essential and has a significant impact on the economic activity of the country. Data from IBGE 2017 shows that Brazil's size is 8.5 million square kilometers, with an estimated population of 208.4 million inhabitants and a population density of 22.43 inhabitants per square kilometer. The territorial dimensions are enormous, and without aviation, it would be impossible to reach some areas. According to the National Civil Aviation Agency – ANAC (2011), the regular carriers companies transported approximately 88 million passengers, and by the end of 2016, the number reached 100 million, an increase of 13% over 2011.

The National Civil Aviation Agency (ANAC) regulates all Brazilian Airlines operating in Brazil. The regulation includes all operational rules and mandatory training in some functions with the primary goal to unify knowledge and ensure safe operation for all. The total number of airline employees is approximately 53,000 at the end of 2016, with regular operations at 126 commercial airports in the country, as stated by ABEAR. According to ABEAR (2017), Brazilian commercial aviation contributes 3.1% of Brazil's economic output, with R$ 312 million added to the Brazilian economy, approximately R$ 39 million produced directly by airlines, airports, and ground services.

Figure 1



The composition of the total operating expenses of an airline company is complicated. It is related to aircraft maintenance, leasing and insurance costs, fuel costs, airport charges, personnel costs, among others. Brazilian aviation saw its operating expenses grow approximately 32% from 2002 to 2015, reflecting one of the most critical components, the fuel cost, which increased by 74% over the same period.

In the first semester of 2019, the airline market in Brazil changed with the exit of Avianca Brazil due to bankruptcy. Now there are three major airlines: LATAM, founded in 1976, GOL founded in 2001, and AZUL, founded in 2008. According to ABEAR 2017, those airlines combined have more than 53.000 employees.

Table 1

Public Air Services Charges

(*) Controlled Airline Companies (Millions of R$)

	2017	2016	2015
(-) Personnel	-5.308,2	-4687,5	-5103,3
(-) Crew and staff costs	-568,4	-392,4	-279,6
(-) TRAINING	-56,4	-35,8	-39,8
(-) Fuel and lubricants	-9.370,4	-8173,9	-10052,1
(-) Passengers assistance	-251,2	-178,7	-342,3
(-) Extrajudicial damages	-57,7	-56,5	-78,9
(-) Court convictions from air services	-300,3	-215,1	-183,1
(-) Cathering	-386,7	-412,1	-479,9
(-) Handling	-829,2	-488,6	-506,7
(-) Aircraft cleaning	-15,1	-18,8	-18,6
(-) Insurances	-95,7	-121,9	-119,8
(-) Leasing	-4.205,1	-4652,5	-4604,4
(-) Aircraft & Engines Maintenance/ Overhaul	-2.253,5	-2652,5	-2429,3
(-) Depreciation/ Amortization/ Exhaustion	-1.087,6	-1091,1	-1053,2
(-) Airport charges	-1.160,4	-1112,8	-1164,7
(-) Specific prices	-47,6	-37,9	-26,5
(-) Communicatio and aid to air navigation charges	-1.311,0	-1293,5	-1157,6
(-) Taxes and fees	-338,7	-379,9	-139,1
(-) Other air public services charges	-1.042,6	-1103,4	-1146,8

Source: ABEAR

Table 2 – Number of Employees (December 31, 2017)

	AVIANCA	AZUL	GOL	LATAM Airlines	Total/Function
Pilots	652	1,706	1,545	1,987	**5,890**
Flight Attendant	1,242	2,419	2,862	4,724	**11,247**
Mechanics	407	1,663	2,105	2,307	**6,482**
Airport Staff	1,511	2,919	5,027	3,587	**13,044**
Other Staff	1,645	2,638	3,869	9,466	**17,618**
Total/Airline	**5,457**	**11,345**	**15,408**	**22,071**	**54,281**

Source: ABEAR (2017)

Alliances and Partnerships

The competitive business environment has required companies to be flexible, innovative, and cost-effective. However, few have the capabilities and resources to adopt this behavior consistently. This is why the number of those that embrace partnerships and alliances is increasing, achieving competitive advantages (Kanter, 2001).

The new competitive processes, as explained by Doz and Hamel (2001), are related to the partnership between companies. The reason for that is that the competition no longer happens only between products or services, but between different business concepts, as well. According to Lewis (1992), the strategic partnership enables companies to gradually achieve the capacity to develop products, reduce costs, acquire new technologies, and get more resources to invest in its core competencies.

The objectives for the formation of an alliance can be classified into two categories. Those oriented to the search for greater efficiency, by the use of shared resources, or those with a market orientation, in which case another classification would also be pertinent: defensive or offensive objectives (Kleymann; Seristo, 2001). This efficiency enhance-oriented concept, as stated by Kleymann, Seristö (2001), also enables the associated companies to develop a more offensive position in the market.

In the United States, many airlines split their training between other companies and sometimes outsource for third parties. Those trainings are not exclusively to their airlines, which is the case of Pan Am Academy. Pan Am international flight academy is a leading provider of training support for airlines and aviation professionals. It is a surviving division of original Pan American World Airways, which was founded in 1980 in Miami, Florida, being one of the most experienced training facilities in the world. Its focus is to provide professional training for pilots, flight attendants, and mechanics. Pan Am has FAA-approved programs that meet the requirements of Part 91,121,125 and 135, which enable the Company to provide training for airlines such as initial and recurrent.

Many airlines flying from and to Miami use Pan Am structure for training, sometimes combining classes between then and creating a partnership. By doing this, they are optimizing their resources by reducing costs and saving time.

The business partnership connects companies with strong exploration intent that seeks, above all, to ensure new gains with the application of resources and specific capabilities available to each partner.

Strategic Role of Training and Personnel Qualification

Currently, the airlines have their schedule of training following internal procedures. In this specific analysis, we will deal with the regular training that is mandatory by the regulatory agency (ANAC).These required training address some of the many functions that airlines have in common, such as airport agents, administrative assistants, ramp operators, cargo operators, operational supervisors, airport coordinators, pilots, mechanics, and flight attendants.

Activities performed near an aircraft represent high complexity and have an increased risk of incidents or accidents. In an attempt to reduce and ultimately prevent these occurrences, mandatory training is required. The training is designed to address all of the activities that will be performed by each job function, as well as their frequency and evaluation.

Companies in Brazil, such as AZUL, GOL, and LATAM, and all the leading airlines in the world, are committed to ensuring that all employees with specific activities are trained and approved to perform their duties with technical capability and with minimal to no risk to the operations. In a macro analysis, it is possible to identify that each airline, operating in the same 126 commercial airports in Brazil, has an individual structure to meet precisely the same training demand. Therefore, we have at least three times the contingent number of instructors and classrooms for applying the same mandatory training regulated by ANAC.

Besides all the workforce and resources needed, it is also necessary to consider the logistics related to moving people all over Brazil. In this environment of constant change, professional education has played a strategic role in organizations. According to Ilyas, Hin, and Adnan (2016), training, previously was seen as an administrative function and has evolved into a

strategic initiative that aims at organizational profitability. Organizations are recognized as knowledge organizations, as much as a worker is a knowledge worker, as people are continuously learning and expanding their skills and capabilities, says Drucker (2008). Santos (2010) presents thirteen learning modalities, as shown in Table 3.

Table 3- Modalities of Learning

Type of Training
In-Person Training
Distance Learning (e-learning)
On Job Training
Job Rotation
Internship
Technical Visit
Informative Meetings
Dissemination Meetings of Training
Conferences
Subscription to journals and magazines
Workshop
Seminar
Congress

In the aviation industry, the most used types of learning are in-person training and on the job training, due to the technical skills and background needed for operational activities.

DATA SOURCE, COLLECTION, AND ANALYSIS

As a source, training data from AZUL, GOL, and LATAM that are the three leading airlines in Brazil was used.

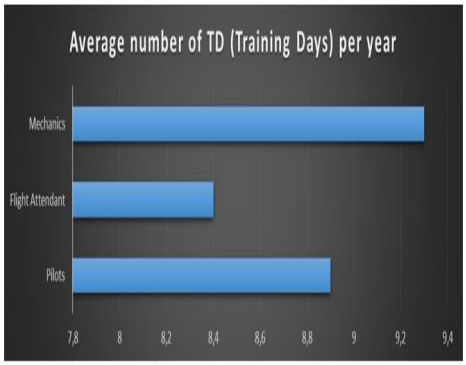
Data collected included:

. A number of seats offered for each function (pilots, mechanics, flight attendants).
. A number of seats used for each function.
. Training Programs approved by ANAC (Training Catalog).

As shown in Graph 2, it is possible to see the average number of days spent on each function. Mechanics has the highest number of TD per year (9.3 days).

Graph 2- Average number of TD (Training Days) per year

Source: Maintenance: Internal Company Resource/2017; Crew: Estimative based on other airlines and ABEAR data

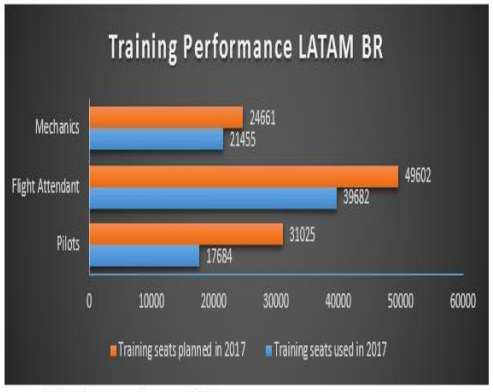


Source: Internal Company Resource/2017

To meet the needs of the days of training for beginners, revalidation, indoctrination, etc., the companies schedule the classes based on the necessity of the different areas. They offer a scale of flexible training in separate schedules, seeking to reduce the number of adjustments of working hours per shift, the number of trips for training, and scheduling the training with the least possible overlap.

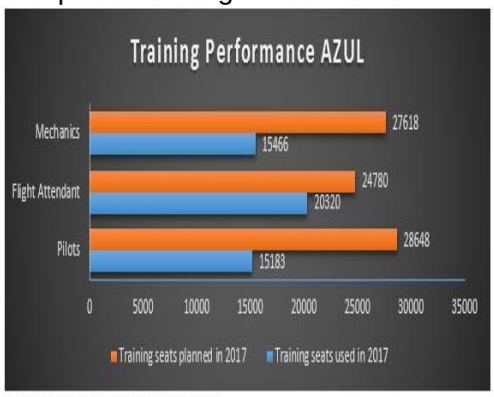
Through Graph 3 and Graph 4, it is possible to see the number of training seats scheduled for 2017 multiplied by their real occupation showing flexibility.

Graph 3- Training Performance LATAM BR



Source: Internal Company Resource/2017

Graph 4- Training Performance AZUL



Source: Internal Company Resource/2017

It is possible to note that each Company has training seats planned versus training seats used, but no one used 100% of its seat availability.

Based on compiled data with all companies' information, the Graph 5 shows that the macro scenario of the training class occupation has a high quantity of seats that are still available and not used (wasted).

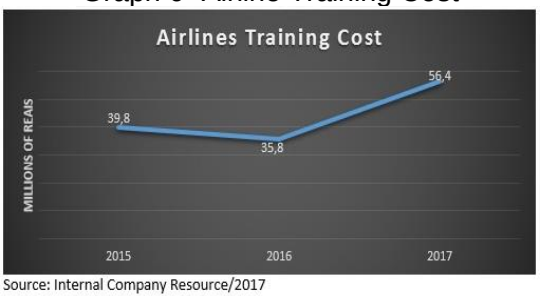
Graph 5- Training Performance of Airlines

Source: Internal Company Resource/2017

To make training seats available, companies need to provide resources, such as infrastructure, instructor, simulators, and so on; all these resources generate costs.

From the data presented in Table 1, we can get to Graph 6 that shows the Brazilians' airline training cost in recent years.

Graph 6- Airline Training Cost

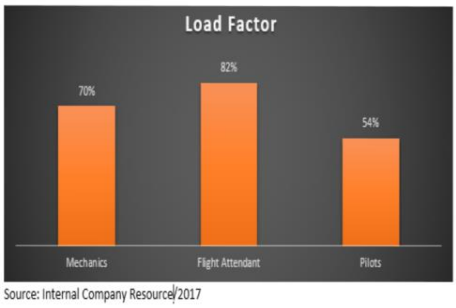


Source: Internal Company Resource/2017

Training seats, as well as the seats of an aircraft, are perishable. In other words, we only have one opportunity to use them. If the occasion is not taken, it will not be possible to recover the resources used to conduct the training.

Based on the data of Graph 5, we can compare the planned training seats against the used training seats, as shown in Graph 7.

Graph 7- Load Factor



Source: Internal Company Resource/2017

Given the characteristics of each Company and the changes inherent to the market, we can describe some factors that contribute to this low load factor: same schedule, operational contingencies, sickness, and turn-over.

Therefore, it is possible to calculate the average between these functions to find a number that represents the Average Load Factor in Brazil, considering airlines presented in this capstone, which is 69%. Considering this information, it is possible to generate Graph 8 that shows the correlation between training cost and load factor.

Graph 8- Training Costs x Load Factor

Source: Internal Company Resource/2017

Considering the high level of misspent resources verified through the data shown in Graph 8, this capstone provides enough information to promote the evaluation of resource sharing between the three airlines.

OUTCOMES

As described in ANAC IS 145-010 (Maintenance Organization Training Program), airlines can share training with other companies to improve the performance against cost and quality, as shown below:

> *"5.3.5.3 To manage costs, companies with similar needs can share training costs. These companies may also want to work as training entities or work with an aeronautical school to develop courses. Such schools and other companies may formalize cooperation agreements."*

Table 4 brings some examples of courses that are similar between airlines and can be shared.

Table 4- Similar Training Examples

Course	Public	LATAM	AZUL	GOL
Low Visibility Operation	Pilots/Mechanics	X	X	X
PBN – Performance Based Navegation	Pilots/Mechanics	X	X	X
ETOPS – Extended Twin Operations	Pilots/Mechanics	X	X	X
RVSM – Reduced Vertical Separation Minimun	Pilots/Mechanics	X	X	X
Meteorology	Pilots	X	X	X
Air Traffic Control Regulations	Pilots	X	X	X
AVSEC – Aviation Security	Pilots	X	X	X
Standard Practices	Mechanics	X	X	X
Standard Practices for Avionics	Mechanics	X	X	X
Standard Practices for Structure	Mechanics	X	X	X
Emergency Procedures	Pilots/Flight Attendant	X	X	X
Dangerous Good	Pilots/Flight Attendant/Mechanics	X	X	X
Damage Assesment	Mechanics	X	X	X
Fire & Smoke Training	Pilots/Flight Attendant	X	X	X
ITC – International Traffic Control	Pilots	X	X	X
Ground School A320	Pilots	X	X	
Level 2 and Level 3 Course for A320 Family	Mechanics	X	X	

In a meeting with Mr. Antonio Augusto do Poço Pereira, ABEAR Chief Financial Officer, held on 08/22/2019 at ABEAR headquarters in São Paulo, we received the confirmation that ABEAR is willing to support this research. It is described below the proposed processes to ensure synergy between the companies to increase the use of training seats and flexibility for operation.

General Considerations

The following procedure only applies to available seat-based training offered by companies on the Training Schedule Form. There is no obligation for a company to provide training and seating for the program. The Decision Board is responsible for deliberating using common sense in the event of disputes related to this program between companies.

Training Schedule

01 – Survey of macro needs for next year

In the third quarter of the year, companies are expected to fill the Training Schedule Form with next year's macro needs. The Decision Board decides the deadline for this submission.

02 – Annual Training Schedule

With the Training Schedule Form completed by companies, next year's schedule will be defined, presenting:

1. Which Company will conduct the training
2. What is the place of training
3. Number of seats available for each Company

After the discussion rounds, the Training and Seating Grid is defined, which is uploaded to the ABEAR website.

03 – Micro needs assessment and adjustments

Companies can readjust their needs by adding or removing participants from each training, always targeting the next quarter.
To this end, the Training Necessity Form is filled by companies with the additional needs that may have arisen for the next quarter.

04 – Review of training and seating availability schedule

The Decision Board meets to deliberate on micro needs for the next quarter, will be reviewed:

1. Which Company will conduct the training
2. What is the place of training
3. Number of seats available for each Company

Each Company chosen to conduct the training must review the proposal, and its Decision Board representative must accept, decline, or make another proposal. After the discussion rounds, the Training and Seating Grid is defined and uploaded to the ABEAR website. There are no limits on removal, but the Seat Value will be charged anyway. For additions, it is considered the Company that completed the Training Necessity Form first.

When there are no vacancies available for a particular Review Period, each Company should either take care of their training or jointly create new classes. The Training Necessity Form should be filled anyway. If any company decides to make training and seating available, it must add the offer to the Training and Seating Grid. Software that manages the Program site on the ABEAR site automatically identifies supply and demand and schedules training.

05 – Financial Agreements:

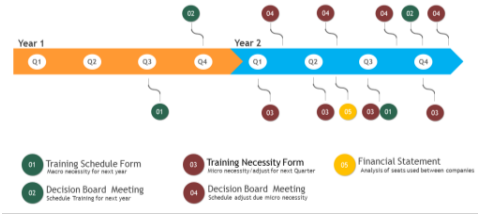
For each seat occupied by one Company in training for another Company, one unit of Seat Value is counted. This control is done automatically by the software that manages the Program website on the ABEAR website. Once a year, the full accounting is made (sums and deductions), and the monetary value per Company for each other Company is defined.

Each Company must generate an invoice against the other Company, which will have a payment term of 90 days from the invoice date.

Training Program Timeframe

To summarize the entire process already detailed in this chapter of the capstone, the macro training program timeframe can be observed in Figure 1.

Figure 1- Training Program Timeframe



Each Company will fulfill the training schedule form or the training necessity form, in case of need to change the schedule, which will be submitted to the decision board to evaluate and then provide the financial statement.

CONCLUSION

A higher amount of training available will bring more agility in the personnel qualification process since it enables to speed up the training process. Consequently, it is possible to have more people capable of maintaining the operation, making it more flexible, so that technicians can accomplish more training, improving more and more their skills.

With the information provided in the methodology section, it was possible to see the number of training resources available being wasted and the opportunity for the improvement of training cost-performance. Based on the process in the analysis section, this opportunity is captured and shared with other airlines, reducing the 31% misspent presented in Graph 8. With the proposed program, airlines supporting this idea, and ABEAR already interested in it, the Training Synergy implementation and its success is highly feasible. Also, this Program concept can be implemented between any airlines in the world or any other industry that has similarities between their training, adjusted to their realities and environment.

RECOMMENDATIONS

We want to recommend for airlines to form a collaborative group and commit resources to the implementation of this program, making the necessary adjustments to the process as required and according to the reality of each Company. Also, the engagement of ABEAR is imperative for the project's triumph so everyone can get benefits from the cost reduction that could be achieved. Topics for future development and implementation

The following topics are a brief list of points that should be analyzed and implemented to grant the Program success:

- Development of the software that manages the Program website. It will need IT resources from airlines and ABEAR. Security and reliability of the application are mandatory for the project's credibility.
- External audit of the software that manages the program on the ABEAR website to grant the project's credibility, pulling away any compliance issues and concerns among companies that are direct competitors.
- A detailed study about the similarity of training which may be offered between airlines: the preliminary analysis showed that companies operating the same type of aircraft have a high degree of training similarity.
- Verification of impacts related to legal and tax aspects, as there will be non-financial (course balance) and financial exchanges between companies. This study should

understand what fees will be applied and how the contract between the parties should be built.

- Related to improving the load factor of training, it is essential for the airlines to manage the schedules of the employees together with ABEAR and the board to increase the load factor.
- With the implementation of this methodology and following the premises presented, we will have a constant growth of the load factor of the classrooms, where we will undoubtedly have a jump from 69% to more than 85% load factor, consequently reducing the misspent by 16% representing then a real economy of R$ 9,020,000.00.

REFERENCES

ABEAR. (n.d.). Associação Brasileira das Empresas Aéreas. Retrieved August 03, 2019, from https://abear.com.br/.

ANAC. (n.d.). Dados e Estatísticas. Retrieved August 03, 2019, from https://www.anac.gov.br/assuntos/dados-e-estatisticas.

Avianca Brasil's Bankruptcy. (2019, July 08). Retrieved August 20, 2019, from https://www.efe.com/efe/english/business/avianca-brasil-s-bankruptcy-creates-challenges-for airline-industry/50000265-4018484#

FAA. (2016, November 22). Training Resources & Guides. Retrieved August 03, 2019, from https://www.faa.gov/training_testing/training/.

IBGE. (n.d.). Retrieved May 18, 2019, from https://www.ibge.gov.br/home/.

ICAO Documents. (n.d.). Retrieved March 15, 2019, from https://www.icao.int/airnavigation/Documents/Forms/AllItems.aspx.

IS 145-010A (ANAC). (2013, August 30). Retrieved March 12, 2019, from https://www.anac.gov.br/assuntos/legislacao/legislacao-1/iac-e-is/is/is-145-010.

Ilyas, M., Hin, C. W., & Adnan, Z. B. (2016). Relationship between strategic training and organizational profitability: A partial least square (PLS-SEM) approach. International Review of Management and Marketing, 6(7S)

Pan Am International: Flight Academy. (n.d.). Retrieved June 22, 2019, from https://panamacademy.com/.

Panorama ABEAR 2017. (2018). *Panorama ABEAR 2017*. Retrieved from https://abear.com.br/wp-content/uploads/2019/03/PanoramaABEAR_2017-atualizacao-29032019.pdf?utm_source=Site Panorama

DECISION SCIENCES INSTITUTE

Stochastic optimization of inventory routing problem with planned transshipment and change of minimum delivery unit constraint in emergency logistics

Tomoki Oshima
Tokyo University of Science
Email: 7419505@ed.tus.ac.jp

Aya Ishigaki
Tokyo University of Science
Email: ishigaki@rs.tus.ac.jp

Ito Takumi
Tokyo University of Science
Email: t-ito@rs.tus.ac.jp

Mori Kenjiro
Tokyo University of Science
Email: ken-mori@rs.tus.ac.jp

Ryuta Takashima
Tokyo University of Science
Email: takashima@rs.tus.ac.jp

ABSTRACT

This study focuses on the emergency logistics of building members. The proposed model aims to minimize the penalties for late delivery and the required costs. When formulating a delivery plan, it is necessary to consider two uncertainties: uncertainty of demand scale and changes in processing conditions of building members. Scenario-based modeling with stochastic programming is used to account for uncertainty of demand scale. Changes in processing conditions in the delivery plan are incorporated as minimum delivery unit constraints. The effects of introducing planned transshipment were investigated and shown to contribute to the minimization of object values in uncertain environments.

KEYWORDS: Inventory routing problem, minimum delivery unit constraint, uncertainty of demand scale, planned transshipment, collaborative strategy

INTRODUCTION

Natural disasters cause great damage to humans and infrastructure and buildings, but it is difficult to predict the scale of damage. In particular, delays in recovering from damage to infrastructure and building will cause serious problems such as population outflows, industrial decline, and regional polarization. In fact, the restoration of damaged buildings by the Great East Japan Earthquake in 2011 continued for more than 8 years, and the early restoration was recognized as an issue (Kazama & Noda, 2012). For the above reasons, it is required to formulate a plan for the early restoration of damaged buildings. On the other hand, it is also necessary to reduce the cost of restoration as much as possible. It is difficult to make such decisions due to the diversification of purposes and the uncertainty of information and data. This

study provides a model that minimizes the supply delay penalty and supply cost of building members to damaged buildings based on uncertain damage data.

Emergency logistics aimed at reducing the damage caused by disasters are drawing attention (Yuan & Wang, 2009). Emergency logistics assumes a situation where supply capacity is not sufficient for demand. Therefore, it is necessary to develop an appropriate delivery plan for the demand with limited resources. Emergency logistics are classified into three types according to response timing: the preparatory phase, the response phase, and the restoration phase (Özdamar & Ertem, 2015). In the preparatory phase, risk reduction before disasters is performed, such as strengthening infrastructure and pre-positioning inventory. In the response phase, rescue operations are carried out immediately after the disaster. These include facility locations, distribution of relief supplies, mass evacuation, and casualty care. In the restoration phase, the damaged system is restored through activities such as repairing and rebuilding infrastructure and delivering members (Anaya-Arenas , Renaud & Ruiz, 2014). This study focuses on the delivery plan of building members in the restoration phase for the purpose of early restoration of the affected house.

Research on emergency logistics has been studied for various purposes, such as length of planning period, satisfaction of demand, high cost, and long distance of delivery. The purpose of Yi & Özdamar (2007) is fairness of demand satisfaction. They used healthcare facilities' location and delivery plans for prioritized demand to minimize delays in providing each service. The purpose of Balcik, Beamon & Smilowitz (2008) is to minimize delivery costs. They looked at relief delivery plans and sought to maximize the benefits of helping recipients. Caunhye, Zhang, Li, & Nie (2016) proposed an emergency delivery model using lateral transshipment. In their model, to respond to uncertain demand in preparation for urgent deliveries, pre-placed inventory locations and subsequent delivery routing were performed.

In this study, inventory routing problem (IRP) are used as a way to reduce delays in the delivery of building materials of building members (Bell et al., 1983). The IRP considers with both inventory planning and routing planning simultaneously.

In emergency logistics, it is necessary to consider the discrepancy between the information of the delivery site and the disaster area. This study considers two types of uncertainty. The first is the uncertainty of demand scale. Demand is given according to probability data. Expected value problem (EV) and stochastic programming (SP) are the methods to handle the demand given by probability. It is known that solution of SP is superior to the solution of EV. They modeled on the scenario given by the discrete probability distribution (Mete & Zabinsky, 2010). The proposed model is developed as a scenario-based model using SP. The proposed model generates an optimal delivery plan when a finite number of scenarios with discrete probability distributions are given. The second is the uncertainty of the processing state of building members. It is difficult for the delivery base to correctly understand the processing form of necessary materials in the demand point. Therefore, building members must be shipped unprocessed from outside the demand point and processed at the demand point. That is, building members should consider the minimum delivery unit constraint. The emergency logistics IRP in previous studies cannot take into account the uncertainty of demand scale and the change in the minimum delivery unit constraint due to the processing by building members. This research propose an IRP model that considers the uncertainty of demand scale and the change of the minimum delivery unit constraint (IRPSPDU).

In an emergency, each demand point should work together to minimize delivery delays to building members. This study proposes a collaborative strategy that is planned transshipment. Planned transshipment has recently attracted attention as a method of reducing delivery costs. Planned transshipment is a way for multiple stores to temporarily share a warehouse to reduce transportation distance and reduce costs [8]. Planned transshipments have been used in the case of Brazilian retailers and have been shown to be effective in reducing the total cost of

multi-period, multi-product IRP [9]. This is called IRP with planned transshipment (IRPT). IRPT in logistics usually requires the consideration of interests between stores, but in emergency logistics it is possible to collaborate without stakes. In this study, planned transshipment is used as an indifferent collaborative strategy with the goal of minimizing delays in the delivery of building materials. The minimum delivery unit constraint for building members changes when processed at the demand point.
However, there is no previous study that uses a planned transshipment model to minimize delays in the delivery of building materials. This method requires consideration of stakes between stores. But, in emergency logistics it is possible to collaboration without stakes. In this study, planned transshipment is used as a collaborative strategy without interest, with the aim of minimizing delays in the delivery of building materials. The minimum delivery unit constraint occurs in the same way as the first strategy, but the planned transshipment is expected to reduce the waste generated by processing at the demand point.
This study proposes an IRPT stochastics programming model with minimum delivery unit constraint (IRPTSPDU). IRPTSPDU have demand uncertainty and vehicle number constraints, capacity constraints. The purpose of IRPTSPDU model is to minimize delay penalty and required costs.

MODEL BUILDING

This section shows the precondition of the IRPSPDU and IRPTSPDU model. The purpose of these model is to minimize delay penalty in the delivery and required costs. These model delivers from delivery base to the demand points in the affected area.

Notation

The following notations are used in this study.

Sets

V^d	Delivery base
V^c	A set of *n* demand points
V	$V^d \cup V^c$ (A set of *n+1* nodes)
K	A set of k_{max} vehicles
T	A set of t_{max} periods
Ξ	A set of ξ_{max}demand scale scenarios $\{\xi^1, \dots, \xi^{\xi_{max}}\}$

Parameters

C_i	Maximum storage capacity of each demand point *i*
C_{base}	Maximum storage capacity of the delivery base
$d_{t,i}$	Demand of each demand point during each time period *t*
Z	Maximum capacity of vehicle
$r_{i,j}$	Distance of the arc (*i, j*)
Lot	Minimum delivery unit in delivery base
Lot'	Minimum delivery unit in processing base
α	Penalty cost due to late delivery of building members
β	Penalty cost due to excess inventory of building members

γ Cost due to shipping distance

Decision variables

$Q_{t,i,k}(\xi)$ Quantity of building member shipped from delivery base by vehicle *k* to demand point *i* in period *t* under disaster scenario ξ

$x_{t,i,j,k}(\xi)$ Decision to use or not use vehicle *k* in arc (*i, j*) in time period *t* under disaster scenario ξ

$y_{t,i,k}(\xi)$ Decision to arrive or not at demand point *i* vehicle *k* in time period *t* under disaster scenario ξ

$yy_{t,k}(\xi)$ Decision to use or not use vehicle *k* in time period *t* under disaster scenario ξ

$I_{t,i}(\xi)$ Inventory level of demand point *i* in period *t* under disaster scenario ξ

$QC_{t,i,k}(\xi)$ Inventory level in vehicle *k* shipped to demand point *i* in period *t* under disaster scenario ξ

Precondition

This section presents IRPSPDU and IRPTSPDU model. The preconditions are as follows.

- These model are represented by a complete graph, with demand base and demand points as vertex vertices, and their vertices connected by straight edges.
- The planning period is a multi-period.
- The purpose is minimization of such period until the satisfaction of all demands.
- The base and each demand point have a limited inventory capacity.
- The finite number of scenarios are denoted $\xi \in \Xi$ in the formulation. The demand scale of the demand point is given according to the scenario.
- The demand given for each demand point is based on very accurate forecasts.
- A single building member is considered in this model.
- Unfulfilled demands are carried over to the next period.
- Building members have a minimum delivery unit constraint, which varies with processing.
- The minimum delivery unit constraint for building members varies at demand points and is smaller than the constraint on the delivery base.
- Loss of building members due to processing is not considered.
- The delivery base has a fixed number of vehicles.
- All vehicles are equal and have inventory capacity.
- A vehicle passes through a maximum of one route per period, and all vehicles depart from the base and return during that period.

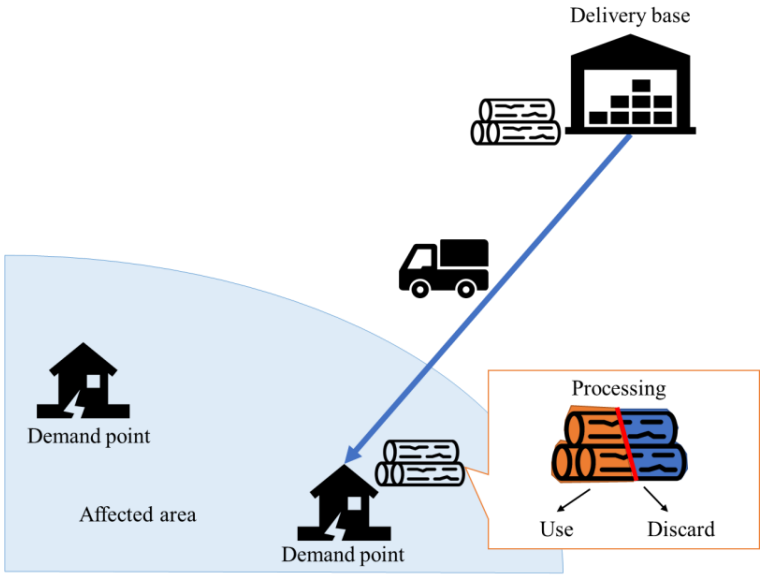
Model description

This section presents IRPSPDU and IRPTSPDU model. The preconditions are as follows.

IRPSPDU model

This model consists of one delivery base and multiple demand points (Fig.1).

Figure 1: IRPSPUD model

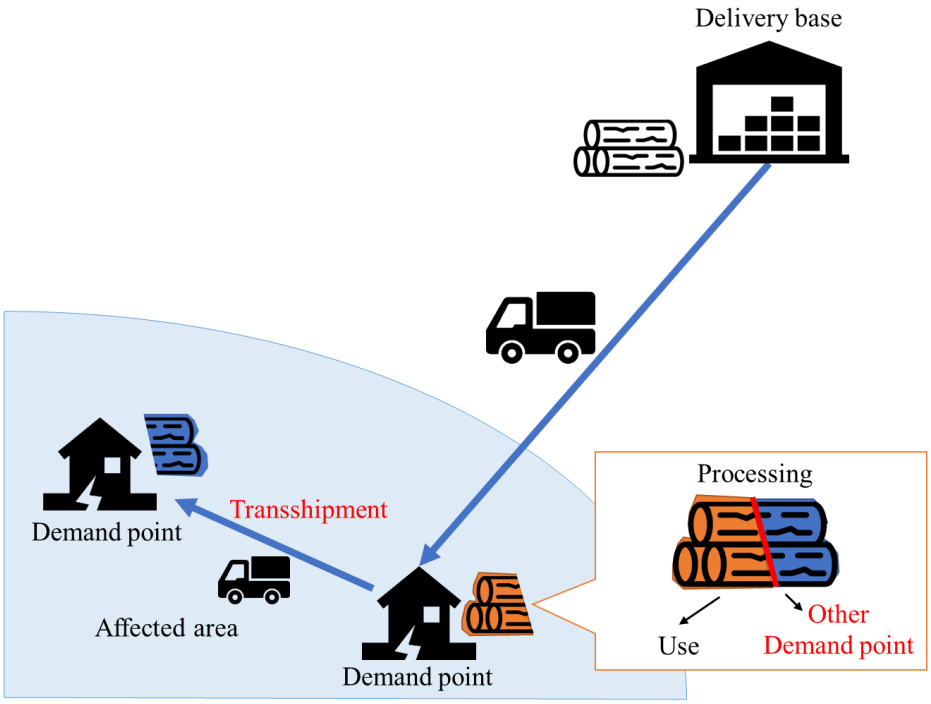


In this model, building members are delivered directly from the delivery base to the demand points. The minimum delivery unite constraint for building members does not change. Also, building members are processed at demand points. Surplus occurs at each demand point. Surplus generated at each point is discarded at the end of the planning period.

IRPTSPDU model

This model is composed of one delivery base and multiple demand points, and performs delivery considering planned transshipment (Fig.2).

Figure 2: IRPTSPUD model



This model considers inventory management, delivery planning and planned transshipment simultaneously. Building members are processed at the demand points, as in the first model. However, in this model, the surplus generated at each demand point can be transported to other demand points by using planned transshipment. This is expected to not only reduce building members waste but also minimizing delays in the delivery of building materials. In addition, it is expected that the ability to respond to the uncertainty of the scale of demand will be improved.

FORMULATION

This section formulates models shown in the previous section. These models are formulated using SP with reference to IRP and IRPT.

IRPSPDU model

First, the IRPSPDU model is formulated.

Minimize

$$E_{\Xi}\left[\begin{matrix}\left(\sum_{t\in T}\sum_{i\in V^c} Max\left(-I_{t,i}(\xi),0\right)\times\alpha + Max\left(I_{t,i}(\xi),0\right)\times\beta\right)\\ +\sum_{i\in V}\sum_{j\in V}\sum_{t\in T}\sum_{k\in K} r_{i,j}\times x_{t,i,j,k}(\xi)\times\gamma\end{matrix}\right] \quad (1)$$

Subject to

$$I_{t,i}(\xi) - I_{t-1,i}(\xi) = \sum_{k\in K} Q_{t,i,k}(\xi) - d_{t,i}(\xi)\,, \forall t \in T, \forall i \in V^c \quad (2)$$

$$I_{t,i}(\xi) \leq C_i \ \forall t \in T, \forall i \in V^c \quad (3)$$

$$Q_{t,i,k}(\xi) \leq y_{t,i,k}(\xi) \times C_i, \forall t \in T, \forall i \in V^c, \forall k \in K \quad (4)$$

$$\sum_{i \in V^c} Q_{t,i,k}(\xi) \leq yy_{t,i}(\xi) \times Z, \forall t \in T, \forall k \in K \quad (5)$$

$$Q_{t,i,k}(\xi) \equiv Lot \ (mod \ Lot), \forall t \in T, \forall i \in V^c, \forall k \in K \quad (6)$$

$$QC_{t,0,k}(\xi) = \sum_{i \in V^c} Q_{t,i,k}(\xi), \forall t \in T, \forall k \in K \quad (7)$$

$$QC_{t,i,k}(\xi) = \sum_{n \in V} x_{t,n,i,k}(\xi) \times QC_{t,n,k}(\xi) - Q_{t,i,k}(\xi) \ \ \forall t \in T, \ \forall i \in V^c, \ \forall k \in K \quad (8)$$

$$Z \geq QC_{t,0,k}(\xi) \geq 0, \forall t \in T, \forall i \in V^c, \forall k \in K \quad (9)$$

$$\sum_{i \in V^c} \sum_{k \in K} Q_{t,i,k}(\xi) \leq C_{base}, \forall t \in T \quad (10)$$

$$\sum_{i \in V^c} x_{t,i,j,k}(\xi) = \sum_{i \in V^c} x_{t,j,i,k}(\xi) = \mathrm{y}_{t,j,k}(\xi) \quad \forall t \in T, \forall j \in V^c, \forall k \in K \quad (11)$$

$$\sum_{i \in V^c} x_{t,0,i,k}(\xi) = \sum_{i \in V^c} x_{t,i,0,k}(\xi) = yy_{t,k}(\xi) \ \ \forall t \in T, \forall k \in K \quad (12)$$

$$y_{t,i,k}(\xi) \leq yy_{t,k}(\xi), \forall t \in T, \forall i \in V^c, \forall k \in K \quad (13)$$

$$\sum_{i \in V} \sum_{j \in V} x_{t,i,j,k}(\xi) = 0, \forall t \in T, \forall k \in K, i = j \quad (14)$$

$$\sum_{i \in S} \sum_{j \in S} x_{t,i,j,k}(\xi) \leq |S| - 1, \forall S \subset V, \forall t \in T, \forall k \in K, (S \in \{1, \dots, n\}, 2 \leq |S| \leq n - 1) \quad (15)$$

$$x_{t,i,j,k}(\xi), y_{t,i,k}(\xi), yy_{t,k}(\xi) \in 0,1, \forall t \in T, \forall i \in V, \forall j \in V, \forall k \in K \quad (16)$$

$$Q_{t,i,k}(\xi) \geq 0, \forall t \in T, \forall i \in V, \forall j \in V, \forall k \in K \quad (17)$$

The IRP model minimize delay penalty and required costs. The objective function (1) minimizes delays in the delivery of building materials length. Constraint (2) ensures that, the left side shows the difference in inventory between the end of period *t* and the end of period *t*-1, and the right side shows the amount of delivery and demand in period *t*. Constraint (3) ensures that, during each time period and for each demand point, the inventory level complies with the inventory capacity. Constraint (4) ensures the amount of product delivered to demand point *i* by vehicle *k* in period *t*. Constraint (5) ensures that vehicle *k* complies with its capacity at each time period *t*. Constraint (6) ensures a minimum delivery unit constraint for building members delivered from the delivery base. The delivery amount to each demand point is an integral multiple of the minimum delivery unit constraint. Constraint (7) confirms the inventory level when the vehicle leaves the delivery base. Constraints (8) and (9) ensure that the inventory level in the vehicle follows the loading constraint when the vehicle leaves each demand point. Constraint (10) ensures that the building members delivered from the delivery base in each period do not exceed the capacity of the delivery base. Constraint (11) ensures the conservation of vehicle flow for each demand point, so that every

vehicle *k* arriving at a node *i* coming from any demand point *j* must leave the demand point *j* going to any node *i*. Constraint (12) ensures that only vehicles carrying a building member can leave the delivery base and that every vehicle must return to the delivery base. Constraint (13) ensures that vehicles not leaving the delivery base will not visit any demand point *i*. Constraint (14) prevents the existence of arcs (*i*, *j*), where *i* equals *j*. Constraint (15) prevents the occurrence of sub routes. Finally, (16) and (17) define the domain of the respective variables.

IRPTSPDU model

Next, the IRPTSPDU model is formulated. This model is formulated based on the IRPSPDU model. To consider planned transshipment, a parameter indicating a decision variable indicating the transshipment amount are added.

Decision variables

$w_{t,i,j,k}$ Transshipment amount by vehicle *k* from demand point *i* to demand point *j* in period *t*

Constraint (2) is replaced as shown in (18).

$$I_{t,i}(\xi) + d_{t,i}(\xi) + \sum_{j\in V^c}\sum_{k\in K} w_{t,i,j,k}(\xi) = I_{t-1,i}(\xi) + \sum_{k\in K} Q_{t,i,k}(\xi) + \sum_{j\in V^c}\sum_{k\in K} w_{t,j,i,k}(\xi), \forall t \in T, \forall i \in V^c \quad (18)$$

Constraint (18) represents an inventory stock formula in addition to inventory amount by transshipment, and (19), (20), (21), and (22) are added.

$$w_{t,i,j,k}(\xi) \le y_{t,i,k}(\xi) \times y_{t,j,k}(\xi) \times I_{t-1,i}(\xi), \forall t \in T, \forall i \in V^c, \forall k \in K \quad (19)$$

$$w_{t,i,j,k}(\xi) \le y_{t,i,k}(\xi) \times y_{t,j,k}(\xi) \times C_j, \forall t \in T, \forall i \in V^c, \forall k \in K \quad (20)$$

$$w_{t,i,j,k}(\xi) \ge 0, \forall t \in T, \forall i \in V^c, \forall j \in V^c, \forall k \in K \quad (21)$$

$$w_{t,i,j,k}(\xi) \equiv Lot'(mod\ Lot'), \forall t \in T, \forall i \in V^c, \forall j \in V^c, \forall k \in K \quad (22)$$

$$QC_{t,i,k}(\xi) = \sum_{j\in V} x_{t,j,i,k}(\xi) \times QC_{t,n,k}(\xi) - Q_{t,i,k}(\xi) + \sum_{j\in V^c}(w_{t,i,j,k}(\xi) - w_{t,j,i,k}(\xi)), \forall t \in T, \forall i \in V^c, \forall k \in K \quad (23)$$

Constraints (19) and (20) ensure that the constraints on the amount of inventory and that of transshipment owing to vehicle loading constraints, respectively, are upheld. Constraint (21) sets the area of the variable. Constraint (22) shows that the minimum delivery unit changes when processed at the demand point. Constraint (23) ensure that the inventory level in the vehicle conforms to the loading constraint when the vehicle leaves each demand point. These changes made it possible to transfer the surplus building members at each demand point to other demand points. In the transfer between demand points, the minimum delivery unit is smaller than the delivery from the delivery base.

SCENARIO ANALYSIS

In this section, the effects of the three models are investigated using numerical experiments. In this research, numerical example were conducted using 5 case studies with 20 demand scale scenarios ($\boldsymbol{\xi_{max} = 20}$) as one case study.

Numerical example

The following numerical example is used to investigate the effect of the proposed model. There is one delivery base, one processing base, four demand points (*n* is 4), a single building member, k_{max} is 2, T_{max} is 6, *Lot* is 5 and $Lot'$is 1. In addition, α, β, and γ, which represent the penalty cost, are 100, 10, and 0.2. The penalty for late delivery is the largest and the cost for delivery distance is the smallest. Limit on inventory of the base is 40, limit on inventory of the vehicles is 30, and limit on inventory of each demand point is 20. The coordination of the delivery base and demand points are shown in Table 1.

Table 1: Location coordination

	Nodes	X	Y
Delivery base	0	139.06	36.39
Demand point	1	139.64	35.45
	2	139.65	35.85
	3	140.12	35.60
	4	139.88	36.57

The distance between each base and demand point was calculated as a straight line. The dataset for this node is clustered in problem sets.

The given demand scale scenario is set as follows using normal distribution N (average μ, standard deviation σ).

- Demand for each demand point is known throughout the period and is set as follows. 1: Demand for N(28, 5) in period 1; 2: Demand for N(28, 5) in period 2; 3: Demand for N(35, 5) in period 3; 4: Demand for N(44, 5) in period 4.
- Twenty scenarios per one case study are generated according to the above normal distribution. Five case studies are created.

Calculations were made using a personal computer (OS: Windows 7 professional 32 bit) with Intel® Core ™ 2 Duo CPU E8600 @ 3.33 GHz CPU and 4 GB RAM on Python 2.7.15. The mathematical planning solver used was Gurobi Optimizer Version 6.0.3. The results of the numerical experiments conducted under the above environment are shown below.

Comparing model effects

The IRPSPDU model and the IRPTSPDU model were compared for the object value (Table 2).

Table 2: COMPATION OF MODEL EFFECTS

Case study	Object value	
	IRPSPDU model	IRPTSPDU model
1	2338.79	2309.57
2	3111.96	2845.73

3	2485.85	2350.44
4	2478.44	2478.44
5	2402.19	2299.83
Average	2563.45	2456.80

In all case studies, the IRPTSPDU model led to better objectives than the IRPSPDU model. Case 4 showed the same results for both models, indicating that there is no reduction effect due to planned transshipment. In this way, the IRPTSPDU model does not lead to worse results than the IRPSPDU model because it utilizes planned transshipment as needed. Since the reduction rate changes depending on the case, it can be seen that the effect of planned transshipment is affected by the demand scale scenario. However, since the planned transshipment achieved an average reduction of 3.9%, it can be said that the IRPTSPDU model is effective in formulating a plan aimed at reducing delivery delays while suppressing costs.

Table 3: COMPATION OF MODEL DETAIL EFFECTS

Case study	Delay amount		Excess inventory		Distance	
	IRPSPDU	IRPTSPDU	IRPSPDU	IRPTSPDU	IRPSPDU	IRPTSPDU
1	4.96	5.83	132.44	123.30	398.40	374.40
2	17.83	15.11	85.31	92.58	355.60	288.60
3	9.74	4.46	107.21	141.94	320.40	365.40
4	4.71	4.71	152.18	152.18	366.00	366.00
5	5.42	3.89	142.90	141.37	310.60	376.60
Average	8.53	6.80	124.01	130.27	350.20	354.20

Next, the two models were compared for the elements of delay amount, excess inventory, and distace. The IRPTSPDU model showed a great effect in reducing the delay amount, and on the contrary, excess inventory and delivery distance tended to increase. It is considered that this tendency changes when the cost parameters α, β, and γ are changed. As a result, it can be seen that the IRPTSPDU model contributes to the reduction of the objective value by balancing the three elements better than the IRPSPDU model. In other words, the IRPTSPDU model can propose a delivery plan that matches the weight of planners.

CONCLUTION

This research propose a delivery model of building materials for the purpose of early restoration of damaged buildings. The proposed model considers two uncertainties. One is the uncertainty of the scale of demand, and the other is the uncertainty of the processing status of building members. The former is formulated by modeling using stochastic programming with a finite number of scenarios. The latter was incorporated as a changing constraint. Furthermore, it is shown that the IRPTSPDU model using planned transshipment in the delivery plan is effective in improving the objective value. The planned transshipment as a collaborative strategy did not have the effect of reducing all the terms of the objective function at the same time, but showed the effect of improving their balance. However, due to the influence of the amount of calculation, the scenario analysis was only able to solve small-scale problems. Future prospects include proposals of approximate solutions for solving larger problems and extensions to two-stage stochastic programming.

REFERENCES

Anaya-Arenas, A. M., Renaud, J., & Ruiz, A. (2014). Relief distribution networks: a systematic review. Annals of Operations Research, 223(1), 53-79.

Balcik, B., Beamon, B. M., & Smilowitz, K. (2008). Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems*, *12*(2), 51-63.

Bell, W. J., Dalberto, L. M., Fisher, M. L., Greenfield, A. J., Jaikumar, R., Kedia, P., Mack, R. G., & Prutzman, P. J. (1983). Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer. *Interfaces*, *13*(6), 4-23.

Caunhye, A. M., Zhang, Y., Li, M., & Nie, X. (2016). A location-routing model for prepositioning and distributing emergency supplies. *Transportation research part E: logistics and transportation review*, *90*, 161-176.

Coelho, L. C., Cordeau, J. F., & Laporte, G. (2014). Thirty years of inventory routing. *Transportation Science*, *48*(1), 1-19.

Kazama, M., & Noda, T. (2012). Damage statistics (Summary of the 2011 off the Pacific Coast of Tohoku Earthquake damage). Soils and Foundations, 52(5), 780-792.

Mete, H. O., & Zabinsky, Z. B. (2010). Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics*, *126*(1), 76-84.

Özdamar, L., & Ertem, M. A. (2015). Models, solutions and enabling technologies in humanitarian logistics. *European Journal of Operational Research*, *244*(1), 55-65.

Peres, I. T., Repolho, H. M., Martinelli, R., & Monteiro, N. J. (2017). Optimization in inventory-routing problem with planned transshipment: A case study in the retail industry. *International Journal of Production Economics*, *193*, 748-756.

Yi, W., & Özdamar, L. (2007). A dynamic logistics coordination model for evacuation and support in disaster response activities. *European journal of operational research*, *179*(3), 1177-1193.

Yuan, Y., & Wang, D. (2009). Path selection model and algorithm for emergency logistics management. Computers & industrial engineering, 56(3), 1081-1094.

DECISION SCIENCES INSTITUTE
Do Different Industries Have Different Sustainability Practices and Impacts?

Yun-Chen Morgan
Department of Management and Business Administration
Southeastern Louisiana University
Hammond, Louisiana 70402
ymorgan@selu.edu

Susan M. L. Zee
Department of Marketing and Supply Chain Management
Southeastern Louisiana University
Hammond, Louisiana 70402
susan.zee@selu.edu

Lillian Y. Fok
Department of Decision Sciences
Western Washington University
Bellingham, Washington 98225
fokl@wwu.edu

ABSTRACT

This research analyzes the relationships among green sustainability orientations, organizational culture, Total Quality Management (TQM), organizational success, and the impact of green initiatives across three industries: manufacturing, retail, and healthcare. Surveys from 331 full-time employees from these three industries were evaluated. Results showed that healthcare organizations embrace more team-orientated cultures and are more participative and proactive than those in manufacturing and retail. Manufacturing organizations use more advanced TQM tools than those in healthcare and retail. Results also showed that a continuous improvement culture is significantly related to the impact of green initiatives in all three industries.

KEYWORDS: Green Movement, Sustainability, Total Quality Management, Organizational Culture

INTRODUCTION

The need to “go green” has been sweeping through all industries in the United States for quite some time. In this research, we examine three of them – manufacturing, retail, and healthcare – to gain understanding of what different environmental and organizational factors may affect the impact brought on by the sustainability movement. We explore each industry individually and also draw comparisons from them to seek understanding of what goes on within the process of balancing organizational success and the desire to address environmental issues. Due to the nature of the differences in these three industries, we expect to find that impact from the sustainability movement is going to be attributed from different organizational factors in each industry. These factors include quality management methods, organizational cultures, and general organizational environmental initiatives.

LITERATURE REVIEW

Total Quality Management (TQM)

An integrative approach to pursuit customer satisfaction is the idea behind Total Quality Management (TQM) (Chin, Pun, & Hua, 2001). The Manufacturing industry, the retail industry, and the healthcare industry, all are facing intense pressure to incorporate sustainability into organizational strategies. TQM is part of the strategy that could ensure companies meeting adequate social and economic performance (Khan & Asim, 2020; Nahas & Chandrasekar, 2019). We suggest that maturity in quality management related issues is important in understanding the impact of going green based on Fok, Hartman, Patti, & Razek's study (2000).

Organizational Culture

Having a corporate culture that is in sync with corporate objectives is important to organizational success. Values and attitudes are part of this culture and in this research, we examine how culture relate to the impact of green initiatives in different industries. Manufacturing performance measure and culture (Seow, Hamid, & Yusof, 2017) and the impact of environmental knowledge and product quality and its relationship to organizational culture (Sun, Teh, & Linton, 2018) have been found to be important factors to organizational success. Research also suggest the role of top management and organizational culture is vital in engaging commitment for environmental performance across industries (Dubey, Gunasekaran, Helo, Papadopoulos, Childe, & Sahay, 2017).

RESEARCH MODEL

In this research, we examine sustainability movement issues to consider organizational culture, employee perceptions of organizational commitment to the sustainability movement and the relationships which may exist between organizational commitment and the QM Maturity of the organization. Additionally, where there is a higher level of perceived commitment to the sustainability movement and where more mature QM systems are in place, we expect that, in the overall, the organization itself will be seen as "doing better" and the impact of the sustainability movement will be perceived as favorable.

> **Research Question 1**: Organizations in different industries will have different levels of organizational green orientation, organizational culture, QM maturity, organizational performance, and report differing impacts of the sustainability movement.
>
> **Research Question 2:** Organizations in different industries will have different relationships among Organizational Green Orientation, Organizational Culture, and QM Maturity towards the impact of the sustainability movement.

METHODOLOGY

Subjects in the sample were approximately 331 full-time employees from a wide variety of industries in the South. The subjects were roughly 51.7 % male and 48% female with an average age of 41.49. Among these 331 subjects, there are approximately nine categories of

industries in our survey instrument. The three largest percentages of the subjects work in retail, healthcare, and manufacturing industries with 22.1%, 20.8%, and 12.7%, respectively.

Subjects responded to a survey which asked about their perceptions and experiences regarding the sustainability movement, quality management, and organizational culture in firms of which they are employees. In this study, we focus on three industry types – manufacturing, retail, and healthcare – and compare their relationships among green orientation and organizational culture, impact of sustainability movement, QM maturity and organizational performance.

DISCUSSION AND CONCLUSIONS

The importance for all businesses to find ways to balance the need to compete and social responsibility in going green is apparent. Our study finds that manufacturing, retail, and healthcare organizations have different levels of green orientation, organizational culture, and organizational performance. These factors lead to different levels of impact from the sustainability movement. We found that healthcare organizations are more team-oriented, and are more likely to exhibit organizational culture that promote participation and proactivity, when compared to organizations in the manufacturing and retail industries. This could be explained by the nature of the tasks within all healthcare facilities and that there is an organic need for individuals within a healthcare facility to be work well with one another. We also found that manufacturing organizations use more advanced quality management tools compared to organizations in retail or healthcare. This could be explained by the need for manufacturing facilities to adopt quality management tools beyond the traditional ones, due to more intense competition and rigid organizational structures.

In addition, we found that continuous improvement culture has significant relationships with sustainability movement impact in all three industries. The use of basic TQM tools is significantly related to the sustainability movement impact only in the retail industry. Organizational success has significant relationships with the sustainability movement impact in manufacturing and retail industries, and green orientation is significantly related to the sustainability movement impact only in the manufacturing industry. Our results suggest that organizational culture is integral in all three of the industries in that cooperation and teams contribute to better business-client relationships, better working environments, and higher organizational reputations. Less complex quality management tools tend to be more effective in a retail environment perhaps due to the attributes that this industry tend to have more traditional views. It is a logical implication that generally positive views of organizational performance and higher morale lead to higher perceptions of impact. Since manufacturing facilities tend to be more rigid and follow routines and protocols, it can explain why this is the only industries showing a clear connection between what organizations are doing as far as their initiatives to be green, and the perceptions of those initiatives to make an impact.

REFERENCES

Reference is available via email to the author.

DECISION SCIENCES INSTITUTE
An Integrated K-means-GP Approach for US Stock Fund Diversification

Dinesh K. Sharma
University of Maryland Eastern Shore
Email: dksharma@umes.edu

H.S. Hota
Guru Ghasidas Central University
Email: proffhota@gmail.com

Vineet Kumar Awasthi
Dr. C.V. Raman University
Email: vineet99kumar@gmail.com

ABSTRACT

The stock fund diversification process is a tedious task due to the erratic nature of the stock market. On the other hand, this process has become more demanding due to the expectation of high annual returns with low risk. This research work explores the potential use of Goal Programming (GP) and K-means Algorithm as an integrated K-means-GP approach for fund diversification, where K-means is used to create stock-based groups on their performance. Then GP is used to diversify the total fund into various stocks to achieve a high annual return. The experimental work has been done on stocks from the DOW30 index. All 30 stocks are divided into different groups using K-means, and then the total fund is diversified into the groups as per their performance with expected annual returns.

KEYWORDS: K-means, Goal Programming (GP), DOW30, Fund diversification.

INTRODUCTION

Diversification of funds to be invested in the stock market is the most challenging task for the fund managers as well as for the investors. It is also challenging to handle the risk and return as per the expectations. The primary objective of this piece of research is to achieve a high return with manageable risk. There are various risks associated with fund diversification, and as a result, it can be complicated to determine how it can be allocated to achieve the best return with minimum risk. An expected value variance model for portfolio optimization was introduced by Markowitz (1952), based on two factors: risk and return (Tamiz et al., 2013). These two factors are common to all the investors, but some common factors heavily impact the assets. Identifying these factors is very difficult and challenging; therefore, a goal should be determined and focused on identifying the most critical fund diversification factors. A fund diversification problem can be viewed as a Goal Programming (GP) model with two objectives: risk and return. GP was first proposed by Charnes et al. (1955) and Charnes et al. (1961) and further development by several authors (Ijiri, 1965; Lee, 1972; Ignizio, 1976; Romero, 1991). This model's premise is to determine an objective function's aspiration levels and minimize any (positive or negative) deviations from these levels (Aouni et al., 2013).

On the other hand, clustering methods like K-means were also applied to cluster stocks and indexes. Kumar et al. (2018) used the K-means clustering algorithm by applying Weka interface

to recognize academic performance and enhance the educational quality of self-financing arts and science college students. The K-means algorithm is also applied in different domains like image segmentation (Qureshi et al., 2018), candidate system (Karimov et al., 2015), center selection (Likas et al., 2003), etc.

To the best of our knowledge, K-means and GP methods were not combined by any authors, at least in the domain of stock fund diversification. Furthermore, it is assumed to be the right combination of these two techniques: K-means and GP may improve the result and produce a high return with low risk. The goal and constraints used in the research work of Hota et al. (2017) are also used in this research work. The main idea of an integrated approach is to find the 'Best' stocks that perform well in terms of annual return and then apply more funds as compared to others. An integrated approach of both K-means and GP as K-means-GP has achieved the expected return level with minimum risk in terms of beta value and standard deviation.

LITERATURE REVIEW

Several researchers used GP as an optimization technique for selecting objects in the various problem domains. Some authors also used the K-means clustering technique for clustering stocks, but no paper found that has applied the integrated technique of K-means and GP for the stock selection. Some of the research works are listed in Table 1, reflecting GP and K-means importance in different problem domains.

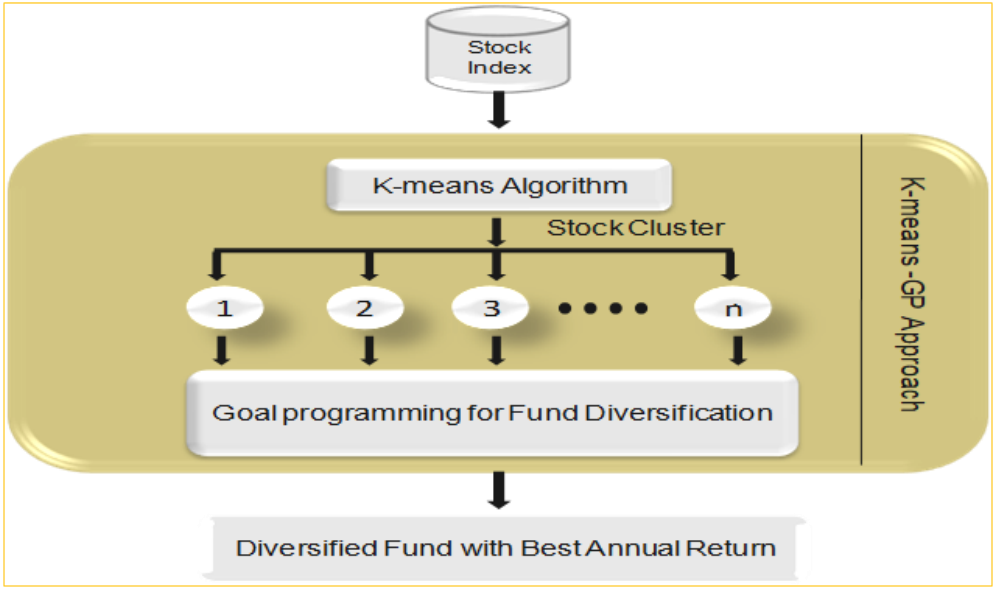
Table 1: List of Research Contribution in the Field of Stock Fund Diversification

Year	Authors	Journal
2010	Nanda	Expert Systems with Applications
2013	Ghahtaraniand Najafi	Economic Modelling
2013	Aouni et al.	European Journal of Operational Research
2013	Tamiz et al.	European Journal of Operational Research
2013	Sharma et al.	Journal of Money, Investment and Banking
2014	Belaid et al.	European Journal of Operational Research
2014	Choudhary et al.	Computers & Industrial Engineering
2014	Jadidi et al.	Int. Journal Production Economics
2015	Halima et al.	Procedia Social and Behavioral Sciences
2015	Jadidi et al	Applied Mathematical Modelling
2015	Momeni et al.	Arabian Journal of Business and Management Review
2017	Hota et al.	Review of Business and Technology Research
2018	Qureshi and Ahamad	Procedia Computer Science
2018	Kumar and Arulanandam	Int. Journal of Computer Sciences and Engineering
2019	Rao et al.	IET Intelligent Transport Systems
2019	Ajayi-Daniels	Int. Journal of Management and Fuzzy Systems
2019	Qahtani et al.	Review at Advances in Operations Research

INTEGRATED K-MEANS-GP BASED FUND DIVERSIFICATION

The proposed integrated system workflow is presented in Figure 1, in which input is the stock index. The purpose of each step is explained as follows.

Figure 1: Integrated K-means-GP Approach



Step 1: In this step, the application of the K-means algorithm is used to create different stocks clusters. MacQueen (1967) introduced K-means, one of the most popular clustering algorithms used in various domains to segregate 'x' observations into 'n' numbers of predefined clusters using Euclidean distance. The algorithm is summarized as follows:

Step 2: This step uses a GP for fund diversification among different groups of stocks obtained from step 1. The fund was diversified in different stocks based on different constraints to achieve a high return and low risk. The GP technique is basically developed to handle multi-objective criteria situations with conflicting criteria. This technique's primary goal is to find the "best possible" solution for achieving the goals with the problem's given constraints. A general format of the GP (Sharma et al., 2005, 2013) model is as follow:

$$\text{Minimize}[P_1(\bar{d}), \ldots., P_k(\bar{d}), \ldots\ldots, P_K(\bar{d})] \quad (1)$$

Subject to,

$$f_i(Z) + \mathrm{d}_i^- - \mathrm{d}_i^+ = \mathrm{b}_i \qquad , \mathrm{i} = 1,2,3 \cdots\cdots \mathrm{m} \quad (2)$$

$$\mathrm{d}_i^-,\ \mathrm{d}_i^+,\ Z >= 0 \text{ and } \mathrm{d}_i^-.\ \mathrm{d}_i^+ = 0 \quad (3)$$

Where $P_k(\bar{d})$ = $P_k(w_{ik}^- d_{ik}^- + w_{ik}^+ d_{ik}^+)$ and P_k is the k^{th} priority structure, w_{ik}^- and w_{ik}^+ are the numerical weights associated with the deviational variables d_{ik}^- and d_{ik}^+ respectively at the priority level P_k, Z is the vector of decision variables, $f_i(Z)$ is the function of i^{th} goal constraint, d_i^- and d_i^+ are i^{th} negative and positive deviation variables respectively.

The goals and constraints are set by the decision-maker to determine a feasible solution. The main idea is to minimize a weighted sum of deviations from the target goal. As decided by Hota et al. (2017), the GP's goals and constraints are considered for stock portfolio management to diversify funds.

The following notifications are used to design the GP model (Sharma et al., 2013):

The Goals

The goals of the decision-makers include:
(1) Utilization of total available funds for investments (2) Maximization of the portfolio's expected annual return, and (3) Minimization of the portfolio's risk.

The Constraint:

If m is the number of stocks, F is the total available fund, and Z_i denotes allocated fund in i^{th} stock, then various constraints can be formulated as follows:

(i) Investments: Decision-maker wants to invest the maximum amount of funds into all stocks (m) of a portfolio, then the constraint for the investment of total fund F is:

$$\sum_{i=1}^{m} Z_i + d^{1-} - d^{1+} = F \quad (4)$$

(ii) Annual return: The main objective of portfolio management is to maximize the total return from all the stocks from the new stock portfolio. If A_i denotes the annual return of i^{th} stock then expected rate of return A is defined as:

$$\sum_{i=1}^{m} A_i Z_i + d^{2-} - d^{2+} = A \quad (5)$$

(iii) Portfolio's risk:

The composite Beta (B) of the stock portfolio can be written as:

$$\sum_{i=1}^{m} B_i Z_i + d^{3-} - d^{3+} = Ba \quad (6)$$

Where B_i is the amount of risk with the i^{th} stock and B_a is the acceptable Beta value.

Standard Deviation:

The composite standard deviation (σ) is referred to as the risk and should be minimized. The equation for the standard deviation is:

$$\sum_{i=1}^{m} \sigma_i Z_i + d^{4-} - d^{4+} = \sigma_a * F \quad (7)$$

Where σ_i is the measure of nonsystematic risk associated with i^{th} stock and σ_a is an acceptable standard deviation level.

(iv) Limits on Investment in all Stocks:

Maximum Limit: The percentage maximum investment limit (max_p) in a single stock can be written as follows:

$$Z_i + d_5^- - d_5^+ = max_p * F \quad , i = 1,2,3 \cdots\cdots m \quad (8)$$

Minimum Limit: The percentage minimum investment limit (min_p) in a single stock can be written as follows:

$$Z_i + d_6^- - d_6^+ = min_p * F \quad , i = 1,2,3 \cdots\cdots m \quad (9)$$

EXPERIMENTAL STUDY

An experiment was done using MATLAB and LINGO software for the simulation of K-means and GP, respectively. The K-means algorithm was applied first to find out the cluster of stokes for the research, and then GP was applied to diversify the fund in various stocks of obtained clusters with an expected rate of return and risk. As stated above, the entire process of clustering and fund diversification is described as follows:

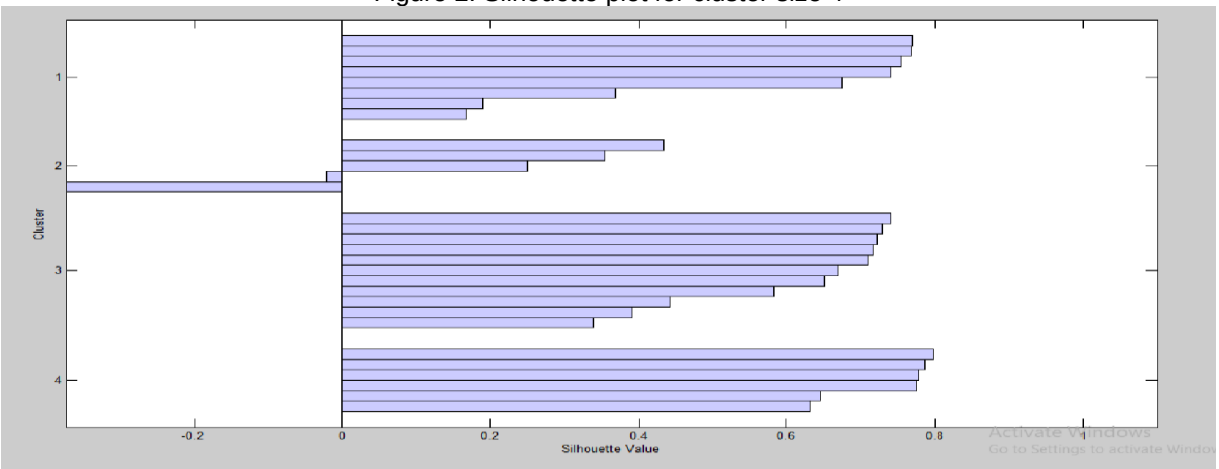
Input: The fund diversification is done into the stock data of DOW30 in 2017-18. DOW30 is a US stock index, having 30 large, public-traded companies presented in Table 2. Stock data is downloaded from http//in.finance.yahoo.com with four criteria; High, Low, Close, and Open. These values were used to derive new useful variables like Beta value, Standard deviation, and yearly return.

Table 2: DOW30 Stock Index

S.No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ticker	AXP	BA	CAT	CSCO	CVX	DIS	DOW	GS	HD	IBM	INTC	JNJ	JPM	KO	MCD
S.No.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Ticker	MMM	MRK	MSFT	NKE	PFE	PG	TRV	UNH	UTX	V	VZ	WBA	WMT	XOM	AAPL

Step 1: In this step, stock data of DOW30 is feeded to the K-means algorithm using MATLAB for creating groups of stocks. The variables Beta value (C_1), Standard deviation (C_2) and annual return (C_3) are considered as criteria for applying the K-means algorithm. Different cluster sizes (3, 4 and 5) are tested to find the minimum value of the cost function. The minimum cost function (5393) is found for cluster size four. A silhouette plot showing four clusters created by the K-means algorithm with 8,5,11 and 6 similar stocks in each cluster is presented in Figure 2.

Figure 2: Silhouette plot for cluster size 4



Based on the average annual return of stocks of each cluster, four groups of stocks were made and named as Best, Better, Good, and Worst as below:

Group 1: Best- average annual return= 51.43%
Group 2: Better- average annual return= 21.51%
Group 3: Good- average annual return= 4.27%
Group 4: Worst- average annual return= -13.26%

The cluster with highest annual return is named as Best; similarly, the cluster showing the lowest annual return is named as Worst.

Step 2 : In this step GP was applied for fund diversification among four groups of stocks obtained from step 1. The objective function and constraints for GP are formulated as per equations 1 to 9. Minimum beta vale and standard deviation, set by the decision maker are 0.8882 and 0.0214 (2.14%) respectively. In order to achieve optimum return with low risk, GP was run with various combinations of maximum and minimum fund allocation in four groups and finally maximum fund allocation in Best, Better, Good, Worst groups are considered as 10%, 5%, 3% and 2% respectively. Similarly minimum fund allocation in Best, Better, Good, Worst groups are considered as 5%, 2%, 1.5% and 1% respectively.

RESULT ANALYSIS AND COMPARATIVE STUDY

Outcomes from the proposed integrated K-means-GP approach were studied using three different cases as per expected return value of investor, as listed below:

Case I: If expected return is 20% of the total investment from the investor's choice then fund diversification of total fund done by GP is shown in Table 3. GP allocated 100% of the funds with objective function = 0.1566, calculated annual return = 0.211(21.2%), beta value = 1.03 and standard deviation = 0.014 (1.4%).

Table 3: Fund Diversification for Expected Annual Return 20%

Stock	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}
Fund(%)	0.05	0.05	0.05	0.05	0.08	0.02	0.02	0.02	0.01	0.02	0.05	0.02	0.05	0.05	0.05
Stock	Z_{16}	Z_{17}	Z_{18}	Z_{19}	Z_{20}	Z_{21}	Z_{22}	Z_{23}	Z_{24}	Z_{25}	Z_{26}	Z_{27}	Z_{28}	Z_{29}	Z_{30}
Fund(%)	0.05	0.03	0.03	0.015	0.03	0.03	0.03	0.03	0.05	0.02	0.02	0.02	0.02	0.02	0.02

Case II: If expected return is 25% of the total investment from the investor's choice then fund diversification of total fund done by GP is shown in Table 4. GP invested 100% fund with objective function value = 0.1771, calculated annual return = 0.25 (25%), beta value = 1.083 and standard deviation=0.014 (1.4%).

Table 4: Fund Diversification for Expected Annual Return 25%

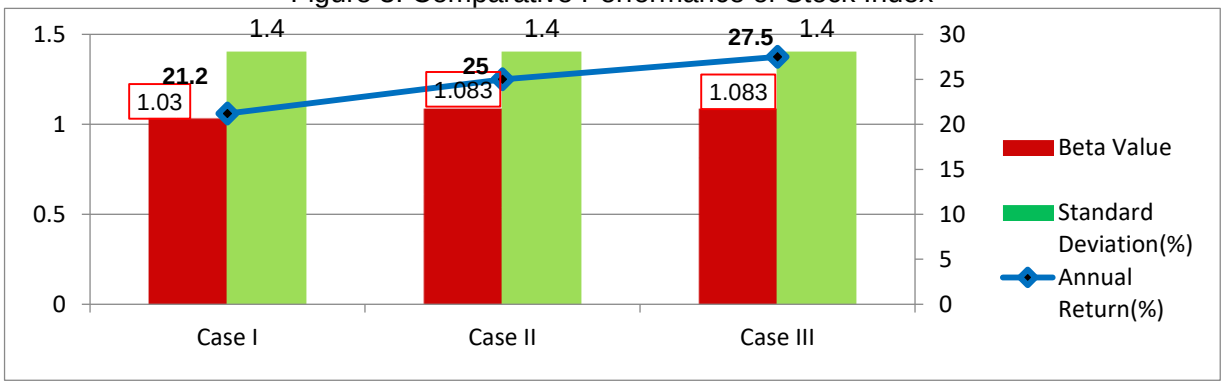
Stock	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}
Fund(%)	0.096	0.05	0.05	0.05	0.08	0.02	0.02	0.02	0.02	0.05	0.02	0.05	0.05	0.05	0.05
Stock	Z_{16}	Z_{17}	Z_{18}	Z_{19}	Z_{20}	Z_{21}	Z_{22}	Z_{23}	Z_{24}	Z_{25}	Z_{26}	Z_{27}	Z_{28}	Z_{29}	Z_{30}
Fund(%)	0.02	0.015	0.015	0.015	0.03	0.03	0.03	0.028	0.03	0.02	0.01	0.02	0.02	0.02	0.02

Case III: If expected return is 30% of the total investment from the investor's choice, then fund diversification of the total fund done by GP is shown in Table 5. GP invested 100% fund with objective function value = 0.2209, calculated annual return = 0.2751 (27.51%), beta value = 1.083 and standard deviation = 0.014 (1.4%). In this case, GP lacks a return with 2.5% and achieved a 27.51% return with the minimum risk level.

Table 5: Fund Diversification for Expected Annual Return 30%

Stock	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}
Fund(%)	0.1	0.05	0.1	0.05	0.08	0.02	0.02	0.02	0.02	0.02	0.05	0.02	0.05	0.05	0.05
Stock	Z_{16}	Z_{17}	Z_{18}	Z_{19}	Z_{20}	Z_{21}	Z_{22}	Z_{23}	Z_{24}	Z_{25}	Z_{26}	Z_{27}	Z_{28}	Z_{29}	Z_{30}
Fund(%)	0.05	0.015	0.03	0.015	0.15	0.03	0.015	0.015	0.025	0.01	0.01	0.02	0.02	0.02	0.01

Figure 3: Comparative Performance of Stock Index



The performance of the integrated K-means-GP-based expert portfolio system for all three cases for the data of 2017-18 is shown graphically in Figure 3. This figure reflects that integrated K-means-GP based system can achieve the highest return up to 27.51% against 30% of the expected return with a minimum value of risks (Beta and Standard Deviation).

CONCLUSION

Investors always wish to invest their funds and receive a high-expected return along with minimum risk. Fund diversification is a challenge for the decision-maker to fulfill the expectation of the investor. This research proposes an integrated approach of the K-means-GP for fund diversification in which K-means is used to create clusters of stocks. Then GP is applied for fund diversification to achieve the best return with an affordable risk. Experiments were done with the data of DOW30 in 2017-18. K-means provides 4 clusters of stocks, and then GP was applied to diversify the total fund with three different expected returns as 20%, 25%, and 30%. The results show that GP can diversify the total available funds with a maximum annual return of 27.5% with a minimum value of risks.

REFERENCES

Belaid, A., Cinzia, & C. and Davide, L.T. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art.*European Journal of Operational Research*, Vol.234, pp.536–545.

Charnes, A., Cooper, W.W., & Ferguson, R.O. (1955.) Optimal Estimation of Executive Compensation by Linear Programming.*Management Science*, Vol. 1 No. 2, pp.138-151.

Charnes, A., & Copper, W. W.(1961). *Management Model and Industrial Application of Linear Programming*, Vol. 1.John Wiley and Sons, New York.

Choudhary, D., & Shankar, R. (2014). A goal programming model for joint decision making of inventory Lot-size, supplier selection and carrier selection. *Computers & Industrial Engineering*, Vol.71, pp.1–9.

Ghahtarani, A., & Najafi, AA (2013). Robust goal programming for multi-objective portfolio selection problem.*Economic Modeling*, Vol.33, pp. 588–592.

Halima, B.A., Karim, H.A., Fahami, N.A., Mahad, N.F, Nordin S., & Hassan N. (2015). Bank Financial Statement Management using a Goal Programming Model.*Procedia - Social and Behavioral Sciences*, Vol. 211, pp.498 – 504

Hota, H.S. Singhai, S.K., & Awasthi, V. K. (2017). Expert Portfolio System Using Integrated MCDM-GP Approach.*Review of Business and Technology Research*, Vol. 14, pp.1-6.

Ijiri, Y.(1965). *Management Goals and Accounting for Control*. North-Holland, Amsterdam.

Ignizio, J.P.(1976). *Goal Programming and Extensions*. Lexington Books, Lexington, Massachusetts.

(A complete list of references is available upon request)

DECISION SCIENCES INSTITUTE
The role of ISO/IEC 27001 for Information Security: a review and research agenda

Giovanna Culot
Polytechnic Department of Engineering and Architecture, University of Udine
Email: culot.giovanna@spes.uniud.it

Guido Nassimbeni
Polytechnic Department of Engineering and Architecture, University of Udine
Email: guido.nassimbeni@uniud.it

Matteo Podrecca
Polytechnic Department of Engineering and Architecture, University of Udine
Email: podrecca.matteo@spes.uniud.it

Marco Sartor
Polytechnic Department of Engineering and Architecture, University of Udine
Email: marco.sartor@uniud.it

ABSTRACT

This paper presents the first review on ISO/IEC 27001, the most renowned on information security management, and the third most widespread ISO certification. Through a systematic approach, 58 articles were analyzed. The themes are presented on five research foci: relation with other standards, motivations, issues in the implementation, possible outcomes and contextual factors. The results are read through the lens of Social Systems thinking to formulate a research agenda. This paper is aimed at inspiring interdisciplinary studies in information security. The findings can support managers in deciding how to implement the standard and policymakers in shaping relevant regulation.

KEYWORDS: ISO/IEC 27001, Information Security, Management Systems, Systematic Literature Review, and ISO standards

DECISION SCIENCES INSTITUTE

Examining the relationship between e-government development, nation's cybersecurity commitment, business usage and economic prosperity: A cross-country analysis

Ben Krishna
Indian Institute of Management, Kozhikode
Email: benkrishna12fpm@iimk.ac.in

M P Sebastian
Indian Institute of Management, Kozhikode
Email: sebasmp@iimk.ac.in

ABSTRACT

The adoption of the cybersecurity measures varies across countries in spite of the benefits. This study proposes a model to examine the relationships between e-government development, cybersecurity commitment, business usage, and economic prosperity of the country. Structural equation modeling was used to analyze the country-level variables to explain the second-order impact of e-government development through the mediating role of cybersecurity commitment. Findings suggest that e-government development demonstrated a stronger association with cybersecurity commitment and business usage. There is preliminary evidence that the improvement of cybersecurity measures initiated by the e-government development will drive business usage and improve the macroeconomic conditions.

KEYWORDS: e-government, Cybersecurity, E-business, Economy, and Cross-country analysis

INTRODUCTION

Various economic theories, including neoclassical theories and new growth theories, argue that technological progress and creativity are determinants of the development of nations (Romer, 1990). E-government development is no exception, and it is the backbone of the digital economy and is often considered as the natural resource, which brings competitive advantages to a country (Srivastava & Teo, 2008; e-government survey, 2018). The most common definition of e-government in the literature is consistent with the framework proposed by Ndou (2004). The three components which characterize the e-government concerning the dissemination of Information communication and technologies (ICT) include (i) transformational nature, (ii) implementation of e-government application domains, and (iii) public, other stakeholders and their relation (Twizeyimana & Andersson, 2019). E-government transformation areas include internal, external, and relational (Ndou, 2004). Internal transformation refers to the use of ICT to improve the internal function and processes of intra-governmental agencies. External transformation refers to the exploration on the possibilities of ICT for the benefits of businesses and citizens. Relational transformation refers to the exploration on how ICT brings fundamental changes in the relationships between state and citizen, such as e-vote. Over the years, the transactional and facilitation potential of ICT created a paradigm shift in the quality, standard, and ethics of the public sector services (e-government survey, 2018), which ultimately benefited the businesses, citizens, and employees (Srivastava, 2011).

Despite having the advances in e-government development, increased digitization brings risks and challenges, including growing threats to businesses and economic development. Cyber threats such as denial of e-services, data integrity breaches, and data confidentiality breaches

are major challenges for information security. According to the official annual report of 2019 on cyber-dependent crime, Steve Morgan, the Editor in chief of Cybersecurity Ventures, commented, “Cybercriminal activity is one of the biggest challenges humanity will face in the next two decades.” Furthermore, the budget for preventing cybercrimes in organizations will hit $6 trillion by 2021. The cyber- dependent crimes around the globe has increased manyfold, and there is a need for extensive research in this area (Maimon & Louderback, 2019). The data breaches majorly affect personal medical data (22 percent of the cases between 2005 and 2012), personal identifiable data (74 percent of the cases between 2005 and 2012), and personal financial data (22 percent of the cases between 2005 and 2012) (Ravi Sen & Sharad Borle, 2015). As the Internet becomes more pervasive, from online banking to smart systems, higher standards of cybersecurity are essential. Along with this line, it is logical to assume that the e-government development will necessarily boost the nation’s resilience in fighting cyber-crimes. In summary, given these frequencies and financial impacts, it is imperative to understand what role a nation could play in safekeeping the interests of its businesses and citizens by committing to cybersecurity measures. This is the primary motivation of this research.

The existing research on the impact of e-government indicates that the development of e-government offers several benefits to businesses and citizens. From the business perspective, these include increased business competitiveness, improved economic performance, reduced corruption, and high government transparency (Srivastava & Teo, 2008; Krishnan et al., 2013). From the citizen’s perspective, benefits include enhanced service delivery, eradication of digital divide, increased democratization, among others (Cho & Choi, 2004; Von Haldenwang, 2004; West, 2004; Wong & Welch, 2004; Srivastava & Teo, 2007; Krishnan et al., 2013). Besides, our literature review reveals that the impact variables of e-government development include intermediate variables (first-order effects) such as corruption, national efficiency (resource spending efficiency and administrative process efficiency), which have an impact on the outcome variables (second-order effect) such as national performance (business competitiveness, reduction of social divide, and less ecological impact) (Srivastava & Teo, 2007; Krishnan et al., 2013). Thus, advances in ICT have enabled the dissemination of government services through various channels, which promises data processing and protection, transparency of government to businesses and citizens, increased trust, reduction of cost, among others (Alshehri & Drew, 2010). However, qualitative assessment of data suggests that increased level of e-government development might not reflect in the cybersecurity commitment because the government does not necessarily invest in cybersecurity with the same level of commitment (Global cybersecurity report, 2018). For example, high profile data breach happened in the United States Office of Personnel Management (OPM) exposed 21.5 million records and has done very little to improvise the security and acceptance of such services (Wagstaff et al., 2015). Thus, it was unclear whether e-government development would lead to cybersecurity commitment and avoid the possible negative impacts on businesses and economic development (Thompson et al., 2020). Although there are perceptible linkages between e-government development, cybersecurity commitment, businesses, and economy, it is surprising that this relationship has not been explored much in the current cybersecurity literature. This gap in the literature is the secondary motivation for this research.

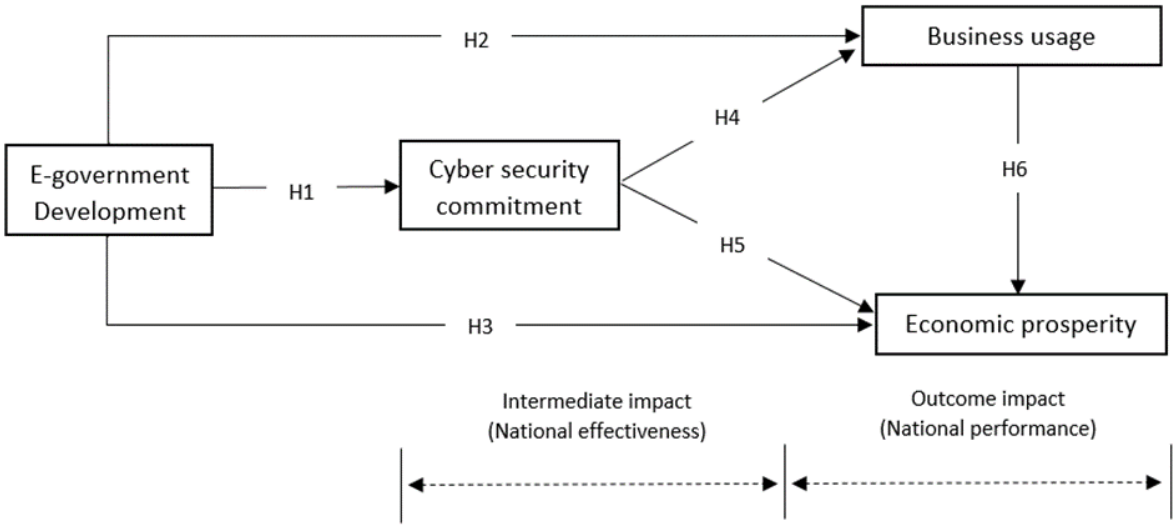
To address this gap, this study conceptualizes the relationship between e-government development and outcome impact through a mediating variable. We construe the second-order effects of e-government development as two variables, namely, business usage (e.g., integrate ICT into their operation (B2B and B2C)) and economic prosperity (e.g., gross national savings, government debt, and inflation), and the first-order impact as the nations’ cybersecurity commitment, defined as “commitment of a nation in building confidence and security in the use of ICT technologies.” To possibly analyze this macro-economic relationship concerning e-

government and cybersecurity commitment, cross-country quantitative empirical studies are best suited (Srivastava & Teo, 2008). Further, most of the research concerning the cybersecurity commitment is anecdotal and descriptive (Luiijf et al., 2013; Awoleye et al., 2014; Ellefsen, 2014, May; Teo & Mahmood, 2017, July). A few studies, which address the macro-economic perspective, only focus on one of the impact variables. For instance, a study conducted by Thompson et al. (2020) focuses on comparing the e-government website's security across two countries whose e-government development levels are different. Thus, there is an imperative need to conduct large-scale quantitative empirical studies exploring the relationships between e-government, cyber commitment, and outcome impacts. To empirically analyze, this study utilizes the archival data published by global resources, namely, World Economic Forum (WEF), International Telecom Union (ITU), and the World Bank. To the best of our knowledge, there is no large-scale study, which aims to understand the impact of e-government development through cyber commitment. The key research question (RQ) that we address in this study is:

RQ: What are the relationships between e-government development, the nation's cybersecurity commitment, business usage, and economic prosperity in a country?

The rest of the paper is organized as follows: Section 2 presents the research model and hypotheses, and Section 3 presents the research design. Section 4 analyses and presents the results. Section 5 discusses the findings of this research. Section 6 presents the implications, limitations, and scope for future research and Section7 concludes the paper.

Figure 1: The Research Model



RESEARCH MODEL AND HYPOTHESES

In this section, we describe the theoretical linkages among e-government development, a nation's cybersecurity commitment, business usage, and economic prosperity. Figure 1 shows the research model with the proposed hypotheses.

Relating e-government development to a nation's cybersecurity commitment

According to the UN, e-government development is defined as how the government in a country has established online service capability, telecommunication infrastructure capability, and human capacity. Previous researches on e-government conceptualized the development through an evolutionary approach like different stages of development (Lee, 2010). Advances in ICT have pushed the government to develop in two fronts, namely, citizen-centric services and integration of government operation digitally, both horizontally and vertically (Lee & Kim, 2007). According to this approach, e-government can assist the government in the innovation of governance process, improving efficiency and effectiveness, while improving more participatory opportunities for businesses and citizens.

Evaluation of the various popular e-government stage models indicates the evolution of two dimensions, namely services of government towards citizens and technical/operational characteristics of government (Lee, 2010). E-government services include transaction, interaction, participation, and democracy, focusing mostly on the access and transparency of the various government operational services. The concept of integration and transformation belongs to the technical/functional characteristics of the government, which emphasizes the technological perspective of the operation of government. Implicitly, e-government development includes stages ranging from presenting information to transforming processes and services for effectiveness. Apart from the technical and managerial perspective, an additional dimension introduced was the political dimension where businesses and citizens are encouraged to interact with the government through polls, forums, e-meeting, among others (Siau & Long, 2005). Thus, advancements in e-government bring technological challenges (Bretschneider, 2003). E-government faces security challenges in the following aspects, namely, information intercepting, information tampering, services denying, system resource stealing, information faking, among others (Zhou & Hu, 2008). E-government must overcome these challenges to secure government resources from the information security threats since e-government systems have linkages to the public and businesses (Zhou & Hu, 2008).

"Cybersecurity is the collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurances and technologies that can be used to protect the cyber environment and organization and user's assets" (Global cybersecurity report, 2018). In the absence of proper cybersecurity measures, cyber intrusion will impair the e-government services, especially e-government web services and ICT infrastructure (Halchin, 2004). An assessment of 212 e-government websites revealed that 82% of the sites worldwide were vulnerable to cross-site scripting and SQL injection (Moen et al., 2007). Cross-cite scripting, also known as frequent web application attacks, is more prevalent in European websites (90%), followed by North American websites (72%). Findings from other studies have agreed that public administration is more vulnerable to spear-phishing attacks (Symantec, 2014). Further, in most of the cybersecurity breaches, the human element is considered as the weakest link (Boss et al., 2009). These human elements invite threats to the system and infrastructure intentionally or unintentionally, which will compromise the effectiveness of services. In 2013, Edward Snowden, a former contractor of the National Security Agency, leaked data that involved cyber espionage, putting the United States and the United Kingdom in the spotlight (Symantec, 2014).

Furthermore, Zhao et al. (2010) assessed 51 US states and found several security flaws, which would expose the IP addresses of the websites. Similarly, other countries were also evaluated, namely Nigeria (Awoleye et al., 2012; Awoleye et al., 2014), Kyrgyz Republic (Ismailova, 2017), Jordan (Alsmadi & Shanab, 2016), Burkina Faso (Bissyandé et al., 2016), and Libya (Ali & Murah, 2018). Results from these studies recommend that policies should be in place to assess

the e-government services to improve authenticity, thereby gaining public and business trust. All the previous studies, except Zhao et al. (2010), studied nations ranked low in e-government development. For instance, the US ranked 11, Thailand ranked 73; Nigeria ranked 143 in the e-government development index (e-government survey, 2018). Recently, the security assessment of the e-government websites of Australia and Thailand have revealed that Australia stands high on the adoption of cybersecurity practices than Thailand (Thompson et al., 2020). Moreover, Australia ranked higher than Thailand in the e-government development index (e-government survey, 2018). Even though there are mixed findings regarding the linkages between e-government development and a nation's cybersecurity commitment, we can argue that the development of the e-government influences a nation's commitment to cybersecurity, such that cybersecurity will safely keep integrity, availability, and confidentiality of the process and data in the information society.

Besides, this can be corroborated by borrowing the neo-institutional theory (DiMaggio & Powell, 1983). The theory argues that institutions become similar over time because of the institutional pressure through the process of isomorphism. Exploratory research in IS security reveals internal and external drivers that motivate senior management commitment to information security (Johnson, 2009). Empirical evidence suggests that external drivers dominated the influence on senior management's commitment to invest in information security. External drivers include compliance with policies, integrity, respect to the customer's data, and external pressure from the competition and business partners (Johnson, 2009). Lee & Larsen (2009) empirically demonstrate that external mechanisms like mimetic, coercive, and normative put pressure on senior management to invest in information security technologies. Institutional theory has been used in studies related to ICT in government settings (Yildiz, 2007; Luna-Reyes & Gil-Garcia, 2011; Zheng et al., 2013). Public organizations are required to adhere to the legitimacy acts set by the external institutions because of their inter-organizational linkages (Gil-Garcia et al., 2007). A normative mechanism is enforced through the professional network and formal education, which will evoke the cognitive base and thus influence the belief (DiMaggio & Powell, 1983). Collaboration and professional association of government organizations with several high-end private and public organizations (suppliers and customers) worldwide will lead to the adoption of similar information systems (Teo et al., 2003; Zheng et al., 2013).

Further, mimetic mechanisms are essential in an uncertain environment, and institutions tend to follow the successful. This is more prevalent in situations where technology is not understood (DiMaggio & Powell, 1983). There is empirical evidence that greater the adoption of information systems by competitors and the success of those competitors influence parent institutions to adopt similar information systems (Zheng et al., 2013). Here in the context of e-government development and cybersecurity commitment, to protect the network and infrastructure of parent nation, similar or better cybersecurity measures are employed than other countries.

Similarly, coercive mechanism occurs when the institution feels pressures from formal or informal sources, where they were forced to adopt processes and structures that are viewed as legitimate in culture and environment (DiMaggio & Powell, 1983). To facilitate online trade and to offer online services in a reliable manner, the firm has to adhere to specific standards like ISO/IEC 27000 family, which is also termed as benchmarking (Johnson & Goetz, 2007). Benchmarking with the industry and between various public institutions can ensure that cybersecurity is at par with its peers. In addition, resource-dependent public institutions experience superior (bi-lateral agreements and international customers) and latent coercive pressures (local suppliers and customers), which will lead them to adopt similar information systems and standards (Teo et al., 2003; Zheng et al., 2013). These mechanisms will either influence the belief or directly influence the legitimacy acts of government organizations. In a similar vein, it is reasonable to believe that when the development of e-government results in

cybersecurity challenges, nations are forced to adopt the process and structure to improve the e-government effectiveness by committing to cybersecurity. Therefore, we hypothesize that,

H1: In a country, e-government development is positively associated with the nation's cybersecurity commitment.

Relating e-government development to business usage and economic prosperity

Business usage "captures the extent to which businesses in a country use the Internet for business-to-business (B2B) and business-to-consumer (B2C) operations, as well as their efforts to integrate ICTs in their operations." (ITU, 2016). Thus, business development in a country more or less dependent on the overall ICT structure and support obtained from the country. Prior research shows that e-government development, especially the services to businesses, are instruments to reduce the private cost at the expense of government (Allers, 1994). Business units are considered as a stakeholder, and through e-government services, business units achieve reduced administrative burden (Rowley, 2011). Primarily, they seek two kinds of benefits, first, to meet operational objectives like access and use of digital infrastructure. For instance, established delivery channels installed by the government mostly benefit Small and Medium Enterprises. The other benefits include meeting specific objectives like increased security, lessening administrative burden, more transparency and accountability, more convenience, increased benefits to the employees and customers. There are shreds of evidence that government-to-business services (G2B) in some countries are more developed than government-to-customer services (G2C). For instance, across the European Union countries, G2B services are more matured than G2C facilities (Rohleder & Jup, 2004). Central to the relationship between e-government development and business usage is the alignment of the interest, mainly communication, collaboration, coordination, commerce, transparency, and accountability (Yildiz, 2007).

The Technology-organization-environment (TOE) model proposed by Tornatzky & Fleischer (1990) has been adopted to study business-to-government relationship (Kuan & Chau, 2001; Zhu et al., 2002; Arendsen et al., 2008). The TOE framework explains the three groups of variables influencing the adoption decision process: technological, organizational, and environmental. The technical context demonstrates the willingness and capability of the firm to adopt technological innovation, which includes current ICT infrastructure and its characteristics. The organizational context explains the governance, size of the firm, quality of human capital, especially the resources and capabilities. The environmental context explains the environmental condition such as market characteristics and external sources like peer pressure, competition, among others. E-government development constitutes of these factors, namely, technological infrastructure and innovation, human competency, the efficacy of public institutions, which mostly drive the adoption of e-government services in the business. Therefore, we hypothesize that,

H2: e-government development is positively associated with business usage in a country.

According to the Resource Based View, a firm is a collector of resources and capabilities (Barney, 1991). The presence of unique, rare resources and capabilities, which cumulate over time, leading to sustainable competitive advantage for a firm. Extending this view along with the evolutionary economic and entrepreneurship, Mathews (2002) put forward the idea of "resource economy," which emphasizes the unique configuration of resources and firms in an economy or nation, not just within the firm. Further, the competitive advantage of a country is contingent upon the presence of fundamental factors like demand and supply of resources, production

function, competitive environment, supporting industries, and advanced and specialized factors like the establishment of digital networks and skilled human personals (Porter, 1990). Along with this line, several researchers argue that e-government is a natural resource and brings a competitive advantage to the nation (Srivastava & Teo, 2008; Srivastava, 2011).

Previous research emphasizes the role of ICT in building a nation's prosperity, especially like the adoption of ICT in government services. For instance, a study estimated that, on the average, 20 to 25 percentage of the GDP is accounted for the digital economy (skills, resources, and capabilities) (Knickrehm et al., 2015). Zhao et al. (2015) point out that there exists a reciprocal relationship between e-government development and economic prosperity, a self-sustainable system. A growing number of nations have implemented the digitization process, which uplifts the e-government capabilities, which plays a pivotal role in boosting the economy and society (Von Haldenwang, 2004; Yildiz, 2007). For instance, the use of ICT in the local government institutions led to the proliferation of e-government, which finally led to economic welfare (Clark et al., 2003). Similarly, the adoption of smart sensors and IoT by local government help them to monitor and manage their operations effectively (Tang & Ho, 2019). Another study conducted by Krishnan et al. (2013) reports that e-government development reduces corruption and favors the growth of the economy. These indicate that there are strong reasons to believe that the development of e-government will improve the macroeconomic conditions of a country. Thus, we hypothesize that,

H3: e-government development is positively associated with the economic prosperity of a country.

Relating a nation's cyber commitment to business usage and economic prosperity

Thirty hackers and a budget of $10 million could bring the United States down (Dynes et al., 2007). This is an example of how a cyberattack could affect every nation and business entities as cyber dependence (utilization of ICT in a network, infrastructure, and services) is increasing. Cyber attackers are powerful enough to crack down and destroy the targeted state or society (Lewis, 2003). In this unstable environment, how could business entities build cyber resilience? Prior research suggests that there are five ways where businesses could bring peace in the cybersecurity context (Evers, 2010; Shackelford et al., 2014). Synthesize of five ways tells us two possible ways a country could fulfill business needs by building confidence and security. One of the possible ways is employing cooperation measures, which include bilateral agreement, public-private partnership, and capacity building. Thus, cybersecurity can be tackled through a socio-political approach that brings more robust cybersecurity capabilities, which help to deter repeated online threats. Another possible way is by employing legal and organizational strategy measures to impose restrictions on the conduct of businesses. This could bring penalties for breach of law or noncompliance. Further, organizational measures or strategies could bring harmonization in cybersecurity efforts (Global cybersecurity report, 2018).

Another aspect is the protection of critical infrastructure (Dynes et al., 2007). ICT infrastructure in a country can be envisioned as encompassing "all computerized networks, applications, and services that citizens can use to access, create, disseminate and utilize digital information" (Selwyn & Brown, 2000, p.662). From the definition of the ICT infrastructure in a country, it can be argued that to communicate and coordinate with the public and business, a secure and sound infrastructure is inevitable. Further, the impact of ICT on the business is twofold, one is direct benefits, and the other is second-order benefits (Arendsen et al., 2014). Immediate benefits include reducing communication costs and time, which is categorized as operational benefits called the electronic integration effect (Malone et al., 1987). The indirect impact is termed as radical integration, which is the horizontal and vertical integration of intra-

departments or automates the whole functions of the business (van den Boer et al., 2011). Thus, as the relevance of the digital economy increases, the presence of cyber-secure resources and cyber-capabilities becomes a necessity.

Theoretical support for the above arguments can be drawn from the Resource Based View, as a nation's cybersecurity commitment can be conceptualized as a complimentary resource. A resource, which enhances the value of an original resource, can be termed as a complimentary resource (Zhu, 2004). This essentially means that the presence of an additional resource that complements the primary resource will bring uniqueness and creates an extra level of sustainable competitive advantage. Besides, the interaction between the primary resource and complimentary resource will create synergies (Srivastava & Teo, 2008). Thus, the presence of the developed e-government will substantially enhance business usage but will have a more significant impact on the presence of a better cybersecurity environment. Therefore, we hypothesize that,

H4: A nation's cybersecurity commitment is positively associated with business usage in a country.

The cyber threats that affect a nation can be categorized as malware or zero-day attack, rampant organized crime, personal information, data breach, state-sponsored attacks, and advanced persistent threats (Teoh & Mahmood, 2017, July). Further, the nature of the attack will also be caused by weakened policies, which can cripple the digital innovation (Burt, 2014). This will affect the cost to the economy. For instance, the US, China, Japan, and Germany incurred a combined cost of 200 billion USD because of cybercrime activities. It is not the actual cost, but the opportunity and recovery loss that matters. For instance, the damage due to cybercrime in Italy recorded as USD 875 million, but the recovery and opportunity cost mounted up to USD 8.5 billion (McAfee, 2014). Cybersecurity and digital economy need to go hand in hand to build cyber resilience. As cybersecurity become prioritized in a country, it will foster economic welfare (Luiijf et al., 2013). Between the period of 2013 to 2016, 48 national cybersecurity strategies had been released based on NATO and ENISA (Teoh & Mahmood, 2017, July). The main objectives of these strategies include "maintaining secure, resilient and trusted electronic operating environment, promoting economic and social prosperity, promoting trust, business, and economic growth, addressing the risk of ICT and strengthening the resilience of infrastructures" (Osula & Kaska, 2013). A nation's commitment to cybersecurity cements the digital economy and improves the macroeconomic condition. Therefore, we hypothesize that,

H5: A nation's cybersecurity commitment is positively associated with the economic prosperity of that country.

Relating business usage to economic prosperity

E-business is often defined as the use of ICT for the buying and selling of products and services. These business transactions include B2C, B2B, B2G, C2B, and C2C. By the 1990s, the wide diffusion of ICT in the industries had contributed to the betterment of business services (Zhu, 2004; Gallouj & Savona, 2009; Arendsen et al., 2014). The literature related to ICT and business stress on the interactions and complementarity between them (Miles, 2004). Furthermore, E-business helps improve the communication between individual firms, thus improving the overall performance of all the stakeholders, namely, the supplier, buyer, intermediary, and end-consumer (Subramani, 2004). Hence, the aggregation of productivity returns of the individual firms will improve national productivity and improve the macro-economic conditions (Srivastava & Teo, 2010). This tells us that the more significant the business usage,

especially with the diffusion of ICT into the business, the more is the productivity of nations, which results in enhanced economic prosperity. Therefore, we hypothesize that,

H6: Business usage is positively associated with the economic prosperity of a country.

Mediating effects of a nation's cyber commitment

After connecting and establishing relationships among the individual variables in the research model, we logically deduce one more hypotheses. We posit that a nation's cyber commitment serves as an intervening mechanism or, at the least, partial conveyor of the effects of e-government development on business usage and economic prosperity. That is, e-government development indirectly affects business usage and economic prosperity through higher levels of the nation's cyber commitment. Therefore, we deduce that,

H7: A nation's cyber commitment mediates the relationships of e-government development with business usage and economic prosperity.

RESEARCH DESIGN

To test our hypotheses, we needed a large dataset gathered from different countries aggregated at the national level. For this, we have taken secondary data available from some reliable data sources that were used in the previous research. Archival data offers advantages such as easy reproducibility, generalization of results, and robustness to common method bias (Krishnan et al., 2013). The hypotheses were tested using the cross-sectional analysis of 127 countries from the year 2014 to 2019. Furthermore, multiyear averages were considered for accurate, stable estimates and are consistent with the previous researches (Wiggins & Ruefli, 2005; Krishnan et al., 2013). The data sources in this research include the Global information technology (GIT) report by the WEF, Global cybersecurity index (GCI) by ITU, Global competitiveness report by WEF, and the e-government survey report by UN (ITU, 2014, 2016,2019; ITU Cyber, 2014, 2017, 2018; UN, 2014, 2016, 2018; WEF Competitiveness, 2014, 2016, 2019). Further, data from these data sources are considered reliable and have been extensively used in previous researches (Srivastava & Teo, 2008). Table 1 presents a summary of the variables and their data sources.

Table 1: Summary of variables and their data sources

Variables	Measure	Data sources
EGD	E-government development is a composite measure of individual components: online service index (OSI), telecommunication infrastructure index (TII), and human capital index (HCI).	E-government survey reports from the United Nations (UN, 2014, 2016, 2018)
CybCom	Global cybersecurity index focuses on the five pillars: legal, technical, organizational, capacity building, and cooperation	International Telecom Union (ITU Cyber, 2014, 2017, 2018)

Macro	This represents the macroeconomic condition of a country, which include gross national savings as a percent of GDP, government debt as a percent of GDP, and among others.	World Economic Forum (WEF Competitiveness, 2014, 2016, 2019)
BusUsage	Business usage captures the extent to which businesses in a country use the Internet for B2B and B2C operations, as well as their efforts to integrate ICTs in their operations.	Global information technology report by ITU (ITU, 2014, 2016, 2019)

Operationalization of constructs

As depicted in the research model (Figure 1), our study consists of four variables: e-government development, nation's cyber commitment, business usage, and economic prosperity. The independent variable e-government development was accessed from the e-government survey reports from the UN (UN, 2014, 2016, 2018). E-government development is a composite measure of individual components: online service index (OSI), telecommunication infrastructure index (TII), and human capital index (HCI). OSI captures the level of maturity attained by the e-government. TII captures the status of the development of telecommunication infrastructure that includes internet users per 100 inhabitants, fixed telephone lines, and more. HCI captures inherent human capital in the country that includes adult literacy rate, average years of schooling, and more. The intermediate variable nation's cybersecurity commitment is taken from the GCI published by the ITU (ITU Cyber, 2014, 2017, 2018). The Global cybersecurity index focuses on the five pillars: legal, technical, organizational, capacity building, and cooperation. The sample list of concepts measured includes cybersecurity regulation, use of the cloud for cybersecurity, cybersecurity metrics, professional training, educational awareness, bilateral agreements, among others.

The dependent variable, Business usage, was taken from the GIT report by the ITU (ITU, 2014, 2016, 2019), and economic prosperity was operationalized as the macroeconomic condition of a country (WEF Competitiveness, 2014, 2016, 2019). Business usage is a composite measure; sample items include (i) use of ICT in B2B transaction, (ii) use of ICT in B2C (iii) firm-level technology absorption, among others. Economic prosperity is operationalized following Porter's productivity paradigm as well as other macro-economic conditions. The sample list of concepts measured here includes (i) gross national savings as a percent of GDP, (ii) government debt as a percent of GDP (iii) government budget balance as a percentage of GDP, among others. Further, we controlled for regional differences. The regional difference indicates the country-level difference across various regions of the world. This classification includes Asia, Africa, America, Oceania, and Europe.

Data validity and reliability

Global data collection agencies like the WEF, the UN, and the ITU followed the standard processes to ensure the reliability and validity of the data. For instance, data for the e-government development were collected through a systematic procedure which included (i) a methodological framework which is based on the literature review and composes of three measures, (ii) data collection through the questionnaire distributed across 193 countries, and (iii) a preparatory process for the survey which included an external evaluation of the previous

surveys, two expert group meetings (contains academicians and practitioners from public, private and civil), and those recommendations were included in the survey. Similarly, for cybersecurity commitment, the questionnaire included binary, pre-coded, and open-ended questions. It involved both primary and secondary data collections through focal point format to ensure a better response rate. Binary answers avoided bias and eliminated opinion biased evaluation. Further, the scoring process was carried out by a panel of experts providing weightage for each question. ITU followed rigor methodological framework to ensure reliability, which includes (i) data providers are credible and reliable organization, (ii) release of data is regular, (iii) at least two data years for each indicator, (iv) data should not be older than three years, and (v) data covers at least two-to-three countries to avoid bias. Moreover, multiple sources of data were used, which include household, individual, business organization, and non- government organizations. In summary, the archival data used in this study were collected from the various data collection agencies with a stringent methodological framework to ensure data reliability and validity. In addition, as noted earlier, the data from these reports have been used in previous studies.

ANALYSIS AND RESULTS

Descriptive statistics and correlations

Descriptive statistics and variable correlation are shown in Table 2.

Table 2: Descriptive statistics and correlation

Variables	Mean	Standard Deviation	Macro	CybCom	EGD	BusUsage
Macro	4.732	0.932	1			
CybCom	0.470	0.224	0.451	1		
EGD	0.578	0.178	0.451	0.745	1	
BusUsage	3.209	0.931	0.522	0.696	0.816	1

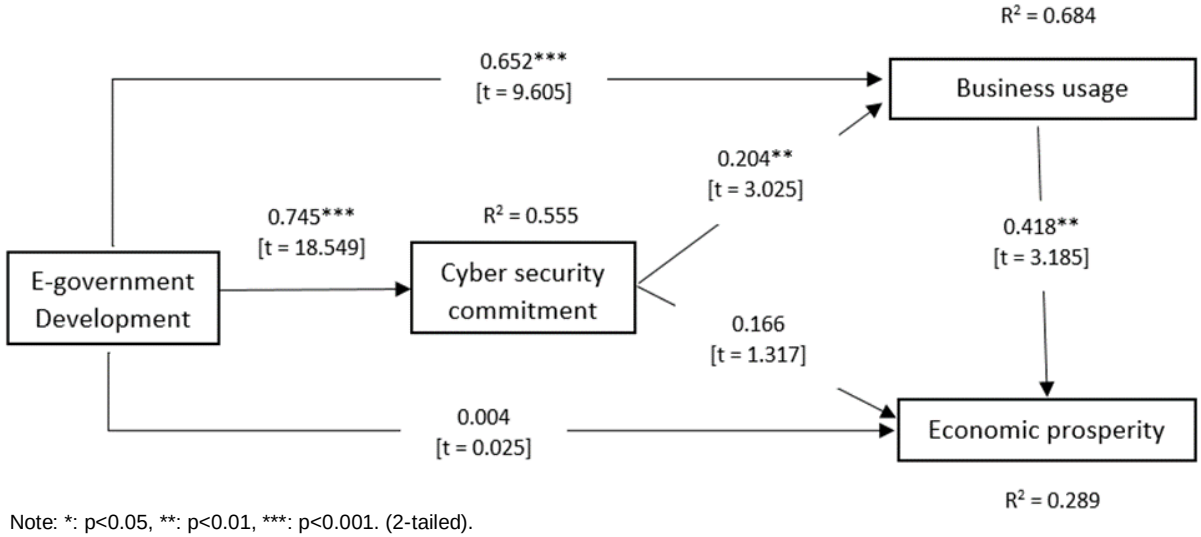
As shown in Table 2, e-government development (EGD) is positively related to nation's cyber commitment (CybCom), Economic prosperity (Macro), and Business usage (BusUsage), and so do the cybersecurity commitment to other dependent variables. Further, as all the correlations among the variables except EGD and CybCom are below the value of 0.8, concern for multi-collinearity is less. Nevertheless, we went for the Variance Inflation Factor (VIF) test to ensure all variables are devoid of collinearity. Our VIF results show that it ranges from 1.33 to 4.2. As a rule of thumb, VIF below 5 are acceptable and avoid the multi-collinearity issue (Hair et al., 2011).

Hypotheses testing

The data analysis was done using structural equation modeling (SEM), a popular multivariate data analysis method among the researchers. This paper uses a Partial Least Square (PLS) approach of SEM using SmartPLS that analyzes variance rather than co-variance based approach. The PLS is beneficial when (i) the research is exploratory where the relationship among the variables is less definite, (ii) a soft modeling approach is used with less assumption

about data distribution or is skewed, (iii) predictive accuracy is paramount, and (iv) minimum restrictions imposed on the scales, measurement, sample size and residual distribution (Hair et al., 2011). Thus, PLS was a suitable choice for analyzing the relationship between the variables in this paper. In the model tested, all the constructs are modeled as reflective constructs. SmartPLS 3 was used to analyze the data and test the hypothesis (Ringle et al., 2015). Furthermore, we used PLS bootstrapping technique to explain the significance of the mediation model (Specific indirect effects and total effects). The results of the PLS analysis of the structured model is depicted in Figure 2.

Figure 2: Results of the PLS analysis



Note: *: p<0.05, **: p<0.01, ***: p<0.001. (2-tailed).

As shown in Figure 2, E-government development is positively associated with cybersecurity commitment (0.745***, t=18.549), Business usage (0.652***, t=9.605), and not related to Economic prosperity (0.004, t=0.0025). Hence, H1 and H2 are supported, and H3 is not supported. Also, as hypothesized, cybersecurity commitment is positively associated with Business usage (0.204**, t=3.025) and not associated with Economic prosperity (0.166, t=1.317). Hence, H4 is supported, and H5 is not supported. Further, in line with the initial prediction, business usage is positively associated with economic prosperity (0.418, t=3.185). Hence, H6 is supported.

To test the mediated effects of a nation's cybersecurity commitment on the relationship between e-government development, business usage, and economic prosperity, our study referred to Preacher and Hayes's method of inspecting indirect effects (Preacher & Hayes, 2008). This method mostly analyzes the total effects, direct effects, and mediated effects through a mediator variable. To study the mediated effects, we choose the bootstrapping technique; a nonparametric resample procedure that does not have any assumption regarding the sample distribution. Through repeated sampling, this technique analyzes the indirect effect through the mediator. SmartPLS 3 provides the total effects and specific indirect effects. The results are tabulated in Table 3.

A total of three mediation or indirect effects were analyzed. The e-government development as the independent variable and business usage as a dependent variable controlling for other variables was examined first. From Table 3, it can be seen that e-government development has a significant indirect effect on business usage through cybersecurity commitment. However, it is to be noted that the direct impact of e-government development does not drop to an insignificant value after introducing the mediator. This means that the nation's cybersecurity commitment

Table 3: Summary of mediation analysis				
Total effects				
Relationship	*Coefficient*	*T-statistics*	*BC 95% CI [lower upper]*	*Hypothesis Supported/Not*
CybCom -> Macro	0.166	1.317	[-0.077 0.418]	H5 -No
CybCom -> BusUsage	0.204**	3.025	[0.069 0.335]	H4-Yes
Ebd -> CybCom	0.745***	18.549	[0.654 0.815]	H1-Yes
Ebd -> Macro	0.004	0.0025	[-0.311 0.318]	H3-No
Ebd -> BusUsage	0.652***	9.605	[0.526 0.791]	H2-Yes
BusUsage -> Macro	0.418**	3.185	[0.148 0.668]	H6-Yes
Specific Indirect effects				*Mediation*
Ebd -> CybCom -> Macro	0.124	1.269	[-0.067 0.319]	No
CybCom -> BusUsage -> Macro	0.085*	2.234	[0.026 0.179]	Yes
Ebd -> CybCom -> BusUsage -> Macro	0.064*	2.236	[0.020 0.136]	Yes
Ebd -> BusUsage -> Macro	0.273**	3.143	[0.102 0.447]	Yes
Ebd -> CybCom -> BusUsage	0.152**	2.979	[0.053 0.257]	Yes

Note: N = 127. 5000 bootstrap samples; e-government development (EGD), nation's cyber commitment (CybCom), economic prosperity (Macro), and business usage (BusUsage); *: p<0.05, **: p<0.01, ***: p<0.001. (2-tailed). 'Total' is the total relation between the independent variable and dependent variable without the consideration of other variables. 'Indirect' is the relationship between variables while considering other relationships among variables.

partially mediates the relationship between e-government development and Business usage. Then the e-government development as the independent variable and economic prosperity as a dependent variable controlling for other variables was examined. Table 3. shows that e-government development has an insignificant indirect effect on Economic prosperity through cybersecurity commitment. We then analyzed the total mediation effect using e-government development as the independent variable, economic prosperity as the dependent variable, cybersecurity commitment, and business usage as mediator variables. From Table 3, it can be observed that the above path is significant. However, the indirect effect has a lower coefficient value. Thus, our hypothesis H7 gets partial support. Further, to ensure the robustness of the results, the regional difference was used as a control variable, which did not have any significant impact on the predicted relationship. This proposed model (including control variable) explained 55.5% variance in cybersecurity commitment, 68.4% variance in Business usage, and 28.9% variance in economic prosperity.

Further, we have conducted SEM for each year of the dataset (2014, 2016, and 2018) as a robustness check. Hypothesized relation between e-government development with cybersecurity (β =0.7, $p<0.01$; β =0.71, $p<0.01$; β =0.76, $p<0.01$) and business usage commitment remained significant (β =0.67, $p<0.01$; β =0.73, $p<0.01$; β =0.89, $p<0.01$) for each year and relationship between e-government development and economic prosperity remained

insignificant for two years. Similarly, indirect effects (cybersecurity commitment as mediator) remained significant between e-government development and business usage (β =0.083, p<0.05; β =0.099, p<0.05; β =0.152, p<0.1) and indirect effects (business usage as mediator) remained significant between e-government development and economic prosperity (β =0.0134, p<0.05; β =0.143, p<0.05; β =0.275, p<0.05). Lastly, we tested for the possibility of reverse causality. However, all the hypotheses were formulated to signify a ''relationship'' rather than a ''causation'' between e-government development and impact variables (i.e., cyber commitment, economic prosperity, and business usage).

Due to the limitation of data and instrument variables, we could not analyze using two-stage least squares (2SLS) regression; rather, we have analyzed the reverse linkage using WarpPLS (Pearl, 2009; Kock, 2015e; Kock & Gaskins, 2016). We have conducted two tests, Simpson's paradox ratio (SPR) index and Warp3 bivariate causal direction ratios. SPR index measures whether the model is free from Simpson's paradox, which is a possible indication of the hypothesized relationship as either implausible or reversed (Kock & Gaskins, 2016). Our SPR value 0.833 (>0.7) indicates that the hypothesized relationship is free from Simpson's paradox. Warp3 bivariate causal direction ratios help to identify whether the causality may be reversed or not. This ratio is calculated by dividing the path coefficients of the reversed link to path coefficients of the hypothesized relationship. A value greater than one indicates reversed linkages, and a value less than 1.3 shows weak support of reverse linkages. Our results indicate that the ratio between e-government development and business usage (0.949), cybersecurity commitment, and business usage (0.981), cybersecurity commitment, and economic prosperity (0.993) have no reverse causal relationship. However, the ratio between e-government development and economic prosperity (1.022), e-government development and cybersecurity commitment (1.003), business usage and economic prosperity (1.061) indicates weak reverse linkages (Kock, 2020).

DISCUSSION

Findings from this study raise several positives as well as concerns that deserve attention. Talking about the positives, e-government development is positively associated with the nation's cybersecurity commitment. Previous qualitative assessments raised doubts regarding the linkages between the e-government developments in a country that will probably result in cyber commitment. For instance, the Global cybersecurity report of 2018 emphasizes the gap with e-government development and cybersecurity commitment; that is, the nations that stand top at e-government development stand bottom in the Global cybersecurity index. For example, Denmark, ranked first in the e-government development index (2018) and was ranked 21 in the GCI index, raises a serious concern. Further, previous studies were not able to answer this concern by assessing the quality of e-government websites (Zhao et al., 2010; Bissyandé et al., 2016; Ali & Murah, 2018). However, in the study conducted by Thompson et al. (2020), the quality of e-government websites was studied in terms of cybersecurity across two countries (Australia and Thailand). The results from that study partially support the role of e-government development in commitment to cybersecurity, as Australian websites follow industry-standard procedures more diligently than their Thai counterparts do. Thus, by studying the relationship between e-government development and the nation's cyber commitment, our study contributed to IS literature on the impact of ICT at the macro level.

Another positive relationship worth mentioning is the relationship between e-government development and business usage. This is a significant result, as it reassures the strong bond between them over the years. This is corroborating with the previous findings, which highlighted the importance of using online services that will enhance the following: staying informed, more transparency, increased trust, among others (Arendsen et al., 2008; Srivastava & Teo, 2010;

Lee et al., 2011). Further, the role of e-government in reducing the administrative burdens of the business has been stressed in the literature that included case studies also, which will improve cost structure and time to achieve productivity returns (Arendsen et al., 2014). Further, the role of information infrastructure and quality of human capital will improve the business usage (van den Boer et al., 2011; Dash & Pani, 2016; Ali & Osmanaj, 2020). Another finding worth mentioning is the positive association between business usage and economic prosperity and is consistent with the previous results. This essentially means that the diffusion of ICT in the business has elevated the national performance. Besides, the positive association between cybersecurity commitment and business usage is highly relevant; as the digital economy progresses and the adoption of ICT services and infrastructure by the businesses increases, a robust and reliable cyber resilience is a precursor. As noted before, theoretically, cyber commitment acts as a complimentary resource that builds confidence in the operationalization of business activities (Wilding, 2016). Further, the role of cyberinfrastructure is of prime importance, as it will protect against terrorism (Halchin, 2004; Burt, 2014). Thus, this study provides quantitative merit in the relationship between cyber commitment and business usage.

The two major concerns with respect to our findings include the insignificant relationship that exists between e-government development and economic prosperity and a nation's cyber commitment and economic prosperity. One possible reason for the insignificant relationship between e-government development and economic prosperity could be the smaller presence of e-government in some countries (Krishnan et al., 2013). Further, we have used macroeconomic conditions, which include inflation, gross nation saving, budget balance for operationalizing economic prosperity. This essentially means that e-government development has not reached the stage where it could affect or improve the macroeconomic conditions. This finding could be considered as a fuel for future research to delve deeper into the above concern. The second concern is on the insignificant relationship between a nation's cyber commitment and economic prosperity. In contrast, there are strong reasons to believe that a nation's cyber commitment can improve macroeconomic conditions. For instance, national cybersecurity strategies, especially security policies, have helped the government to build a digital economy and have an impact on various stakeholders in the society (Ellefsen, 2014, May; Teoh & Mahmood, 2017, July). Thus, considering the implications of cyber commitment, the findings should have been favorable. Nevertheless, one of the possible reasons could be the aggregate measure of macro-economic conditions used for operationalizing economic prosperity. A simple measure like GDP per capita might have yielded a favorable result. In addition, the relationship between a nation's cybersecurity commitment and economic prosperity could be realized in the presence of other intermediary or contingent factors.

Turning into the mediation effect, two relationships are worth discussing. Firstly, e-government development was associated with business use through the mediated effects of a nation's cybersecurity commitment. This means that benefits from the development of e-government in terms of e-government services, technical infrastructure, and quality of human capital can be realized indirectly through cybersecurity commitment (better laws and regulation, international agreements, technical support, and protection of cyberinfrastructure) and, in turn, impact the business usage. Secondly, e-government development was not associated with economic prosperity through the mediated effects of the nation's cybersecurity commitment. As mentioned in the previous section, this could be attributed to an aggregate measure of macroeconomic conditions, a multiyear average dataset, or no alleged significant relationship exists between them at this point. However, e-government development is indirectly related to macroeconomic conditions through the double mediation effect of a nation's cybersecurity commitment and business usage and is supported by PLS analysis (See Table 3). However, this relationship might vary across countries, and thus needs further investigation. In the following section, we discuss the implications and limitations of this research and propose future research directions.

IMPLICATIONS, LIMITATIONS, AND FUTURE RESEARCH

The present study makes three significant theoretical contributions. First, while the existing studies on cybersecurity linking e-government development, business usage, and economic prosperity are often descriptive and mostly undertaken through qualitative case study approaches or website analysis, our research, using archival data, has identified quantitative merit in the relationships among them. Second, while different global reporting agencies have raised doubts about the relationships between cybersecurity commitment and other indices, especially e-government development and ICT development index, our results have shown the research community the theoretical underpinning of the relationship, with empirical evidence. Further, this study conceptualizes cybersecurity commitment as a complimentary resource to support e-government development, rooted in the resource-based view. Third, by conceptualizing an e-government development impact model as having a first-order association with cybersecurity commitment, it relates to second-order outcome variables of business usage and economic prosperity. Thus, we show the importance of considering the mediating influence of cybersecurity commitment for analyzing the impact of e-government development on macro parameters. An analysis using the mediator will lead to a fuller understanding of the paths.

From a practical standpoint, our study makes three contributions. Firstly, the findings highlight the positive and strong influence of e-government development on the cybersecurity commitment. This essentially tells the practitioners and policymakers how the higher level of e-government development will result in higher cybersecurity commitment and justify investments in the development of e-government. However, prior research warns about a few outliers. Nevertheless, the acceptability of e-government service, better cyberinfrastructure, and quality of the human capital will have a considerable impact on the various dimensions of cybersecurity commitment, especially technical sophistication, awareness about security threats, and capacity building. Second, the positive and strong influence of cybersecurity commitment on business usage. This could be a useful consideration for policymakers and practitioners, as diffusion of ICT into business is happening at a higher pace, every nation needs to move from cybersecurity to cyber resilience (Wilson, 2014; Wilding, 2016). Thus, policymakers and practitioners need to put serious efforts into building the cyber resilience culture in a country. Third, the influence of e-government development on macroeconomic conditions through the double mediation effect of a nation's cybersecurity commitment and business usage. Even though this association is weaker, this can initiate an action plan for the policymakers and practitioners to check whether the improvement of cybersecurity measures will drive business usage and drive macroeconomic conditions. A qualitative assessment of this relationship has been studied previously, in which the authors have assessed how the legal and political strategy would fight against cybercrime (Mohamed, 2013). Mitigating cyber-crimes is one way of saving cost in billions of USD for the nation and thus keeping a tab on the budget balance. One interesting finding from this study deserves mention was the insignificant impact of e-government development on macroeconomic conditions. This could be because of the fact that the development of e-government is still not a significant contributor in improving the macro-economic conditions, despite the whole world moving to the digital economy.

There are two limitations to this study. One is the use of secondary data, where we have less control over the definitions and measures of the construct. Thus, primary data collection has always the upper hand in this area. However, as the research question requires a large amount of cross-country data where primary data collection is not feasible, the use of secondary data could be justified. The other limitation is in the use of the control variable. Previous researchers have considered the income status of a country as one of the variables to influence e-government development. However, there are mixed findings regarding this (Zhao et al., 2010)

and not included in the current study. Despite these limitations, our research is one of the few studies that shed light on the macroeconomic perspective, analyzing the realization of cyber commitment through e-government development directly influencing nation-level parameters (business usage and economic prosperity).

The current study offers several directions for future research. As the present study focuses on the cross-sectional data, the association among these macro parameters is better understood using the panel data. Previous researchers have indicated this so that temporal trends could be better captured. For instance, a global pandemic like Covid-19 could influence the business and economy. Such breaks could be included in the model, and the real impact of cybersecurity commitment on business usage and economic prosperity could be assessed. Thus, researchers could extend this model. Another avenue for future research is the inclusion of interaction variables other than mediation. For instance, it has been pointed out that there are few outliers who may be low in e-government development and still high in cybersecurity commitment. One possible explanation is the role of governance could play, and it could moderate the relationship between e-government development and cybersecurity commitment. For instance, the state-level quality of the e-government websites in terms of cybersecurity could vary in a country in correspondence with political stability or agenda (Zhao et al., 2010). Thus, our study could be extended by formulating moderated models and empirically evaluate the relationships as contingent variables could explain the insignificant relationships.

CONCLUDING REMARKS

Despite an extensive recognition of the need for cybersecurity in a country, both research and practitioner communities know relatively little about the association between e-government development and cybersecurity commitment, how it effectively assists businesses and enhances economic prosperity. As an initial step, the present study highlights the pivotal role of e-government and its positive influence on cybersecurity commitment at the country level. Further, this study also recognizes the role of cyber commitment to boost ICT usage in business, the use of e-government services for the profitability of the firm, and effectively influence economic prosperity. We have constructed a theoretical model and validated it using publicly available archival data. Further, we hypothesized and demonstrated empirically the direct as well as indirect relationships among e-government development, cybersecurity commitment, business usage, and economic prosperity. To summarize, our study unearths the role of a nation's cybersecurity commitment and how it is associated with the other macro parameters in a country.

REFERENCE

Ali, A. A., & Zamri Murah, M. (2018). Security Assessment of Libyan Government Websites. *2018 Cyber Resilience Conference (CRC)*, 1–4. https://doi.org/10.1109/CR.2018.8626862

Ali, O., & Osmanaj, V. (2020). The role of government regulations in the adoption of cloud computing: A case study of local government. *Computer Law & Security Review*, *36*, 105396. https://doi.org/10.1016/j.clsr.2020.105396

Allers, M. A. (1994). *Administrative and compliance costs of taxation and public transfers in theNetherlands.* (Groningen theses in economics, management & organization). Groningen: Wolters-Noordhoff.

Alshehri, M., & Drew, S. (2010). Implementation of e-government: Advantages and Challenges. *CONFERENCE PROCEEDINGS*, 8.

Alsmadi, I., & Shanab, E. A. (2016). e-government website security concerns and citizens' adoption. *Electronic Government, an International Journal*, *12*(3), 243. https://doi.org/10.1504/EG.2016.078417

Arendsen, R., Peters, O., ter Hedde, M., & van Dijk, J. (2014). Does e-government reduce the administrative burden of businesses? An assessment of business-to-government systems usage in the Netherlands. *Government Information Quarterly*, *31*(1), 160–169. https://doi.org/10.1016/j.giq.2013.09.002

Arendsen, R., van Engers, T. M., & Schurink, W. (2008). Adoption of High Impact Governmental eServices: Seduce or Enforce? In M. A. Wimmer, H. J. Scholl, & E. Ferro (Eds.), *Electronic Government* (Vol. 5184, pp. 73–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85204-9_7

Awoleye, O. Michael, Ojuloge, B., & Siyanbola, W. O. (2012). Technological assessment of e-government web presence in Nigeria. *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV '12*, 236. https://doi.org/10.1145/2463728.2463774

Awoleye, Olusesan M., Ojuloge, B., & Ilori, M. O. (2014). Web application vulnerability assessment and policy direction towards a secure smart government. *Government Information Quarterly*, *31*, S118–S125. https://doi.org/10.1016/j.giq.2014.01.012

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Bissyandé, T. F., Ouoba, J., Ahmat, D., Ouédraogo, F., Béré, C., Bikienga, M., Sere, A., Dandjinou, M., & Sié, O. (2016). Vulnerabilities of Government Websites in a Developing Country – the Case of Burkina Faso. In R. Glitho, M. Zennaro, F. Belqasmi, & M. Agueh (Eds.), *E-Infrastructure and e-Services* (Vol. 171, pp. 123–135). Springer International Publishing. https://doi.org/10.1007/978-3-319-43696-8_14

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., & Boss, R. W. (2009). If someone is watching, I'll do what I'm asked: Mandatoriness, control, and information security. *European Journal of Information Systems*, *18*(2), 151–164. https://doi.org/10.1057/ejis.2009.8

Bretschneider, S. (2003). Information Technology, e-government, and Institutional Change. *Public Administration Review*, *63*(6), 738–741. https://doi.org/10.1111/1540-6210.00337

Burt, D. (2014). *Cyberspace 2025 today's decisions, tomorrow's terrain*. Microsoft.

Cho, Y. H., & Choi, B.-D. (2004). e-government to Combat Corruption: The Case of Seoul Metropolitan Government. *International Journal of Public Administration*, *27*(10), 719–735. https://doi.org/10.1081/PAD-200029114

Clark, D. D., Gillett, S. E., Lehr, W., Sirbu, M. A., & Fountain, J. E. (2003). Local government stimulation of broadband: Effectiveness, e-government, and economic development. E-

Government, and Economic Development (January 2003). KSG Working Papers Series RWP03-002.

Cyber-crime report (2018). https://www.herjavecgroup.com/wpcontent/uploads/2018/12/CV-HG-2019-Official-Annual-Cybercrime-Report.pdf (received on October 22, 2019)

Dash, S., & Pani, S. K. (2016). E-Governance Paradigm Using Cloud Infrastructure: Benefits and Challenges. *Procedia Computer Science*, *85*, 843–855. https://doi.org/10.1016/j.procs.2016.05.274

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147. https://doi.org/10.2307/2095101

Dynes, S., Goetz, E., & Freeman, M. (2007). Cyber Security: Are Economic Incentives Adequate? In E. Goetz & S. Shenoi (Eds.), *Critical Infrastructure Protection* (Vol. 253, pp. 15–27). Springer US. https://doi.org/10.1007/978-0-387-75462-8_2

e-government survey. (2018). *United Nations e-government Survey 2018—Now available— Operating Unit on Policy-Driven Electronic Governance*. https://egov.unu.edu/news/news/un-egov-survey-2018.html

Ellefsen, I. (2014, May). The development of a cyber security policy in developing regions and the impact on stakeholders. In 2014 IST-Africa Conference Proceedings (pp. 1-10). IEEE.

Evers, T. (2010). *Doing business and making peace.* UI Occasional Papers (No. 3). https://www.ui.se/globalassets/ui.se-eng/publications/ui-publications/doing-business-and-making-peace-min.pdf

Gallouj, F., & Savona, M. (2009). Innovation in services: A review of the debate and a research agenda. *Journal of Evolutionary Economics*, *19*(2), 149–172. https://doi.org/10.1007/s00191-008-0126-4

Global cybersecurity report (2018). https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity index.aspx

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Halchin, L. E. (2004). Electronic government: Government capability and terrorist resource. *Government Information Quarterly*, *21*(4), 406–419. https://doi.org/10.1016/j.giq.2004.08.002

Ismailova, R. (2017). Web site accessibility, usability and security: A survey of government web sites in Kyrgyz Republic. *Universal Access in the Information Society*, *16*(1), 257–264. https://doi.org/10.1007/s10209-015-0446-8

ITU. (2014). *Global information technology report*. http://reports.weforum.org/global-information-technology-report-2014/

ITU. (2016). *Global information technology report*. https://www.weforum.org/reports/the-global-information-technology-report-2016

ITU. (2019). *Network readiness report*. https://networkreadinessindex.org/wp-content/uploads/2020/03/The-Network-Readiness-Index-2019-New-version-March-2020-2.pdf

ITU Cyber. (2014). *Global cyber security index*. https://www.itu.int/en/ITU-D/Cybersecurity/Pages/GCI-2014.aspx

ITU Cyber. (2017). *Global cyber security Index*. https://www.itu.int/en/ITU-D/Cybersecurity/Pages/GCI-2017.aspx

ITU Cyber. (2018). *Global cyber security Index*. https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf

Johnson, M. E., & Goetz, E. (2007). Embedding information security into the organization. IEEE Security & Privacy, 5(3), 16-24.

Johnson, A. M. (2009). Business and Security Executives Views of Information Security Investment Drivers: Results from a Delphi Study. *Journal of Information Privacy and Security*, *5*(1), 3–27. https://doi.org/10.1080/15536548.2009.10855855

Kock, N. (2015e). How likely is Simpson's paradox in path models? International Journal of eCollaboration, 11(1), 1-7.

Kock, N., & Gaskins, L. (2016). Simpson's paradox, moderation, and the emergence of quadratic relationships in path models: An information systems illustration. International Journal of Applied Nonlinear Science, 2(3), 200-234.

Kock, N. (2020). WarpPLS User Manual: Version 7.0. Laredo, TX: ScriptWarp Systems

Knickrehm, M., Berthon, B., & Daugherty, P. (2015). *Digital disruption: The growth multiplier: Accenture strategy*. https://www.accenture.com/_acnmedia/pdf-14/accenture-strategy-digital-disruption-growth-multiplier-brazil.pdf

Krishnan, S., Teo, T. S. H., & Lim, V. K. G. (2013). Examining the relationships among e-government maturity, corruption, economic prosperity and environmental degradation: A cross-country analysis. *Information & Management*, *50*(8), 638–649. https://doi.org/10.1016/j.im.2013.07.003

Kuan, K. K. Y., & Chau, P. Y. K. (2001). A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Information & Management*, *38*(8), 507–521. https://doi.org/10.1016/S0378-7206(01)00073-8

Lee, Jooho, Kim, H. J., & Ahn, M. J. (2011). The willingness of e-government service adoption by business users: The role of offline service quality and trust in technology. *Government Information Quarterly*, *28*(2), 222–230. https://doi.org/10.1016/j.giq.2010.07.007

Lee, Jungwoo. (2010). 10 year retrospect on stage models of e-government: A qualitative meta-synthesis. *Government Information Quarterly*, *27*(3), 220–230. https://doi.org/10.1016/j.giq.2009.12.009

Lee, Jungwoo, & Kim, J. (2007). Grounded theory analysis of e-government initiatives: Exploring perceptions of government authorities. *Government Information Quarterly*, *24*(1), 135–147. https://doi.org/10.1016/j.giq.2006.05.001

Lee, Y., & Larsen, K. R. (2009). Threat or coping appraisal: Determinants of SMB executives' decision to adopt anti-malware software. *European Journal of Information Systems*, *18*(2), 177–187. https://doi.org/10.1057/ejis.2009.11

Lewis, J. A. (Ed.). (2003). *Cyber security: Turning national solutions into international cooperation*. CSIS Press, Center for Strategic and International Studies.

Luiijf, E., Besseling, K., & Graaf, P. D. (2013). Nineteen national cyber security strategies. *International Journal of Critical Infrastructures*, *9*(1/2), 3. https://doi.org/10.1504/IJCIS.2013.051608

Luna-Reyes, L. F., & Gil-Garcia, J. R. (2011). Using institutional theory and dynamic simulation to understand complex e-Government phenomena. Government Information Quarterly, 28(3), 329-345.

Maimon, D., & Louderback, E. R. (2019). Cyber-dependent crimes: an interdisciplinary review. Annual Review of Criminology, 2, 191-216.

Malone, T. W., Yates, J., & Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM*, *30*(6), 484–497. https://doi.org/10.1145/214762.214766

Mathews, J. A. (2002). A resource-based view of Schumpeterian economic dynamics. *Journal of Evolutionary Economics*, *12*(1–2), 29–54. https://doi.org/10.1007/s00191-002-0106-z

McAfee. (2014). *Net losses: Estimating the Global Cost of Cybercrime*. https://csis-website-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/attachments/140609_McAfee_PDF.pdf

Miles, I. (2004). Innovations in services. In *The Handbook of Innovations*. J. Fagerberg, D.C. Mowery, R.R. Nelson (Eds.). Oxford University Press.

Moen, V., Klingsheim, A. N., Simonsen, K. I. F., & Hole, K. J. (2007). Vulnerabilities in e-governments. *International Journal of Electronic Security and Digital Forensics*, *1*(1), 89. https://doi.org/10.1504/IJESDF.2007.013595

Mohamed, D. binti. (2013). Combating the threats of cybercrimes in Malaysia: The efforts, the cyberlaws and the traditional laws. *Computer Law & Security Review*, *29*(1), 66–76. https://doi.org/10.1016/j.clsr.2012.11.005

Ndou, V. D. (2004). E - Government for Developing Countries: Opportunities and Challenges. *The Electronic Journal of Information Systems in Developing Countries*, *18*(1), 1–24. https://doi.org/10.1002/j.1681-4835.2004.tb00117.x

Osula, A.-M., & Kaska, K. (2013). *NATO, National Cyber Security Strategy Guidelines*. CCDCOE, Editor.Tallinn, Estonia. https://ccdcoe.org/uploads/2018/10/NCSS-Guidelines_2013.pdf

Pearl, J. (2009). Causality: Models, reasoning, and inference. Cambridge, England: Cambridge University Press

Porter, M. E. (1990). *The Competitive Advantage of Nations*. Free Press: New York, N.Y.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Ramon Gil-Garcia, J., Chengalur-Smith, I., & Duchessi, P. (2007). Collaborative e-Government: impediments and benefits of information-sharing projects in the public sector. European Journal of Information Systems, 16(2), 121-133.

Ravi Sen & Sharad Borle (2015) Estimating the Contextual Risk of Data Breach: An Empirical Approach, Journal of Management Information Systems, 32:2, 314-341, DOI: 10.1080/07421222.2015.1063315

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *Smartpls3*. SmartPLS GmbH. http://www.smartpls.com

Rohleder, J. J., & Jup, V. (2004). EGovernment leadership; high performance, maximum value. *Accenture*.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102. https://doi.org/10.1086/261725

Rowley, J. (2011). e-government stakeholders—Who are they and what do they want? *International Journal of Information Management*, *31*(1), 53–62. https://doi.org/10.1016/j.ijinfomgt.2010.05.005

Selwyn, N., & Brown, P. (2000). Education, nation states and the globalization of information networks. *Journal of Education Policy*, *15*(6), 661–682. https://doi.org/10.1080/02680930010000245

Shackelford, S. J., Fort, T. L., & Prenkert, J. D. (2014). How businesses can promote cyber peace. *University of Pennsylvania Journal of International Law*, *36*(2), 353–431.

Siau, K., & Long, Y. (2005). Synthesizing e-government stage models – a meta-synthesis based on meta-ethnography approach. *Industrial Management & Data Systems*, *105*(4), 443–458. https://doi.org/10.1108/02635570510592352

Srivastava, S. C. (2011). Is e-government providing the promised returns? A value framework for assessing e-government impact. *Transforming Government: People, Process and Policy*, *5*(2), 107–113. https://doi.org/10.1108/17506161111131159

Srivastava, S. C., & Teo, T. S. H. (2007). e-government Payoffs: Evidence from Cross-Country Data. *Journal of Global Information Management*, *15*(4), 20–40. https://doi.org/10.4018/jgim.2007100102

Srivastava, S. C., & Teo, T. S. H. (2008). The Relationship between e-government and National Competitiveness: The Moderating Influence of Environmental Factors. *Communications of the Association for Information Systems*, *23*. https://doi.org/10.17705/1CAIS.02305

Srivastava, S. C., & Teo, T. S. H. (2010). e-government, E-Business, and National Economic Performance. *Communications of the Association for Information Systems*, *26*. https://doi.org/10.17705/1CAIS.02614

Subramani. (2004). How Do Suppliers Benefit from Information Technology Use in Supply Chain Relationships? *MIS Quarterly*, *28*(1), 45. https://doi.org/10.2307/25148624

Symantec. (2014). *Symantec internet Security Threat Report.* https://docs.broadcom.com/doc/istr-14-april-volume-19-en

Tang, T., & Ho, A. T. K. (2019). A path-dependence perspective on the adoption of Internet of Things: Evidence from early adopters of smart and connected sensors in the United States. Government Information Quarterly, 36(2), 321-332.

Teo, H. H., Wei, K. K., & Benbasat, I. (2003). Predicting intention to adopt interorganizational linkages: An institutional perspective. MIS quarterly, 19-49.

Teoh, C. S., & Mahmood, A. K. (2017). National cyber security strategies for digital economy. *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, 1–6. https://doi.org/10.1109/ICRIIS.2017.8002519

Thompson, N., Mullins, A., & Chongsutakawewong, T. (2020). Does high e-government adoption assure stronger security? Results from a cross-country analysis of Australia and Thailand. *Government Information Quarterly*, *37*(1), 101408. https://doi.org/10.1016/j.giq.2019.101408

Tornatzky, L. G., & Fleischer, M. (1990). *The process of technological innovation.* Lexington Young Books.

Twizeyimana, J. D., & Andersson, A. (2019). The public value of e-government – A literature review. *Government Information Quarterly*, *36*(2), 167–178. https://doi.org/10.1016/j.giq.2019.01.001

UN. (2014). *e-government survey*. https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf

UN. (2016). *e-government survey*. https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2016-Survey/e-government%20Survey%202016.pdf

UN. (2018). *e-government survey*. https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/e-government%20Survey%202018_FINAL%20for%20web.pdf

van den Boer, Y., Pieterson, W., & Arendsen, R. (2011). Channel Choice in the G2B Context: Towards a Research Agenda. In T. Skersys, R. Butleris, L. Nemuraite, & R. Suomi (Eds.), *Building the e-World Ecosystem* (Vol. 353, pp. 14–27). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-27260-8_2

Von Haldenwang, C. (2004). Electronic Government (e-government) and Development. *The European Journal of Development Research*, *16*(2), 417–432. https://doi.org/10.1080/09578810420002208 86

Wagstaff, K., Eng, J., & DeLuca, M. (2015). *OPM: 21.5 million people affected by background Check Breach.* https://www.nbcnews.com/ tech/security/opm-hack-security-breach-n389476.

WEF Competitiveness. (2014). *Global Competitiveness Report*. http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2013-14.pdf

WEF Competitiveness. (2016). *Global Competitiveness report*. http://www3.weforum.org/docs/GCR2016-2017/05FullReport/TheGlobalCompetitivenessReport2016-2017_FINAL.pdf

WEF Competitiveness. (2019). *Global Competitiveness report*. http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf

West, D. M. (2004). e-government and the Transformation of Service Delivery and Citizen Attitudes. *Public Administration Review*, *64*(1), 15–27. https://doi.org/10.1111/j.1540-6210.2004.00343.x

Wiggins, R. R., & Ruefli, T. W. (2005). Schumpeter's ghost: Is hypercompetition making the best of times shorter? *Strategic Management Journal*, *26*(10), 887–911. https://doi.org/10.1002/smj.492

Wilding, N. (2016). Cyber resilience: How important is your reputation? How effective are your people? *Business Information Review*, *33*(2), 94–99. https://doi.org/10.1177/0266382116650299

Wilson, N. (2014). Australia's National Broadband Network – A cybersecure critical infrastructure? *Computer Law & Security Review*, *30*(6), 699–709. https://doi.org/10.1016/j.clsr.2014.09.003

Wong, W., & Welch, E. (2004). Does e-government Promote Accountability? A Comparative Analysis of Website Openness and Government Accountability. *Governance*, *17*(2), 275–297. https://doi.org/10.1111/j.1468-0491.2004.00246.x

Yildiz, M. (2007). e-government research: Reviewing the literature, limitations, and ways forward. *Government Information Quarterly*, *24*(3), 646–665. https://doi.org/10.1016/j.giq.2007.01.002

Zhao, F., Wallis, J., & Singh, M. (2015). e-government development and the digital economy: A reciprocal relationship. *Internet Research*, *25*(5), 734–766. https://doi.org/10.1108/IntR-02-2014-0055

Zhao, J. J., Zhao, S. Y., & Zhao, S. Y. (2010). Opportunities and threats: A security assessment of state e-government websites. *Government Information Quarterly*, *27*(1), 49–56. https://doi.org/10.1016/j.giq.2009.07.004

Zheng, D., Chen, J., Huang, L., & Zhang, C. (2013). E-government adoption in public administration organizations: integrating institutional theory perspective and resource-based view. European Journal of Information Systems, 22(2), 221-234.

Zhou, Z., & Hu, C. (2008). *Study on the e-government Security Risk Management*. 6.

Zhu, K. (2004). The Complementarity of Information Technology Infrastructure and E-Commerce Capability: A Resource-Based Assessment of Their Business Value. *Journal of Management Information Systems*, *21*(1), 167–202. https://doi.org/10.1080/07421222.2004.11045794

Krishna & Sebastian	e-government, cybersecurity, and national performance

Zhu, K., Kraemer, K. L., & Xu, S. (2002). *A Cross-Country Study of Electronic Business Adoption Using the Technology-Organization-Environment Framework*. 14.

DECISION SCIENCES INSTITUTE
Information Technology Focused Executive Education: What Employers Need

Alanah Mitchell
Department of Information Management and Business Analytics
College of Business and Public Administration
Drake University
Email: alanah.mitchell@drake.edu

ABSTRACT

Today's Information Technology (IT) leaders are looking to executive education to develop future technology leaders. This paper looks to identify and understand the needs of business leaders in IT as well as identify a means for addressing these needs through executive education. Based on a focus group of organizational leaders and hiring managers, executive education content and delivery needs are identified and introduced.

KEYWORDS: Information Technology Management, Executive Education, Decision Makers, Corporate Training, Workforce Development, Education

INTRODUCTION

Information technology (IT) plays a critical role in the strategy of today's organizations (Pearlson et al., 2019). Because of this, the role of executives and their understanding of IT is more important than ever. Organizational strategies tied to technology, the intricacies and interdependence of business operations and processes, and the complexities of new and emerging technologies further complicate this challenge. Executives in the area of technology are looking for organizational leaders to set priorities and directions, develop commitment, and become adaptive in response to organizational needs and business goals. Furthermore, IT leaders are looking for ways to educate and develop future technology leaders. In fact, previous research has suggested using and mastering new technologies is going to be more critical than ever for organizational employees as digital technologies transform the workplace (Foroughi, 2019). Executive education offers a mutually beneficial solution to these issues.

Executive education allows for integration between the actual practice of management and business and a business school's scholarly research. Specifically, executive education impacts organizations and universities through: 1) engaging practitioners and academics, 2) inspiring innovation and research, and 3) overall university influence (Waller & Fawcett, 2015). AACSB recognizes the value of executive education for business schools and has encouraged programs to explore ways business programs can offer value (Association to Advance Collegiate Schools of Business, 2008). One possible area is the field of IT which, as mentioned above, is a constantly evolving discipline due to the frequent technology and skill changes and developments. This type of partnership can benefit organizations looking to create new learning and growth opportunities for their employees as well as universities through new outreach and partnership opportunities as well as the establishment of research collaborations.

Based on this background, the goal of this work is to identify and understand the needs of executives in the area of IT while also identifying a means for addressing these needs so universities and educators can prepare and provide solutions for organizational leaders. In the research for this paper, a focus group of senior executives, business leaders, and hiring

managers was assembled and interviewed about their perceptions regarding the skills and gaps they see as missing in their existing workforce. The interviewees held various roles within the IT organization, including Chief Technology Officer (CTO), Chief Information Security Officer (CISO), representatives from Human Resources (HR), and others. The following sections present the key ideas and themes that emerged from this focus group integrated with existing research in each area.

EXECUTIVE EDUCATION NEEDS

Today's employees are expected to be proficient in the area of technology in order to use various organizational systems, follow organizational processes, or collaborate electronically. This is not only true for employees within an IT department as nearly all employees today need to have a digital fluency (Colbert et al., 2016). Indeed, previous research has identified twelve competencies needed by a digital workforce (Petter et al., 2018). These competencies include: accountability, adaptability, business acumen, collaboration/teamwork, conceptual thinking, decisiveness, digital literacy, global mindset, innovation, openness to learning, results orientation, and risk taking.

Once organizations have taken the time to identify and hire digitally fluent talent that meets the needs of their organization, they want to be able to retain these employees. In fact, attracting and retaining talent is a top priority for organizations and one of the most basic issues that leaders struggle with (Keller & Meaney, 2017). Furthermore, research has identified career development as one important tool that can be offered to employees as a part of a retention plan (Pflügler et al., 2018). Executive education offers a solution to both of these challenges. Specifically, executive education programs allow for organizations to develop the identified skills and behaviors they need in their workforce while also allowing for existing employees to improve their skills and behaviors as a part of their own career development.

Based on the focus group comprised of organizational employers and leaders for this study, two key ideas stood out. First, the leaders mentioned being interested in executive education for the development of new behaviors as new skills training is something they could develop in house. Secondly, the leaders were interested in the development of T-shaped leaders (Boynton, 2011), or employees with broad knowledge and skills as opposed to highly specialized skills in one particular field (i.e., I-shaped). With these two key ideas in mind, four areas of employee development needs emerged: 1) technology strategy, 2) data driven, 3) connecting and communicating, and 4) design thinking (see Figure 1). The following sections outline these specific needs in more detail.

Figure 1: Technology Focused Executive Education Areas of Need

Need #1: Technology Strategy

Established companies who are working in non-technology industries are being transformed by recent technology advancements. Many of the leaders in this focus group were concerned their employees did not see themselves as employees of a technology company. However, technology is a key component to an overall business strategy today. In fact, the argument has been made that every company is a technology company (Mims, 2018). While the overall business strategy should drive the organizational and technology strategies of a company, the organizational and technology strategies should support the business strategy (Pearlson et al., 2019).

An executive education program designed to teach senior leaders about the importance of understanding technological innovations and exploring successful cases of technology transformation within an organization would be helpful to change the viewpoints and behaviors of existing employees in well-established organizations. These employees need to learn how to increase their digital proficiency and see how technology can be used to deliver new organizational solutions.

Who Should Attend:

Managers and team leaders from all organizational levels looking to utilize technology to support the organization or as a competitive advantage.

Need #2: Data Driven

Businesses today have an overwhelming amount of data at their disposal. The concept “big data” has been coined as a way to discuss this abundance of data. Organizational leaders are interested in using this wealth of data to make better and smarter decisions or become data driven decision makers. However, employees need to think broadly about the kind of data that is being tracked versus the kind of data that is worth being tracked and how this can and should relate to the business plan. In addition, today’s employees need to understand the different types of business analytics methods and tools available as well as how to communicate the findings from this type of analysis.

An executive education program designed to teach employees about business analytics (descriptive, predictive, and prescriptive) as well as the methodologies and tools available in this space would be helpful for transitioning current employees into data driven leaders. Students in this type of program would learn the behaviors necessary to become curious leaders and identify existing data within their organization. They would then be able to identify valuable data through the exploration of successful cases of business analytics. In addition, executive education on data driven leadership would need to focus on the ethics of this topic as well as how to clearly communicate their message upstream within an organization.

Who Should Attend:

Managers and team leaders from all organizational levels interested in using data to drive their decision making.

Need #3: Connecting and Communicating

Leaders today want employees who are comfortable connecting outside their office walls. This can be through the development of partnerships with other organizations (e.g., Barnes & Noble and Starbucks, Uber and Spotify, etc.), utilizing resources outside their organization (e.g., Software as a Service SaaS, Amazon Web Services AWS, etc.), or even developing coopetition arrangements (e.g., Microsoft and Intel, the Kindle app and the iPad, etc.). The focus group business leaders in this study want to encourage their employees to leverage more resources outside of their organization. These high demand employees who can work with people outside of the organization in diverse, complex teams, using different technology systems are highly desirable. Global virtual teams offer a number of challenges related to geographic locations, times zones, diverse cultures, and technology issues (Pinsonneault & Caya, 2005). Employees are often not trained in the best ways to deal with the various issues that can occur in this type of work arrangement (Davis & Zigurs, 2008). Furthermore, leaders in this environment need to be proactive when identifying challenges that might disturb or halt business productivity (Thomas et al., 2007).

Executive education courses can provide participants with the behaviors and skills needed to identify and develop partnerships and be successful. This type of technology supported, collaborative, virtual team environment will be helpful for connecting and communicating within and outside an organization. Employees need to learn strategies for becoming successful virtual team leaders and members when connecting outside office walls.

Who Should Attend:

Managers or leaders working with partners or vendors from other organizations and/or other locations, virtual team leaders or members of virtual teams.

Need #4: Design Thinking

Businesses are interested in brainstorming to find new ideas or identify problems that do not yet have solutions. They are then interested in solving these problems. The concept of design thinking can help in both generating ideas and problem solving through five key steps: 1) empathize, 2) define, 3) ideate, 4) prototype, and 5) test (Himsworth, 2018). Employees who are trained in design thinking have the behavior of a lifelong learner in that they are curious and open to new ideas and solutions. IT leaders are looking for employees to navigate the consequences of the changing digital economy and frequent changes in addition to leading technological innovation initiatives. Behaviors and skills tied to design thinking can help with this.

An executive education course which teaches business leaders about the process of design thinking as well as how to be curious in problem discovery and innovative in problem solving would be very valuable. Employees exposed to this type of education would be able to analyze current business processes and practices within their organization and think differently about how and why these practices are in place. In addition, employees in this type of program could learn to be curious as they imagine and brainstorm while developing and designing new products or solutions.

Who Should Attend:

Managers and team leaders from all organizational levels looking to increase their skills in brainstorming and problem solving.

TECHNOLOGY FOCUSED EXECUTIVE EDUCATION DELIVERY

Once leaders have determined the gaps in their current workforce and the priorities, they have in educating their employees, the next step is to identify the means for delivery. Previous research on executive education has identified a few different strategies for successful re-tooling. The following discussion presents three proven methods for executive education: 1) in-house, 2) off-campus, and 3) omni-learning (see Figure 2).

Figure 2: Executive Education Delivery Options

Delivery #1: In-house

One option for the delivery of executive education is to bring outside educators and facilitators into an organization to work with executives within their own organization and environment. This type of education would team employees up with their co-workers as they work and explore various challenges and topics of interest. Previous research suggests that this type of in-house education is preferable as it allows for customization to the organization and its individual challenges (Hassell, 2016). Specifically, this type of strategy can be useful if the executives need the personal experience and customization in order to benefit from the learning process.

Delivery #2: Off-campus

An alternative option to executives staying within their own organization in classes filled with their co-workers is to move the learning off the corporate campus and into a neutral learning space (e.g., such as a university campus). Executive education programs that take place on college campuses generally enroll a number of students from different industries and organizational roles, teaming them up to work together during the learning process.

This type of off-campus approach allows for executives to learn from the executive education faculty as well as their peers in the classroom. Another benefit from working with practitioners from other organizations is that executives can see how other organizations operate and learn from them instead of just learning about concepts and theories (Clinebell & Clinebell, 2008).

Delivery #3: Omni-learning

Finally, the concept of omni-learning has been introduced as a way for executives to learn by blending face-to-face and online tools (Auricchio & Káganer, 2015). This research has suggested that current executive development experiences are time constrained and challenging and executives are not prepared for e-learning alone. Therefore, this idea of combining online and in-person learning might be an option for achieving better results. Specifically, omni-learning encourages the executive student to learn every day and everywhere with the integration of journaling and data driven insights engrained into the executive's day to day life.

CONCLUDING COMMENTS

The goal of this paper was to provide guidance on the knowledge executives are looking for in relation to IT, as well as how to deliver this knowledge. A focus group was used to identify four topics of interest for executive education in relation to IT. In addition, previous research was presented as a way of illustrating how these topics might be delivered. Organizational leaders can determine what their specific education and delivery needs might be and create a customized plan of action to meet these needs. What this work has shown is that executive education can be a useful solution for organizations looking to change the behaviors and increase the skills and knowledge of their workforce. Furthermore, executive education can be an appealing tool for employees looking for growth opportunities within their current positions. Finally, academic institutions and educators can benefit from executive education as it creates new outreach and partnership opportunities as well as can potentially lead to the establishment of research collaborations.

This exploration of technology focused executive education and what employers are looking for allows for academic, pedagogical, and practical benefits. In terms of academic implications, faculty and researchers can use the recommendations from this work to gain a better understanding of what executives need and want to have their employees. Furthermore, this work provides a foundation for future education and research studies in this area. In terms of pedagogical implications, executive education faculty might try to utilize some of the different methods of delivery presented above. Finally, in terms of practical implications, this work provides a guide to those looking for executive education in the field of IT, which ultimately benefits the participating organizations.

REFERENCES

Association to Advance Collegiate Schools of Business. (2008). *Final Report of the AACSB International Impact of Research Task Force*. https://www.aacsb.edu/-/media/aacsb/publications/research-reports/impact-of-research.ashx

Auricchio, G., & Káganer, E. (2015). The Road to Omni-Learning: How Digitalization is Changing the Way Executives Learn. *IESE Insight*, *26*(3), 31–38.

Boynton, A. (2011). Are You an "I" or a "T"? *Forbes*. https://www.forbes.com/sites/andyboynton/2011/10/18/are-you-an-i-or-a-t/

Clinebell, S. K., & Clinebell, J. M. (2008). The Tension in Business Education Between Academic Rigor and Real-World Relevance: The Role of Executive Professors. *Academy of Management Learning & Education*, *7*(1), 99–107.

Colbert, A., Yee, N., & George, G. (2016). The Digital Workforce and the Workplace of the Future. *Academy of Management Journal*, *59*(3), 731–739.

Davis, A., & Zigurs, I. (2008). *Teaching and learning about virtual collaboration: What we know and need to know*. 1–9.

Foroughi, A. (2019). Supply chain workforce training: Addressing the digital skills gap. *Proceedings of the 50th Annual Conference of the Decision Sciences Institute*, 546–569.

Hassell, B. (2016). The Problem with Executive Education. *Chief Learning Officer*, 32–37.

Himsworth, J. (2018). Why Design Thinking Should Also Serve as a Leadership Philosophy. *Forbes*. https://www.forbes.com/sites/forbesagencycouncil/2018/07/19/why-design-thinking-should-also-serve-as-a-leadership-philosophy/

Keller, S., & Meaney, M. (2017). *Leading Organizations: Ten Timeless Truths*. Bloomsbury Publishing.

Mims, C. (2018). Every Company is Now a Tech Company. *The Wall Street Journal*. https://www.wsj.com/articles/every-company-is-now-a-tech-company-1543901207

Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2019). *Managing and Using Information Systems: A Strategic Approach* (7th ed.). Wiley.

Petter, S., Barber, D., Barber, C., & Berkley, R. (2018). Using Online Gaming Experience to Expand the Digital Workforce Talent Pool. *MIS Quarterly Executive*, *17*(4), 315–332.

Pflügler, C., Wiesche, M., Becker, N., & Krcmar, H. (2018). Strategies for Retaining Key IT Professionals. *MIS Quarterly Executive*, *17*(4), 297–314.

Pinsonneault, A., & Caya, O. (2005). Virtual teams: What we know, what we don't know. *International Journal of E-Collaboration*, *1*(3), 1–16.

Thomas, D. M., Bostrom, R. P., & Gouge, M. (2007). Making knowledge work in virtual teams. *Communications of the ACM*, *50*(11), 85–90.

Waller, M. A., & Fawcett, S. E. (2015). Executive Education is an Implementation of Supply Chain Management in Collegiate Schools of Business. *Journal of Business Logistics*, *36*(2), 135–138.

DECISION SCIENCES INSTITUTE
Financial Sustainability of IoT Platforms: The Role of Quality and Security

Anurag Garg
University of Florida
Email: anurag.garg@warrington.ufl.edu

Emre Demirezen
University of Florida
Email: emre.demirezen@warrington.ufl.edu

Kutsal Dogan
University of Florida
Email: kutsal.dogan@warrington.ufl.edu

Hsing Kenneth Cheng
University of Florida
Email: kenny.cheng@warrington.ufl.edu

ABSTRACT

The economic significance of Internet of Things (IoT) platforms is evident from the proliferation of smart home assistant platforms like Amazon Alexa and Google Home that center themselves around providing unique and connected experiences to users. We study the dynamics of the financial sustainability of the platform provider and app developers and find that the collaborative nature of the IoT platform has a significant role in the behavior and profitability of all parties. We verify our results in the presence of competition or complementarity among apps and further derive additional interesting results and managerial insights.

KEYWORDS: IoT platform, collaboration, quality improvement, security improvement, differential game

DECISION SCIENCES INSTITUTE

Analyze the Barriers of Circular Economy initiatives in Developing Economies: A Novel Grey-DEMATEL Approach

Chitranshu Khandelwal
Department of Management Studies
Indian Institute of Technology Roorkee, India
Email: ckhandelwal@bm.iitr.ac.in

Mukesh Kumar Barua
Department of Management Studies
Indian Institute of Technology Roorkee, India
Email: baruafdm@iitr.ac.in

ABSTRACT

The Circular Economy (CE) has gathered significant momentum across the globe due to its potential to optimize resource usage and minimize waste. Developing nations lag in the adoption of CE practices due to the presence of several barriers. This study assesses the relationship among fifteen barriers of circular economy initiative in the plastic supply chain in the context of developing nations. A Grey-DEMATEL methodology is used to analyze the causal interactions among selected barriers. Lack of advanced technology & innovation and poor coordination among SC members are identified as prominent barriers.

KEYWORDS: Circular economy, Barriers, Plastic industry, Grey-DEMATEL, India

DECISION SCIENCES INSTITUTE

Helping Business Students to Build their Analytics Portfolios to Improve their Marketability

Cody Baldwin
Brigham Young University-Hawaii
Email: cbaldwin@byuh.edu

ABSTRACT

In the world of analytics, much has been written about the need to prepare students for more technical positions in data science – and less about more business-focused positions in analytics. At BYU-Hawaii, within our business program, we teach students to supplement their business skills (in finance, marketing, human resources, or operations) with analytical skills. As such, in our efforts to improve the marketability of our business students, we are identifying the best ways for them to showcase their skills using online portfolios. For this research, we started by reviewing hundreds of job postings in business analytics to identify the skills most needed across the world. We then evaluated a variety of avenues for students to display these skills.

KEYWORDS: Business Analytics, Analytics Education, Student Portfolios

INTRODUCTION

Business analytics is a growing field, and many business schools are working to build analytics coursework into their curriculum. As we, at Brigham Young University-Hawaii, were preparing our analytics coursework, we asked two important questions: (1) What roles might our business students fill in analytics? (2) What skills are needed to be marketable for those roles? Answers to these questions provide context to the remainder of the paper, which discusses how to help students build a portfolio that demonstrates they have developed those skills in business analytics.

"What roles might our business students fill in analytics?"

There are several possible career paths in data science and analytics. However, emphasis and focus has been on more technical roles, such as data scientists and statisticians, and less on more business-focused roles, such as business analysts and business intelligence analysts. Within our program, we seek to develop the latter of the two. In other words, we aim to build "data savvy professionals" as they are referred to in a report from the McKinsey Digital Institute (Brown, 2011). These are often business managers and analysts that can apply analytics in their business role. The McKinsey report describes them in the following way: they are "people who have basic knowledge of statistics and or machine learning and define key questions data can answer." They would be distinguished from "deep analytical talent," such as data scientists and statisticians, and "supporting technology personnel," such as computer scientists and database administrators. *Figure 1* below shows common job titles in each of those categories.

Figure 1: Categories of Roles in Analytics

Deep Analytical Talent

- Data scientists
- Actuaries
- Statisticians
- Economists

Data Savvy Professionals

- Business and functional managers
- Financial analysts
- Market researchers

Supporting Technology Personnel

- Computer programmers
- Database administrators

"What skills are needed to be marketable for those roles in analytics?"

To identify the skills required for "data savvy professionals," we reviewed hundreds of job postings on Indeed, which aggregates job postings from a variety of other sites, such as major job boards, newspapers, associations, etc. Across those job postings, there were several skills and tools that kept reappearing. (Those skills and tools are introduced later in this paper.)

Once those skills and tools were identified, we begin considering how to provide students with (1) hands-on experience in these tools and (2) ways to demonstrate that experience using online portfolios. Within this paper, details are provided on how we fulfill this second point – to support them in building their own online analytics portfolio.

LITERATURE REVIEW

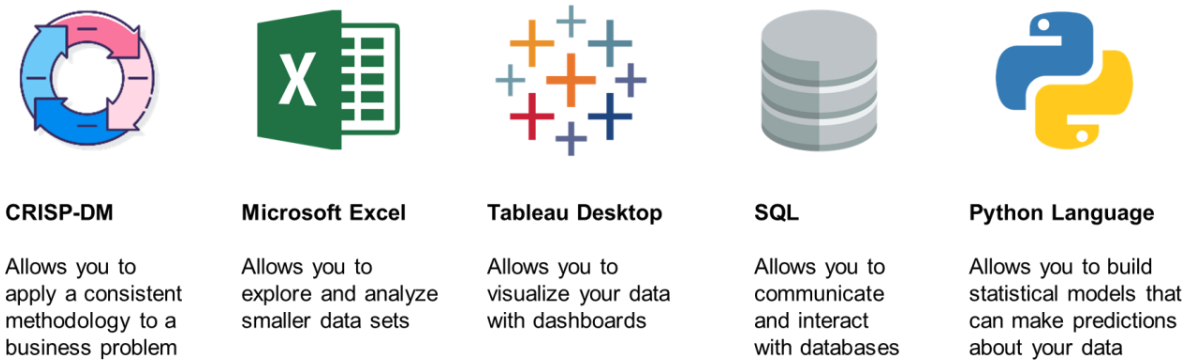
Much has been written about student portfolios and their affect on learning outcomes. However, less has been written about how portfolios can affect a student's job prospects and application performance. This is noteworthy in today's economic climate. A report by Glassdoor (2015) stated the following: "On average, each corporate job opening attracts 250 résumés. Of these candidates, four to six will be called for an interview and only one will be offered a job." Our students need every advantage they can get to help move their resume up a seemingly endless stack. Portfolios can play a role in this. Lievens (2014) argued that in difficult economics times, electronic student portfolios can help improve working employability and mobility. Additionally, a

study performed by Brackett (2017) found that students who engaged in electronic portfolio pedagogy had higher quality interview skills.

SUMMARY OF TOOLS, SKILLS, AND PORTFOLIO ITEMS

Across the hundreds of job postings reviewed during this research, several skills and tools kept reappearing. Those tools and brief explanation of their role in analytics are shown in *Figure 2*.

Figure 2: Popular Business Analytics Tools



Each of these tools are now being used in our business analytics course (BUSM 361), which all business major students are required to take at BYU-Hawaii. Although there were other tools listed in some job postings, such as Microsoft Power BI and R, they were not as common as those listed above.

In addition to identifying tools in high demand, we also prepared learning outcomes ("skills") and ways for demonstrating the achievement of those outcomes using an online portfolio ("portfolio items.") Those skills and portfolio items are summarized in the table below and will be discussed in detail in the next section of this the paper.

Table 1: Summary of Tools, Skills, and Portfolio Items

Tool	Skill	Portfolio Item
CRISP-DM	The ability to apply a standard analytics methodology (CRISP-DM) to a business problem.	Apply the methodology to the assignments below.
Microsoft Excel	The ability to use Microsoft Excel for basic data analysis – includes pivot tables and macros.	Build/share an Excel dashboard.
Tableau	The ability to use Tableau to visualize and communicate data and analysis – and pursue the Tableau Desktop Specialist certification.	Build/share a public Tableau dashboard.
SQL	The ability to use SQL to interact with databases and extract data for further analysis – includes joins.	Build/share a SQL "fiddle."

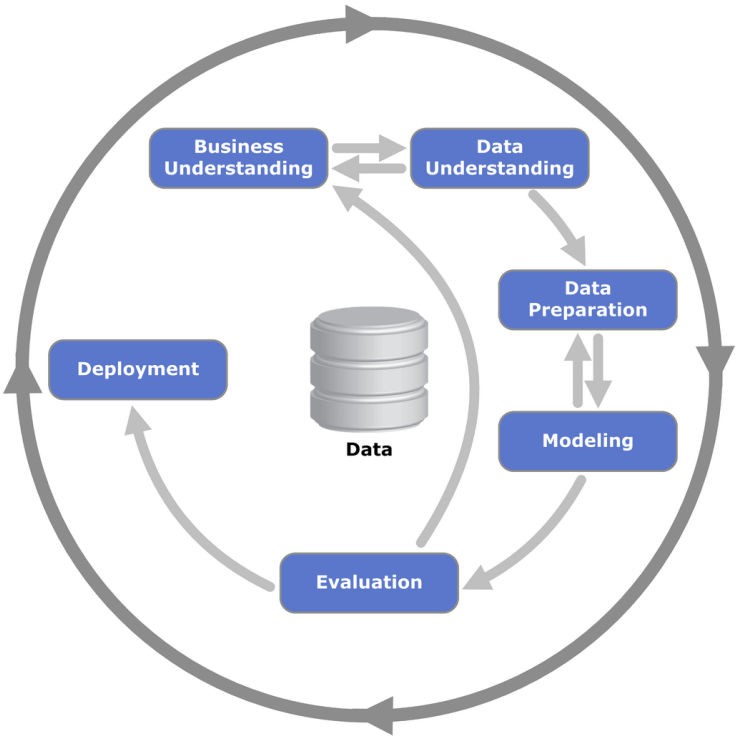
Python	The ability to use Python and Jupyter Notebooks to explore, data, clean data. and build predictive models – includes regression and classification.	Build/share a notebook in Google Colab.

DETAIL OF TOOLS, SKILLS, AND PORTFOLIO ITEMS

CRISP-DM

CRISP-DM, or the cross-industry standard process for data mining, provides a common framework for solving analytics problems in business. There are other frameworks, but research suggests that this is the most common (Mariscal, 2010). A graphical depiction of CRISP-DM is shown in *Figure 3*. An important highlight of CRISP-DM is that it starts with an understanding of the business problem ("business understanding"), which reminds students that they should avoid jumping into an analytics problem and expending resources before understanding the business problem that they hope to solve.

Figure 3: An Overview of the CRISP-DM Process



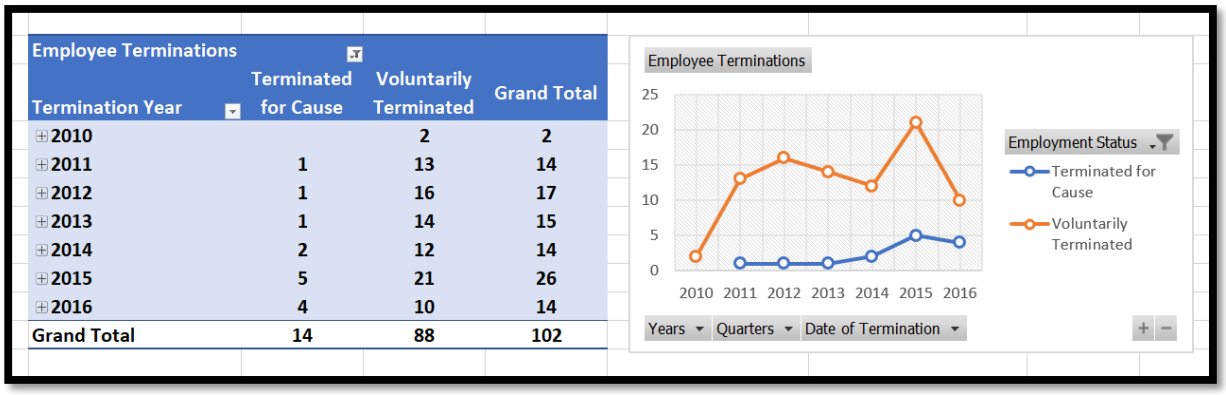
Although CRISP-DM is rarely referenced explicitly on job postings, a solid understanding of the process can help students address more "softer" skills requested on the posting. This could include, as one job posting described it, "using analysis… to identify strategic and operational opportunities to drive significant value for the business." CRISP-DM places an emphasis on first understanding how analytics can support the business, i.e. "drive significant value."

Microsoft Excel

Microsoft Excel was referenced on most job posting for “data savvy professionals.” It continues to be a tool in high demand. In analytics, Excel allows us to explore and analyze smaller data sets. (However, Microsoft is continuing to improve Excel to allow the handling of larger datasets.) Some of the most common Excel features referenced in the job postings were pivot tables and macros, and it is interesting that tools like Tableau (discussed later) use a “pivot-style” approach for interacting with visualizations.

To help students demonstrate their understanding of Microsoft Excel within an online portfolio, students can be provided with a real-world dataset, which they are then required to use pivot tables to explore and extract findings. An example of one of those pivot tables can be seen in *Figure 4*. They can share their pivot tables (or “dashboards”) by uploading the file to Google Drive or OneDrive and sharing a link.

Figure 4: Example Pivot Table in Microsoft Excel

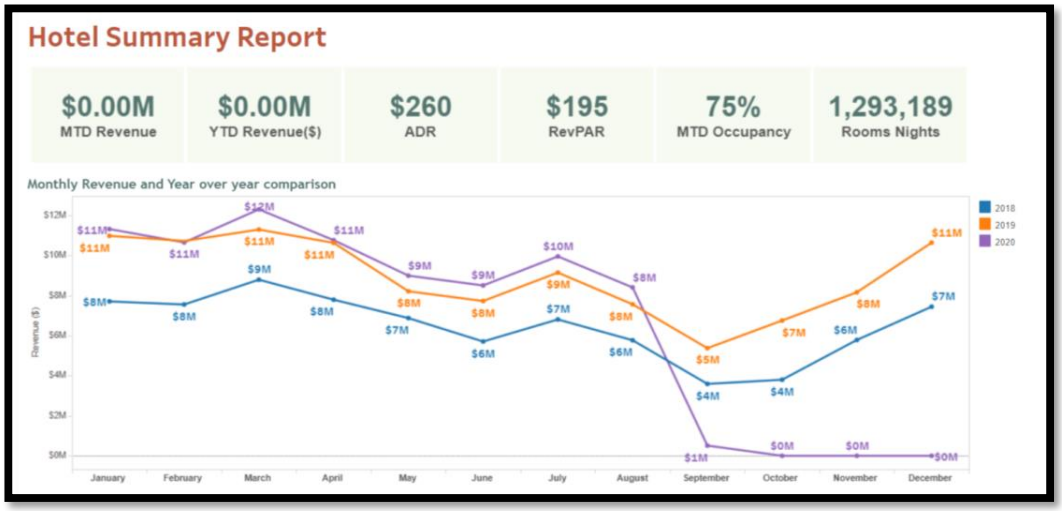


Employee Terminations			
Termination Year	Terminated for Cause	Voluntarily Terminated	Grand Total
2010		2	2
2011	1	13	14
2012	1	16	17
2013	1	14	15
2014	2	12	14
2015	5	21	26
2016	4	10	14
Grand Total	14	88	102

Tableau

Tableau is a popular analytics tool that allows students to visualize their data using dashboards and then share those dashboards with others. To show evidence that they have developed skills with the tool, students can be given a real-world dataset to build a dashboard, and then they can share that dashboard using Tableau Public, a free service for sharing publicly-available Tableau dashboards. *Figure 5* shows an example of a public dashboard (Ruiz, 2020).

Figure 5: Example Public Tableau Dashboard



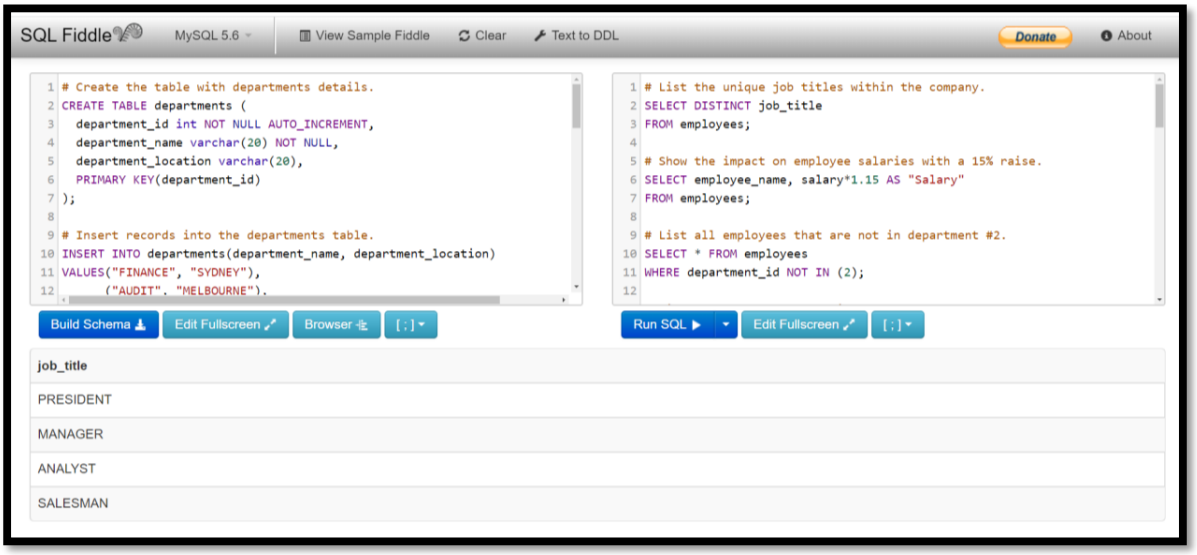
There are two additional benefits to using Tableau to teach business analytics: (1) students can obtain a free license, and (2) students can also pursue the Tableau Desktop Certification, which, in addition to sharing a Tableau dashboard in their portfolio, can help students improve their marketability. It should also be noted that a similar visualization tool – Microsoft Power BI – was referenced on many job postings but not as frequently as Tableau. If appropriate, Power BI could be considered as a replacement for Tableau within your curriculum. However, a limitation of Microsoft Power BI is that there is no Mac version of the software.

SQL

The structured query language (SQL), another analytics tool requested on job postings allows us to communicate and interact with databases. In analytics, most of the data that we use is drawn from databases, so the ability to interact with those databases is critical.

One way to help students demonstrate their SQL skills is to have to them create a “fiddle” using SQL Fiddle (www.sqlfiddle.com). “Fiddles” are primarily used by database administrators to share SQL issues that they are facing with other experts, who can help them find a solution. However, it can also be used to by students to demonstrate their understanding of SQL. To create a “fiddle,” students write SQL to build their database schema (e.g. create their tables and add records) and then write SQL to pull information from those database (i.e. using SELECT statements). As students are building their “fiddle,” a URL is created, which is unique to them, and then updated with a version number as changes are made. Students can then share that unique URL within their portfolio.

Figure 6: Example SQL “Fiddle”

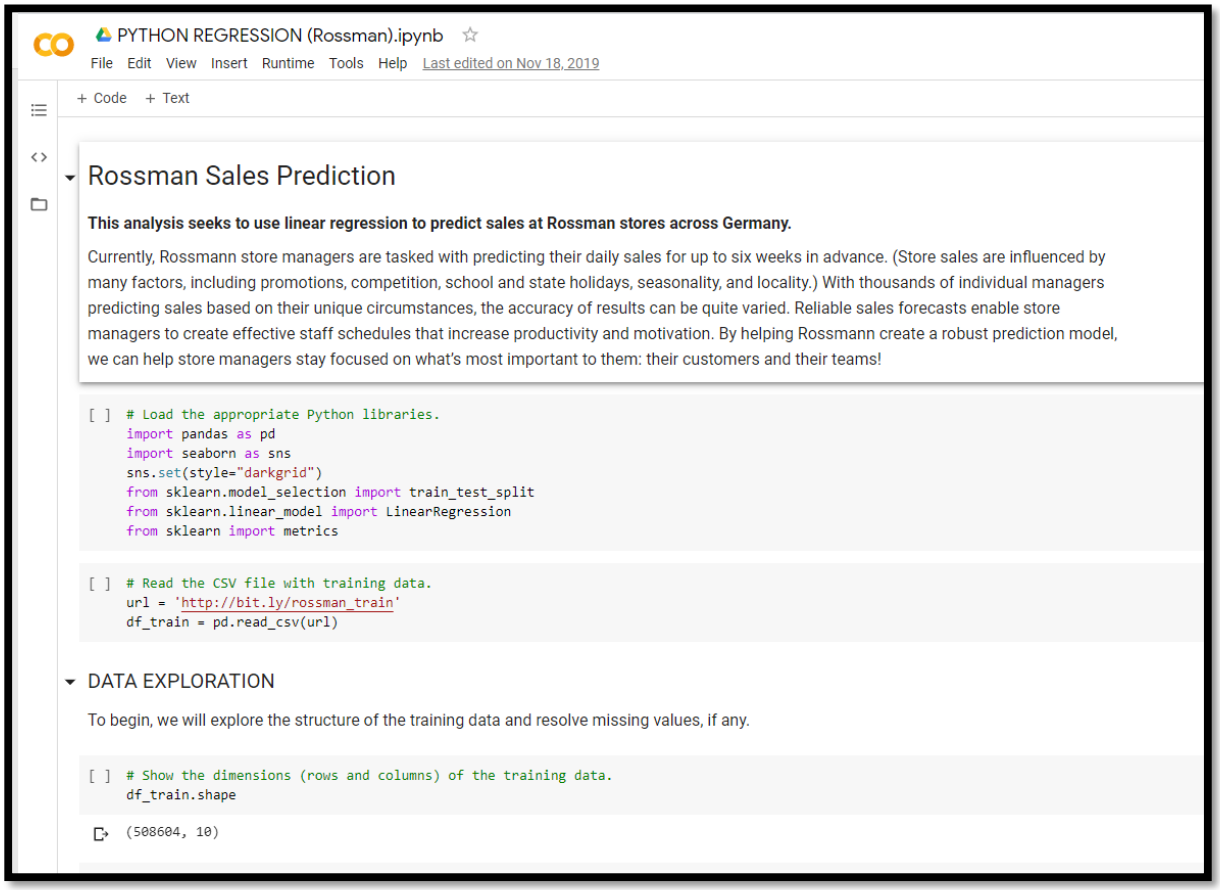


Python

Python, another popular tool on job postings for “data-savvy professionals,” allows us to build statistical models that can make predictions about our data. As a note, another language was often seen on job postings – R – but not nearly as frequently as Python. Additionally, Python has the highest score on the “PYPL PopularitY of Programming Language Index,” which analyzes how often people search for tutorials of various programming languages on Google. According to the index, Python represented the an estimated 31.6% of these searches (the highest on the list), whereas R only represented 3.86% (Carbonnelle, 2020).

In order for students to demonstrate their skills with Python, they can create a Jupyter Notebook using Google Colab. Jupyter Notebooks, which have become very popular in analytics, can be thought of as a cross between a Word document and a Python code file. It allows us to mix text and code into a useful report that can be understood by more than just technical professionals. Google Colab is essentially a Jupyter Notebook in the cloud, which allows users to avoid installing Python and Jupyter on their own machine. This helps students to get up-and-running more quickly. *Figure 7* below shows an example of a notebook in Google Colab.

Figure 7: Example Jupyter Notebook in Google Colab



METHODS FOR SHARING STUDENT PORTFOLIOS

With all of the “portfolio items” referenced in earlier in this paper, a link can be generated that allows students to share their work. The next logical question becomes, “where do they put those links?” Students should be encouraged to (1) consider adding a shortened version of the link to their resume, (2) share the link on a LinkedIn post, and (3) add the link as a “project” within LinkedIn. Let’s discuss each of those sharing options further.

Adding Links on their Resume

If a student wants to add a link to one of their portfolio items on their resume, they should provide context as to what the link includes and avoid using the long form of the link. They can shorten the link by using a URL shortening tool, like *bit.ly*. This is critical as a short URL is easier to type into a browser.

Sharing Links on a LinkedIn Post

By sharing the link to their portfolio item on a LinkedIn post, students can accomplish two objectives: (1) It can demonstrate to the connections in their network that they are developing analytics skills, which might cause some of those connections to keep the student in mind for future job opportunities. (2) It can provide an opportunity for their connections to praise them on

their new skills, which can keep students motivated and excited about continuing to develop those skills.

Adding Links as a "Project" on their LinkedIn Profile

On LinkedIn, a "project" can stay permanently on a student's profile, whereas a "post" may fade away over time (or be out of view). It can also allow students to provide more detail about the portfolio item and the business context.

CONCLUSION

As a concluding thought for business programs that are considering these tools and skills, it is important to understand that most of these will be new for students. As such, if we are not careful, considerable time both inside and outside of class can be spent on troubleshooting installation and other technical issues. At BYU-Hawaii, we have found success by setting up a lab with computers that have the tools already installed and that is managed by one or more teaching assistants that took the analytics coursework in previous semesters. Having these in place have ensured that more class time is focused on skill development. Along those same lines, in the selection of the portfolio assignments referenced previously, emphasis was placed on finding ways to help students get up-and-running quickly to keep them motivated and excited about analytics. If the students are motivated, the likelihood of achieving the learning outcomes increase.

REFERENCES

Brackett, B., Ring, G. L., Waugaman, C. (2017). The Value of Career ePortfolios on Job Applicant Performance: Using Data to Determine Effectiveness. International Journal of ePortfolio v7 n2 p225-236.

Brown, B., Bughin, J., Byers, A. H., Chui, M., Dobbs, R., Manyika, J., & Roxburgh, C. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Digital Institute. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation. Last accessed on June 15, 2020.

Carbonnelle, P. (2020). PYPL PopularitY of Programming Language. Retrieved from http://pypl.github.io/PYPL.html. Last accessed on June 15, 2020.

Glassdoor (2015). 50 HR & Recruiting Stats That Make You Think. Retrieved from https://www.glassdoor.com/employers/blog/50-hr-recruiting-stats-make-think. Last accessed on June 22, 2020.

Lievens, R. (2014). A proposal: Mitigating effects of the economic crisis with career ePortfolios. International Journal of ePortfolio, 4(2), 157-168.

Mariscal, G., Marbán, Ó, & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137-166.

Ruiz, Mark Angelo. (2020). Hotel Revenue Management. Tableau Public. Retrieved from https://public.tableau.com/profile/mark.ruiz#!/vizhome/HotelRevenueManagement/Overview. Last accessed on June 15, 2020.

DECISION SCIENCES INSTITUTE
Stock Market Portfolio Selection Using Linear Programming

Hesham K. Alfares
King Fahd University of Petroleum & Minerals
Email: alfares@kfupm.edu.sa

Mohammed A. Al-Marhoun
King Fahd University of Petroleum & Minerals
Email: mohammed.almarhun@gmail.com

ABSTRACT

This paper presents an integer linear programming model to optimize portfolio investment in the stock market for maximizing expected return. The model considers risk tolerance, expected dividends, and limits on both the capital investment and the number of stocks for each company. As a new feature, the model ensures that the companies whose shares are selected in the portfolio produce a sufficient variety of different products. A case study is presented, illustrating the model's application to optimize investment in shares of petrochemical companies in Saudi Arabia. Sensitivity analysis is performed to characterize the model's performance, and to make appropriate conclusions.

KEYWORDS: Portfolio selection, Linear programming, Investment Analysis, Risk Management

INTRODUCTION

Strategic investment decisions are difficult, especially in the stock market shares. This is due to the high volatility and uncertainty of the stock market, and its sensitivity to various unpredictable economic, political, and technological factors. The degree of difficulty of such decisions intensifies with larger capital investments and longer-term investment horizons. Several uncertain objectives must be balanced, including expected return and expected risk. In order to deal with risk, it is necessary to diversify the investment. The usual diversification strategy is to invest in several companies by purchasing a portfolio (group) of several company stocks. This paper adds a new dimension of diversification, by requiring the investment in several products. In other words, the companies included in the portfolio must produce a sufficient variety of different products. The proposed diversification approach thus avoids both dependence on a single company and dependence on a single product.

The portfolio selection is an important investment problem for both individuals and organizations such as banks and brokerage firms. Given a limited amount of capital, the main objective is to divide capital among different investment options to maximize the expected return resulting from the resale of purchased assets. Other investment objectives include high dividends, low risk, and fast growth. Different investors may emphasize different objectives according to their financial situations, risk-taking preferences, and investment time horizons. Since high return is associated with high risk, the fundamental question in the portfolio selection problem is to find the optimum balance between risk and return.

The classical portfolio problem either maximizes return subject to a given risk limit or minimizes risk subject to a given return limit. A popular approach to reduce investment risk is to diversify by buying the stocks of several companies. Sometimes, however, this approach may not work if all of the companies in the portfolio produce one product or a few similar products. This is applicable if the investor limits the portfolio selection to a specific market sector, such as construction, transportation, or IT. In that case, a downturn in the demand or prices of the limited set of products will negatively affect the stock prices of all companies in the portfolio. To overcome this limitation, this paper proposes two-fold diversification by investing in multiple products as well as multiple companies.

This paper presents a novel mixed-integer linear programming of the portfolio selection problem with two diversification dimensions: number of companies, and number of products. As far as the authors know, this is a new approach that has not been addressed previously in published literature. The model integrates several practical factors, including risk allowance, dividends income, and bounds on both the capital investment and the number of stocks for any individual company. To demonstrate the model's applicability and usefulness, it is applied to a real case of investment in the petrochemical sector companies in the Saudi Arabian stock market.
The remainder of this paper is organized as follows. Relevant recent literature is reviewed in Section 2. The integer programming model is formulated in Section 3. The case study application is presented in Section 4. Finally, conclusions and recommendations for future research are provided in Section 5.

LITERATURE REVIEW

Portfolio selection problem is a classical and well-studied problem in economics, finance, and operations research problem. Markowitz (1952) proposed the original mean-variance portfolio model, in which he mathematically proved how diversification can reduce the investment risk. Defining risk as the standard deviation of returns, he expressed the variance of the whole portfolio's return as a function of the individual variances of the returns of selected stocks. Using Markowitz portfolio selection model, investors can either maximize return for a given risk, or minimize risk for given return. By varying the levels of given risks and returns, the Pareto efficient front of investment alternatives can be identified. Depending on their personal levels of risk-taking or risk-aversion, each investor can a select an appropriate point on the efficient front corresponding to a specific investment portfolio (Alvani et al., 2004).

Over the years, there have been numerous extensions, variations, and applications of the Markowitz original model. In different occasions, these works have been reviewed and classified in several comprehensive surveys of literature, such those by Markowitz (1999), Rubinstein (2002), and Zhang et al. (2018). Zhang et al. (2018) classify and review several types of improvements of Markowitz's mean–variance model. These improved models include dynamic, robust, and fuzzy portfolio optimization, in addition to portfolio optimization with practical factors. The focus of this section is recent linear programming and diversification approaches to the portfolio selection problem, especially in the stock market investment decisions.

In the literature, portfolio diversification to reduce investment risk is usually considered only across firms and across geographical regions (Luigi and Jappelli, 2008). Diversification across firms means buying assets of several companies, while geographical (international) diversification means buying assets in several countries. Hui (2005) uses the statistical technique of factor analysis to select international markets to invest in. For investors interested in international diversification, he analyzes co-movement and interdependence between the US

and the Asia-Pacific stock markets. Topaloglou et al. (2008) develop a stochastic dynamic programming model to optimize an international portfolio selection problem. Considering multiple periods with uncertain asset prices and currency exchange rates, the model divides available capital among different markets, and selects the assets to buy in each market.

In addition to the number of companies included in the portfolio, entropy has been used as an alternative measure of diversification. Bera and Park (2008) maximize cross-entropy as the objective function, subject to constraints derived from the mean–variance matrix. Huang (2012) develops an entropy-based method to maximize diversification in a fuzzy portfolio selection problem, using triangular fuzzy numbers to replace variances by semi-variances. Yu et al. (2014) use multiple criteria to compare the dynamic performance of several entropy-based models in the portfolio selection problem. The comparison shows that models with entropy have better diversity and asset allocation than models without entropy.

Linear programming (LP) and other optimization approaches have been extensively used to solve various versions of the portfolio selection problem. Mansini et al. (2014) survey and classify LP-based portfolio optimization techniques since 1994. Many of these are mixed integer linear programming (MILP) techniques, used after linearizing Markowitz quadratic programming model of the portfolio selection problem. Sawik (2013) reviews linear and mixed integer programming techniques for portfolio selection optimization with multiple objectives. Multi-criteria portfolio optimization techniques are classified into weighting methods, lexicographic methods, and reference point methods. Li et al. (2000) formulate a portfolio selection with transaction costs as a parametric non-quadratic programming model. Using a linear approximation of the utility function, the model is transformed to a quadratic programming model and then solved by linear programming.

Papahristodoulou and Dotzauer (2004) use LP to compare three formulations of the portfolio selection problem: (i) classical quadratic programming model (ii) maximin model, and (iii) minimum mean absolute deviation model. Approximating the variance by the conditional value at risk, Mansini et al. (2007) formulate an LP model of the portfolio optimization problem and analyze its risk aversion performance. Sadati and Nematian (2013) use an LP model to represent a fuzzy portfolio selection problem aiming to maximize the degree of both the possibility and the necessity. The model is optimally solved by using a two-level LP procedure that calculate two bounds on the objective function value. Xidonas et al. (2018) integrate three objectives and several real-world constraints in a portfolio LP model, which they apply to the European stock market.

Based on the above literature review, the model presented in this paper has a unique feature, which is two-dimensional diversification: both across firms and across products. The problem is defined and the model is formulated in the next section.

PROBLEM DEFINITION AND FORMULATION

The problem considered here is to optimize portfolio selection for a single investor with a limited budget and many investment options to choose from. These options vary widely in terms of expected returns (resale profits), expected risks, and expected dividends. The objective is to maximize total expected return subject to risk and budget limits, in addition to several diversification conditions. The problem is formulated as a linear programming model, which is presented below.

Indices

i Company index (i = 1, 2,…, *I*), where *I* is the total number of companies.
j Product index (j = 1, 2,….., *J*), where *J* is the total number of products.

Input Parameters

R_i Return of buying one share in company i.
C_i Cost of buying one share in company i, excluding commission.
D_i Dividend of one share in company i.
S_i Risk of buying one share in company i.
G_j Set of indices *i* of companies that make product j.
b Trading Commission rate in Saudi Stock Market.
B Total Budget.
L Minimum number of products in the portfolio.
d Investor minimum dividend income, as a percentage of the investment.
s Investor maximum expected risk per share.
p Maximum cost proportion of each company out of the total cost of the portfolio.
M A large positive number.
N Minimum number of shares to buy from each company selected in the portfolio.

Decision Variables

TC Total cost of the investment, excluding purchase commission.
TR Total return of the investment.
X_i Number of company i shares to buy
Y_i Equal to 1 if shares of company *i* are bought; equal to 0 otherwise.
Z_j Equal to 1 if shares of companies producing product *j* are bought, equal to 0 otherwise.

Objective Function

The objective function (1) of the portfolio selection model is to maximize *TR*, the total return of the whole investment.

$$Maximize\ TR = \sum_{i=1}^{i} R_i X_i \quad (1)$$

Constraints

Objective function (1) is optimized subject to constraints (2)-(13) listed below. Constraint (2) defines the total cost *TC* as the sum of purchase costs of all company shares. Constraint (3) ensures the sum of purchase and commission costs does not exceed the available budget *B*. Constraint (4) guarantees the total income from dividends is at least equal to a given proportion *d* of the purchase cost. Constraint (5) assures the average risk index per share does not exceed a given threshold *s*. Constraints (6) make sure the investment in each individual company not exceed a given proportion *p* of the total investment.

The two dimensions of portfolio diversification are imposed by constraints (7)-(11). Constraints (7) relate variables X_i and Y_i, by making sure that Y_i is equal to 1 if $X_i > 0$, i.e. if shares are bought of company *i*. Constraints (8) impose a lower bound, *N*, on the number of shares bought from each selected company. Constraints (9) and (10) are logical constraints relating variables Y_i and Z_j. The set of constraints (9) and (10) ensures that Z_j is equal to 1 if shares are bought of any company producing product *j*, and that Z_j is equal to 0 if no such shares are bought. Constraint (11) impose a lower bound, *L*, on the number of the number of products produced by the companies selected in the portfolio. Finally, constraint (12) specifies non-negativity and binary restrictions on relevant decision variables.

$$TC = \sum_{i=1}^{i} C_i X_i \tag{2}$$

$$TC\,(1+b) \le B \tag{3}$$

$$\sum_{i=1}^{i} D_i X_i \ge d\,TC \tag{4}$$

$$\sum_{i=1}^{i} S_i X_i \le s \sum_{i=1}^{i} X_i \tag{5}$$

$$C_i X_i \le p\,TC \qquad for\ (i = 1,2,\dots\dots,I) \tag{6}$$

$$X_i \le M\,Y_i \qquad for\ (i = 1,2,\dots\dots,I) \tag{7}$$

$$X_i \ge N\,Y_i \qquad for\ (i = 1,2,\dots\dots,I) \tag{8}$$

$$\sum_{i \in G_j}^{j} Y_i \ge Z_j, \qquad for\ (j = 1,\dots,J) \tag{9}$$

$$\sum_{i \in G_j}^{j} Y_i \le MZ_j, \qquad for\ (j = 1,\dots,J) \tag{10}$$

$$\sum_{j=1}^{J} Z_j \ge L \tag{11}$$

$$TC \ge 0,\ \ X_i \ge 0,\ \ Y_i = (0 \text{ or } 1),\ \ Z_j = (0 \text{ or } 1)$$

$$for\ (i = 1,2,\dots\dots,I),\qquad (j = 1,2,\dots\dots,J) \tag{12}$$

CASE STUDY

The above integer programming model has been applied for portfolio selection of investment in the stock market in Saudi Arabia. In particular, the case study application is limited to 10 publicly-owned petrochemical companies whose shares are traded in the Saudi Arabian stock market. For investors interested in a single market sector, investing in several companies may not provide sufficient diversification. This is because the selected companies in the same economic sector may produce a limited number of similar products that face the same set of market forces. Therefore, it is important to have a second dimension of diversification, namely diversification in the number of products made by the selected companies. Using real historical data, the inputs for the model are prepared and presented below.

First, relevant publicly available data from Bloomberg and the Saudi Exchange (Tawaul) Websites was collected for the last three years (2016-2018) for the 10 companies under study. Unfortunately, this is the only period during which all the data required for the model can be extracted. From this data, the expected returns R_i and expected dividends D_i were calculated as the mean returns and dividends during these three years. For the unit stock prices C_i, we simply used the latest values available at the end of 2018. To determine the expected risk for each company, S_i, input positive differences (expected return - actual return) for each year were averaged for the last three years, and then the average market risk for the petrochemical sector was added. The annual values of the expected returns, as well as the average market risk value, were provided by expert financial analysts. All of these input data values are shown in Table 1.

Another important component of input data is the set of products produced by the 10 companies included in the case study. Altogether, these 10 companies produce a total of 22 petrochemical products that are listed in Table 2. The company-product matrix in shown in Table 3. The rows in Table 3 show the set of companies that make each product, while columns show the set of products made by each company. It should be noted that the number “1” in a cell in Table 3 indicates the given (row) company makes the corresponding (column) product, while a “0” in a cell indicates the company does not. Therefore, for each column, the set of indices *i* of rows with “1” represents the set G_j of all companies that make a given product *j*.

The final input parameter relevant to the Saudi stock market is b, the trading commission rate, which is 0.16% of the stock purchase cost. The five remaining input parameters are flexible, because they represent the values to be specified by each individual investor. These are the values of parameters *B*, *L*, *d*, *s*, and *p*. For the case study, the specific values of these parameters are given in Table 4. These values are typical for a class of high-end investors in the Saudi stock market.

Solving the linear programming model of the portfolio selection problem with the values shown in Table 4 leads to the optimum solution shown in Table 5.

For the optimum number of shares shown in Table 5, the corresponding objective function, i.e. the expected return *TR*, is equal to 142,669.87. The optimum solution calls for selecting four out of the ten petrochemical companies to invest in. Although shares are bought in only four companies, the total number of products made by these companies is 14. For this optimum investment, the income from dividends is equal to 1.13% of the total cost *TC*, which is greater than the minimum proportion, d = 1%. Moreover, the average risk index per share is equal to 7.51%, which is less than given threshold, s = 10%.

Table 1. Petrochemical sector companies and their input data values

i	Company Name	Abbreviation	C_i	D_i	R_i	S_i
1	National Industrialization Co	TASNEE	18.5	0.00	0.29	0.085
2	Saudi Arabian Fertilizer Co	SAFCO	73.5	2.25	0.15	0.051
3	Yanbu National Petrochemicals Co	YANSAB	62.5	3.33	0.15	0.101
4	Advanced Petrochemical Co	APC	52.75	2.56	0.14	0.070
5	Saudi Basic Industries Corp	SABIC	98.5	4.20	0.05	0.060
6	Saudi Arabian Mining Co	MAADEN	48.5	0.00	0.04	0.067
7	Saudi International Petrochemical Co	SIPCHEM	19.5	0.55	0.29	0.040
8	National Petrochemical Co	PETROCHEM	26	0.50	0.25	0.085
9	Nama Chemicals Co	NAMA	20.25	0.58	0.20	0.106
10	Saudi Kayan Petrochemical Co	KAYAN	14.5	0.00	0.20	0.086

Table 2. list of products produced by the petrochemical companies.

j	Product	j	Product
1	Polyethylene	12	Polycarbonate
2	Polypropylene	13	Bisphenol A
3	Polystyrene	14	Diammonium Phosphate (DAP)
4	Mono Ethylene Glycol (MEG)	15	Epoxy Resin
5	Methyl Tert-Butyl Ether (MTBE)	16	Hydrochloric Acid
6	Benzene	27	Liquid Caustic Soda
7	Urea	28	Soda Granule
8	Ammonia	29	Methanol
9	Polyvinyl Chloride (PVC)	20	Butanol
10	Purified Terephthalic Acid (PTA)	21	Acetic Acid
11	Propylene	22	Vinyl Acetate Monomer (VAM)

Table 3. Matrix showing the relationship between each company (i) and each product (j)

$i \setminus j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
8	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
10	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0

Table 4. Given input values for an individual investor

s	d	p	L	B
10%	1%	30%	12	10 Million

Table 5. Optimum number of shares to buy for each company

i	1	2	3	4	5	6	7	8	9	10
X_i	161,903	0	0	0	0	0	153,600	0	49,305	206,565

CONCLUSIONS

This paper considered the portfolio selection problem with two dimensions of diversification: diversification across companies, and diversification across products. A linear programming (LP) model has been formulated to maximize expected return. In addition to enforcing two dimensions of diversification, several practical constraints are incorporated in the LP model, including budget, dividends incomes, risk index, and limits on individual company's number of shares and proportion of the total investment. The model has been successfully applied to a case study involving investment in petrochemical companies listed in the Saudi Stock Market. For future research, the LP model can be extended to consider multiple time periods, multiple objectives, or multiple international markets. Possibly, new dimensions of portfolio diversification may also be proposed.

REFERENCES

Alvani, S. M., Azar, A., & Danayeefar, H. (2004). *Quantitative Research Methodology in Management: A Comprehensive Approach*. Saffar, Tehran, Iran.

Bera, A. K., & Park, S. Y. (2008). Optimal portfolio diversification using the maximum entropy principle. *Econometric Reviews*, 27(4-6), 484-512.

Bloomberg professional services. https://www.bloomberg.com/professional/solution/content-and-data/.

Huang, X. (2012). An entropy method for diversified fuzzy portfolio selection. *International Journal of Fuzzy Systems*, 14(1), 160-165.

Hui, T. K. (2005). Portfolio diversification: a factor analysis approach. *Applied Financial Economics*, 15(12), 821-834.

Li, Z. F., Wang, S. Y., & Deng, X. T. (2000). A linear programming algorithm for optimal portfolio selection with transaction costs. *International Journal of Systems Science*, 31(1), 107-117.

Luigi, G., & Jappelli, T. (2008). *Financial Literacy and Portfolio Diversification*. ECO 2008/31, European University Institute (EUI), Fiesole, Italy.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2007). Conditional value at risk and related linear programming models for portfolio optimization. *Annals of Operations Research*, 152(1), 227-256.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2014). Twenty years of linear programming based portfolio optimization. *European Journal of Operational Research*, 234(2), 518-535.

Markowitz, H. (1952). Portfolio selection*. The Journal of Finance*, 7(1), 77-91.

Markowitz, H. M. (1999). The early history of portfolio theory: 1600–1960. *Financial Analysts Journal*, 55(4), 5-16.
Papahristodoulou, C., & Dotzauer, E. (2004). Optimal portfolios using linear programming models. *Journal of the Operational Research Society*, 55(11), 1169-1177.

Rubinstein, M. (2002). Markowitz's "Portfolio Selection": A Fifty-Year Retrospective. *The Journal of Finance*, 57(3), 1041-1045.

Sadati, M. E. H., & Nematian, J. (2013). Two-level linear programming for fuzzy random portfolio optimization through possibility and necessity-based model. *Procedia Economics and Finance*, 5, 657-666.

Sawik, B. (2013). Survey of multi-objective portfolio optimization by linear and mixed integer programming. In *Applications of Management Science* (pp. 55-79). Emerald Group Publishing Limited.

Tadawul website. https://www.tadawul.com.sa/wps/portal/tadawul/home/.

Topaloglou, N., Vladimirou, H., & Zenios, S. A. (2008). A dynamic stochastic programming model for international portfolio management*. European Journal of Operational Research*, 185(3), 1501-1524.

Xidonas, P., Hassapis, C., Mavrotas, G., Staikouras, C., & Zopounidis, C. (2018). Multiobjective portfolio optimization: bridging mathematical theory with asset management practice. *Annals of Operations Research*, 267(1-2), 585-606.

Yu, J. R., Lee, W. Y., & Chiou, W. J. P. (2014). Diversified portfolios with different entropy measures. *Applied Mathematics and Computation*, 241, 47-63.

Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with Markowitz's mean–variance framework: a review of literature. *Fuzzy Optimization and Decision Making*, 17(2), 125-158.

DECISION SCIENCES INSTITUTE
Digitalizing Supply Chains' Impact on Waste in Lean Manufacturing

Abubaker Haddud
Department of Business Innovation and Strategy
Southwest Minnesota State University
Email: abu.haddud@smsu.edu

Anshuman Khare
Faculty of Business
Athabasca University
Email: anshuman@athabascau.ca

ABSTRACT

This paper examines potential impacts of digitalizing supply chains on the seven common types of manufacturing wastes by identifying key benefits to help businesses minimize them. Data were collected from selected experts using an online survey. Most of the examined potential impacts were found significant which may directly reduce the seven types of manufacturing wastes and ultimately improve the overall supply chain and business performance. The results can be beneficial to academic scholars interested in lean operations and digital supply chain management, business professionals specializing in relevant areas, and other parties interested in understanding potential benefits from digitalizing supply chain.

KEYWORDS: Digital supply chains, Lean operations, Manufacturing wastes, Digital transformation, Supply chain management

INTRODUCTION

Supply chain management (SCM) is defined as "the management of the relationships and flows between the 'string' of operations and processes that produce value in the form of products and services to the ultimate consumer. It is a holistic approach to managing across the boundaries of companies and of processes" (Slack et al., 2016, p.399). Supply chains mainly include; raw materials, suppliers, production and manufacturing, distribution, retails, and customers. SCM involves several business areas including operations, production, logistics, procurement, information technology (Meier, 2016). Over the years, businesses invested heavily in adopting disruptive technologies in these areas and businesses continue to capture more benefits as a result. More adoption of smart technologies within supply chains has shifted value creation from the physical artefact to the data created by smart products. Also, there has been a radical change in business models and competitive structures (Pflaum et al., 2018). Such continuous investment and adoption of more technologies led to the emergence of the 'digital supply chain' term. Digitalizing supply chains is one of the fastest growing, and disruptive, trends in the industry (Büyüközkan and Göçer, 2018; Nabhani et al., 2018) and it is defined as "an intelligent process that leverages innovative solutions with emerging technological means to generate new forms of revenue and business value for organizations". In this study, 'digitalizing supply chains' is referred to as the adoption of one or more of the seven key technology trends within supply chains that Meier (2016) identified including; Mobile Technologies, Big Data, Cloud Computing, Internal Social Media, Predictive Analytics, Internet of Things, and 3D Printing. All these

technologies are enabling technologies of what is known as Industry 4.0. The participants were presented with these seven key technologies and they were asked to share their perception about the impact of the use of one or more of these technologies on the seven types of manufacturing wastes.

Although organizations within different industries continue to 'digitalize' their supply chains, it is uncertain how it will unfold (Handfield, 2016). Also, more digitalization of supply chains increases organizations' complexities and it is important to understand the potential benefits such digitalization will bring to businesses (Brinch et al., 2018; Pflaum et al., 2018). Several studies explored aspects of lean production and Industry 4.0 (Mrugalska and Wyrwicka, 2017; Wagner et al., 2017). Mrugalska and Wyrwicka (2017) concluded that there is a possibility of linking lean production and Industry 4.0 approaches. In addition, Kolberg and Zühlke (2015) indicated that there is a lack of comprehensive framework that combines Industry 4.0 solutions with methods of lean production. Furthermore, Wagner et al. (2017) concluded that Industry 4.0 applications can stabilize and support lean principles. In this paper, the aim is to explore potential impacts of digitalizing supply chains resulting from adopting one or more of the above mentioned seven technology trends on the seven types of manufacturing wastes within the context of lean operations. Previous studies suggested further research to better understand supply chain digitalization and its potential (e.g. Bär et al., 2018; Brinch et al., 2018; Iddris, 2018; Pflaum et al., 2018). In general, there is a very limited scientific literature about digital supply chain of the future (Pflaum et al., 2018) and we endeavor to close this gap. More specifically, the study attempts to answer the following main research question:

RQ1: How does digitalizing supply chains impact the seven common types of waste in lean manufacturing?

This paper is organized as follows: Section 2 provides a literature review about digital supply chain and the seven common types of manufacturing wastes. One null and seven alternative hypotheses are developed in this section as well. This section also provides the research framework diagram. Section 3 includes information about the used research methodology and method used for data collection. This includes: Instrument development; Sample and data collection; Profile of participants; Data analysis methods; and reliability and validity of research instrument. Section 4 includes analyses of the collected data, results and findings, and discussions of these findings linking them to relevant literature. Finally, Section 5 discusses research conclusions, theoretical contribution and managerial implications, and research limitations and future research recommendations.

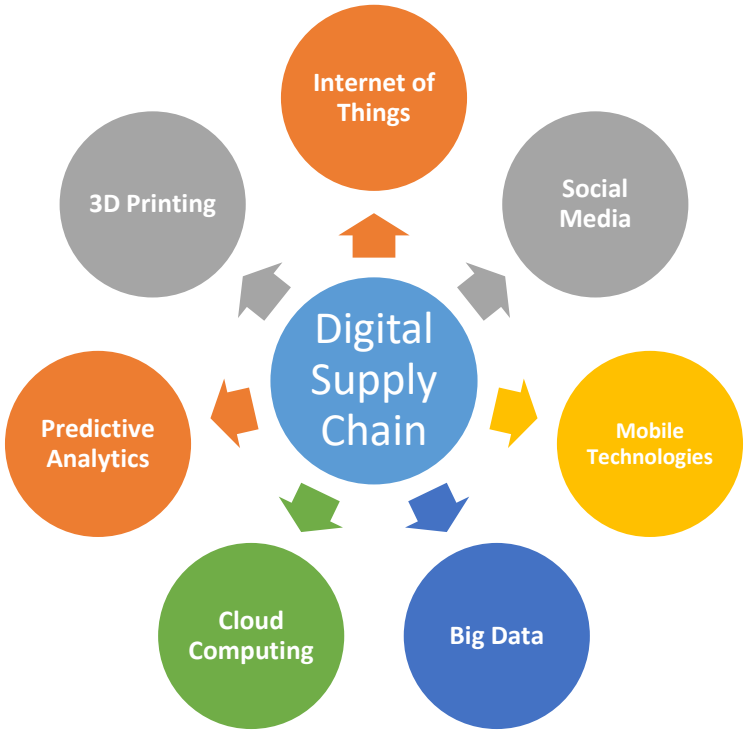
LITERATURE REVIEW

Overview of Digital Supply Chain

Digital supply chain refers to the exchange of information between supply chain stakeholders to enhance communication (Eljazzar et al., 2018) and to integrate various manufacturing processes to create a transparent system (Chaudhary et al., 2018). Digitalizing supply chains is one of the hottest topics in the industry (Büyüközkan and Göçer, 2017) and it directly impacts how retailers, manufacturers, distributors and logistics providers manage their operations particularly inventory (Yearling, 2018). Meier (2016) identified seven digital technology trends, which have been supported with other studies, in supply chain management (see Figure 1) that include: Mobile Technologies (Ivanov et al., 2019), Big Data (Brinch et al., 2019), Cloud

Computing (Suguna and Kumar, 2019), Internal Social Media (Haddud et al., 2015), Predictive Analytics (Zsidisin and Henke, 2019), Internet of Things (Haddud et al., 2017), and 3D printing (Farahani et al., 2017). Digitalizing supply chains provides cost-effective solutions and creates value in the value chain to many actors in the ecosystem including firms and their suppliers, employees and customers (Korpela et al., 2017). In this study, how digitalizing supply chains, using one or more of the above-mentioned seven technologies, impacts the seven types of manufacturing wastes is explored.

Figure 1 Key technologies powering digital supply chains



Overview of The Seven Common Types of Manufacturing Wastes

Lean manufacturing is a Japanese concept that seeks to identify value-adding activities within manufacturing facilities and to eliminate any non-value adding activities that contribute to generating any type of the manufacturing wastes known as; Overproduction, Unnecessary waiting, Unnecessary transportation, Inappropriate processing, Unnecessary inventory, Unnecessary motion, and Defects. The concept of lean thinking was developed based on the Toyota Production System in 1950's. Since then, this concept began to make its way into different business operations. The emerging of technological applications and software, particularly during the 1970's, made it possible to assist organizations with enhancing lean. One of the areas that has seen a huge adoption of technology trends is the supply chain management. In this study, how digitalizing supply chains impacts the seven types of manufacturing wastes is explored. Therefore, the following null hypothesis was developed:
H0: Digitalizing supply chains does not directly reduce the levels of the seven common types of manufacturing wastes.

Overproduction

Overproduction is one of the seven common lean manufacturing wastes. Overproduction occurs when goods are made in excess volumes, made too early, or made 'just-in-case' there would be a demand for them (Wahab et al., 2013). Overproduction has been considered as a deadly waste because it creates other manufacturing wastes (Ohno, 1988) such as unnecessary waiting, redundant transportation, unnecessary inventory, and excess motions. Overproduction occurs when producing items for which there are no orders or producing more than required at a certain time. This can be intentional or unintentional due to inaccurate demand forecasting (Hart et al., 2016). Therefore, companies have sought different ways to ensure that only required production levels are made to avoid overproduction issues. One of predominant used ways were the adoption of sophisticated technologies that can accurately predict demand and develop robust manufacturing plans resulting in efficient manufacturing practices. In this study, the potential impact of digitalizing supply chains overproduction waste is explored. Therefore, the following alternative hypothesis was developed:

H1: Digitalizing supply chains positively impacts lean operations by reducing the level of overproduction waste

Unnecessary waiting

The waiting waste is the time when the production flow stops or is disrupted (Wahab et al., 2013) due to causes related to; manpower, lack of available materials, faulty machines, and any disruptions in processes and operations. When goods are not moving, this affects the workers and goods and the time is not used effectively; thus, stoppage is considered as a waste (Womack and Jones, 2010). This waste is one of the biggest challenges supply chain members face leading to long waiting times or delivery times and can cause serious delays (Chen et al., 2013). Different factors can contribute to the waste of time such as poor information management regarding product or service demand, forecasts, capacity, and production/service plans (Aitken et al., 2016). Other causes of waiting waste are due to uncertainties including; uncertain demand (arrival of new orders and order cancellation), changes in order priority, processing delays, changes in release dates, machine breakdowns, unavailability of raw material/machine/personnel or organizational constraints/priorities etc. (Kang et al., 2014). To design workplaces that run smoothly without any disruptions, companies require a great level of process and operation integration within internal operations and across external supply chain parts. In this study, potential impacts digitalizing supply chains may have on reducing unnecessary waiting waste are explored. Therefore, the following alternative hypothesis was developed:

H2: Digitalizing supply chains positively impacts lean operations by reducing the levels of unnecessary waiting waste.

Unnecessary transportation

Waste of transportation is the inefficient transportation of materials (Harris et al., 2016) and all the associated needless movements (Okpala, 2014) and this can involve raw components, sub-assemblies, empty boxes or just about anything that is required for production (Kumar et al., 2015). Major causes of excessive transportation may include maintaining different warehouses, difficult material handling processes, over-production, improper layouts (Okpala, 2014), shipping

wrong materials, delivering materials to the wrong facility/location, wrong materials pick-ups, and human errors in material handling. Lean thinking seeks to achieve a smooth and continuous flow of inventory through the entire operations and processes (Cawley et al., 2013). Effective supply chain integration, that allows different parties to share accurate information, continues to improve the overall supply chain performance. In this study, potential impacts of supply chain digitalization on unnecessary transportation waste are explored. Therefore, the following hypothesis was developed:

H3: Digitalizing supply chains positively impacts lean operations by reducing the levels of Unnecessary transportation waste.

Inappropriate processing

Inappropriate processing, or over-processing, is referred to as taking unnecessary steps that are not required in a production process. Such unnecessary processing is considered as non-value-added processing and it can be due to the rework, testing, sampling, inspection etc. (Kang et al., 2014). Over-processing can also mean manufacturing products with a quality more than the actual requirements (Kulkarni et al., 2014). Causes of over-processing wastes may include product change without process change, lack of communication, over-processing to accommodate downtime, and redundant approvals (Agrahari et al, 2015). Therefore, it is very important that companies find better ways to ensure only the necessary processing practices are done. In this study, potential impacts of digitalizing supply chain on waste of inappropriate processing are explored. Therefore, the following hypothesis was developed:

H4: Digitalizing supply chains positively impacts lean operations by reducing the level of Inappropriate processing waste.

Unnecessary inventory

Unnecessary inventory refers to excess of stock from raw materials to finished goods (Kumar, 2014). Excessive inventory leads to; insufficient stock conditions, possible material thefts, eminent material quality deterioration, occupies space, adversely affects capital, and incurs costs (Ansah et al., 2016). In lean manufacturing, inventory should only be pulled through the different manufacturing processes when it is needed (Gîfu et al., 2014) and this applies to the entire supply network as well. Procurement and inventory management have improved dramatically over the years using different sophisticated software and technological applications that improved materials demand forecasts, material handling, and inventory management in general. The trend of using more disruptive technologies continues establishing what is currently known as 'procurement 4.0'. The digitalization of procurement process can bring important benefits within supply chains (Bienhaus and Haddud, 2018). In this study, potential impacts digitalizing supply chain may have on unnecessary inventory waste were examined. Therefore, the following hypothesis was developed:

H5: Digitalizing supply chains positively impacts lean operations by minimizing the waste of Unnecessary inventory.

Unnecessary motion

The unnecessary motion of resources across the production facilities is considered waste (Kang et al., 2014). Excess motion can also mean the unnecessary movement of equipment (equipment redundant, ineffective, and nonvalue added to the production process) or people (operator, foreman, and the people who are directly related to production) which is caused by poor workflow, poor layout, housekeeping, and inconsistent or undocumented work methods (Haghirian, 2016). The use of technology and the continuous adoption of digital transformation in workplaces have minimized/eliminated some of these causes. In this study, potential impacts of digitalizing supply chains the waste of unnecessary motions were examined. Therefore, the following hypothesis was developed:

H6: Digitalizing supply chains positively impacts lean operations by reducing the levels of Unnecessary motion.

Defects

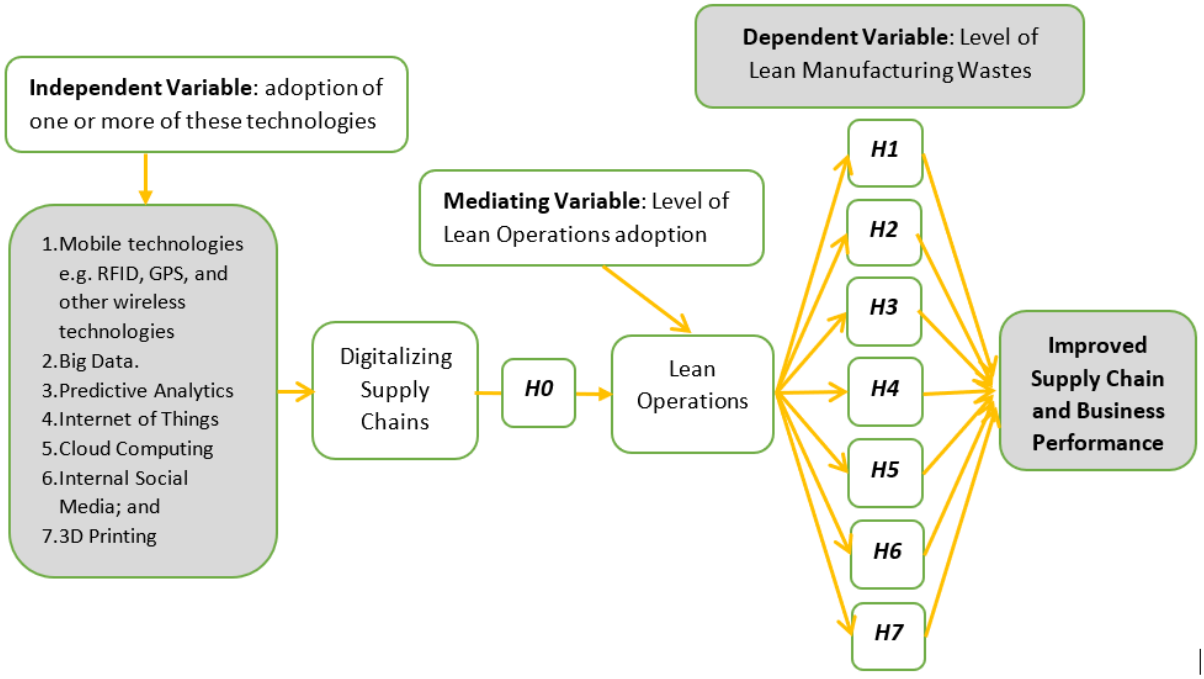
The waste of defects includes any errors in the production process that require either repairing/rework or scraping the produced defective items (Hassan, 2013). Common causes of producing detective items may include lack of 5S tool implementation, poor production line flow and layout, overproduction, insufficient workers training, and inaccurate standards and instructions (Sultana and Islam, 2011). Defective items are considered waste because production lines stop unnecessarily, and this leads to more manufacturing wastes e.g. unnecessary waiting time and accumulated work-in-progress stocks. In this study, potential impacts of digitalizing supply chains on defects waste are explored. Therefore, the following hypothesis was developed;

H7: Digitalizing supply chains positively impacts lean operations by reducing the levels of defects waste.

Conceptual Research Framework

Based on the above literature review on digital supply chains and the seven types of manufacturing wastes; a conceptual model was developed to examine potential impacts of digitalizing supply chains on these seven types of manufacturing wastes (see Figure 2). The dependent variable of this research is the adoption of one or more of the seven identified technologies. The mediating variable is the level of lean operations adoptions, and the dependent variable is the level of seven identified lean manufacturing wastes.

Figure 2 Conceptual framework of the research

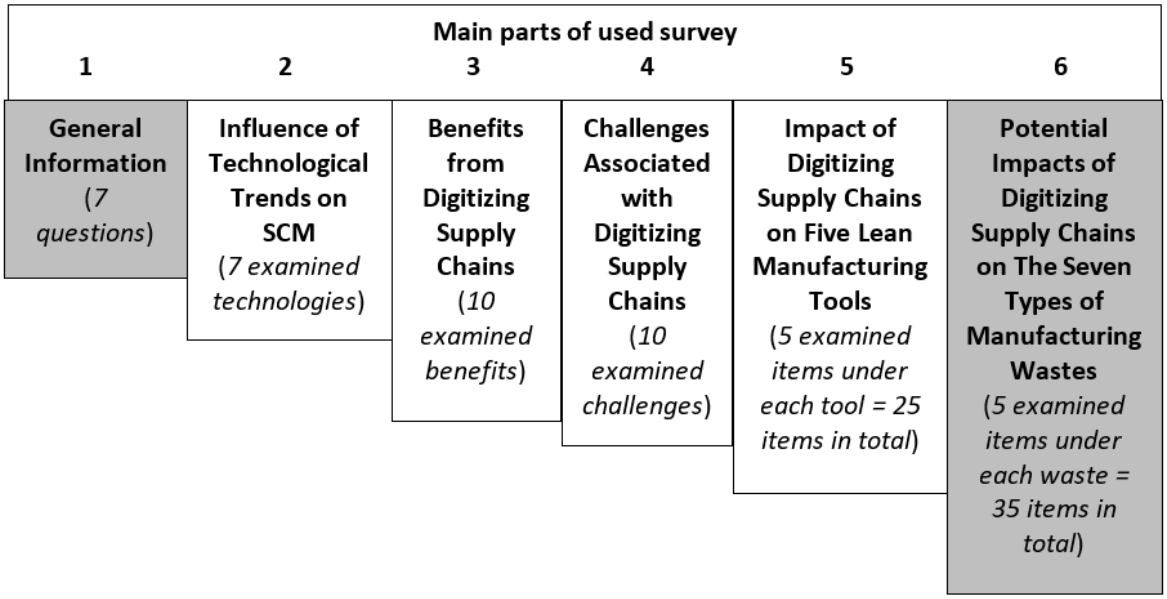


RESEARCH METHODOLOGY

Instrument Development

An online survey was utilized to collect primary data for this research capturing targeted participants' perceptions about the explored potential impacts of digitalizing supply chains on seven manufacturing wastes as related to lean operations. The survey instrument was developed using information extracted from around 20 journal articles appeared in a book chapter discussing the concept of digital supply chain (Haddud and Khare, 2018).
This study was approved by Southwest Minnesota State University's Institution Review Board in September 2018 and utilized an online survey that consisted of six main parts (see Figure 3) and parts 1 and 6 are included in this paper. First, General Information section, consists of seven questions including years of experience in current field; field and area of expertise; work sector; knowledge of supply chain management, knowledge of lean options; and knowledge of digital supply chain. Second, seven constructs (one for each waste) were used that included five potential impacts under each one.

Figure 3 Main Components of Survey (greyed parts included in this paper)



Sample and Data Collection

Data collection process took place in Fall 2018 using a web-based survey developed on Google Forms. Two potential sources of primary data were identified; 1) Academics who specialize in areas related to digitalizing supply chains and lean operations, and 2) Industry professionals who have direct involvement with digitalizing supply chain aspects and implementation. Although it would be more useful to gain insights from the second group, it was not possible to get access to a good sample size from this group. A total number of 923 email invitations were sent to participants who published, or co-published, article(s) in the Decision Science Institute (DSI) Annual Conference proceedings in 2015 to 2018 and in the Emerald's Journal of Manufacturing Technology Management in 2015 to 2018. With 79 completed surveys, the response rate was 8.55%. Six of the submitted survey responses were incomplete and were excluded from the data analysis process. Therefore, the total number of received usable survey responses was 73.

Profile of Participants

The profile of participants is provided in Table 1. The participants had work experience that spans from 5 to 21+ years. Furthermore, most participants worked in supply chain and operations management, lean manufacturing, and production/logistics. Other fields were decision science, executive management, and information technology, etc. Furthermore, the participants worked for academia, industry, or both. To ensure the included sample, particularly the academics, possess the required knowledge to answer the survey questions, participants were asked to report their level of knowledge of supply chain management, lean manufacturing, and digital supply chains as shown in Table 1.

Table 1 Profile of participants

Question	Frequency (n=73)	Percentage (%)
Years of work experience in current field		
• 1 - 5	13	17.80
• 6 - 10	20	27.39
• 11 - 20	19	26.02
• 21 +	21	28.76
Field and main area of expertise		
• Supply Chain and Operations Management	31	42.46
• Lean Manufacturing	9	12.32
• Manufacturing/production/logistics	19	26.02
• Other	14	19.17
Sector		
• Academia	51	69.86
• Industry	10	13.69
• Both	12	16.43
Level of knowledge – used scale (Low) 1 – 7 (High)		
	Mean values	
• Knowledge of Supply Chain Management	5.66	
• Knowledge of Lean Operations	5.46	
• Knowledge of Digital Supply Chains	4.73	

Data Analysis Methods

The IBM Statistical Package for the Social Sciences (SPSS) version 24 was used to analyze the collected data. First, the reliability and validity of the used survey were examined. Items' consistency and reliability used in the survey were examined. Cronbach alpha values were calculated for each construct as well. Also, an exploratory factor analysis was used to examine item and construct validity. Second, descriptive analyses were conducted to calculate frequency means for each item that used the five-point Likert scale. Finally, Chi-square test (χ^2 test) was used for hypotheses testing.

Reliability and Validity of Research Instrument

Reliability is defined as the proportion of variance in observed test score that is related to true scores (Cronbach, 1951; McDonald, 2013) of the scale items in the same construct (DeVellis, 1991). Cronbach alpha values were used to examine the consistency and reliability of the items used in each of the survey constructs. As shown in Table 2, the Cronbach alpha coefficient values for the 7 used constructs ranged from 0.801 to 0.897. Rivard and Huff (1988) suggest that Cronbach's values exceeding alpha coefficient of 0.7 thresholds provide reliability evidence for internal consistency of the measurement scales. Although an acceptable reliable coefficient is normally 0.7 or higher, lower thresholds are sometimes used in the literature (Santos, 1999). For new instruments, constructs with reliability values as low as 0.5 is also acceptable (O'Leary-Kelly and Vokurka, 1998). Because the closer to 1 the Cronbach alpha coefficient value is, the more reliable the used construct becomes (Cohen et al., 2003), it can be concluded that the developed survey instrument for this research study be considered as an extremely reliable research tool that; 1) provides a thorough picture of the examined themes; 2) used items/statements that are directly related to the main theme of the construct; and 3) all of the used statements are inter-related.

Table 2 Used survey constructs' validity and reliability tests results

Used Scale	Number of items	Items factor loadings ranges	Construct Cronbach Alpha value
• Overproduction	5	0.515 to 0.691	0.826
• Unnecessary waiting	5	0.520 to 0.637	0.807
• Unnecessary transportation	5	0.610 to 0.819	0.801
• Inappropriate processing	5	0.539 to 0.814	0.891
• Unnecessary inventory	5	0.559 to 0.716	0.875
• Unnecessary motion	5	0.632 to 0.792	0.897
• Defects	5	0.619 to 0.758	0.883

In addition, an exploratory factor analysis (Principal Component) was used to examine the survey's construct validity. This analysis helps to determine how, and to what extent, each used item within survey constructs is linked to their underlined factors (Hair et al., (2006) state that the rule-of-thumb is that factor analysis values greater than 0.30 should be considered significant, values greater than 0.40 should be considered more important, and values that are 0.50 or greater should be considered very significant. As shown in Table 3, all the used constructs had item factor loading values greater than 0.5. Therefore, the construct validity of the used survey constructs is very significant.

Hypotheses Testing Method

Chi-square test (χ^2 test) was used to test the used research hypotheses in this study. The χ^2 test is widely used to examine data for differences, associations, and relationships to answer hypotheses (Waller, 2012). It compares observed frequencies (N) with the corresponding expected frequencies (N) under each of the used five levels of agreement. The expected N can be determined by dividing the number of total usable responses (73 in this study) by the number of used Likert-scale levels of agreement options (5 in this study). By dividing 73 by 5, the result is 14.6 and this was the used expected N to calculate residual values for the used survey constructs.

ANALYSIS AND RESULTS

Potential Impacts of Digitalizing Supply Chain on Seven Manufacturing Wastes

Overproduction

The used survey construct for overproduction waste consists of five items (see Table 3). The overall mean values ranged from 4.05 to 4.32. Information sharing within the entire supply chain came in the first place with (mean = 4.32). This result goes in line with statements mentioned in other articles stating that digitalizing supply chains helps establishing effective information sharing platforms integrating multiple parties (Zsidisin and Henke, 2019). Providing the required visualizations and trackability came in the second place with (mean = 4.22). Use of advanced smart technologies has provided robust capabilities to improve supply chain visualization and trackability (Merlino and Sproǵe, 2017; Suguna and Kumar, 2019). Therefore, digitalizing supply chains will continue to improve this area. The other three examined potential impacts were related to stock and materials management and had mean values as 4.18, 4.15, and 4.05

consecutively. Therefore, digitalizing supply chains leads to better inventory management. Keeping only the required stock would minimize the possibility of producing more than is needed and this may prevent producers from failing into the 'overproduction' trap. Thus, digitalizing supply chains minimizes/eliminates the manufacturing waste of overproduction.

Table 3 Impact of Digitalizing Supply Chains on overproduction waste.

Examined Impact	Mean (n=73)	Cronbachα	Factor loading	χ² test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Sharing information within the entire supply chain helps determine if there are any bottleneck in the system and to adjust production levels accordingly	4.32	0.787	0.618	-14.6	-14.6	-8.6	23.4	14.4	Yes
2. Providing the required visualization and goods traceability to ensure only the required goods are produced.	4.22	0.807	0.522	-14.6	-14.6	-6.6	26.4	9.4	Yes
3. Knowing how much of goods are in stock and enabling automated JIT strategies	4.18	0.766	0.691	-14.6	-11.6	-11.6	30.4	7.4	Yes
4. Providing more accurate demand forecasting and knowledge about changes in the marketplaces	4.16	0.785	0.632	-14.6	-14.6	-7.6	17.4	4.4	Yes
5. Ordering materials required for the production process, get the exact needed quantities, make them available only when they are needed	4.05	0.812	0.516	-14.6	-12.6	-3.6	26.4	4.4	Yes

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 3, the highest residual values under each of the examined five potential impacts were under the 'Agree' option. Therefore, all impacts were supported and digitalizing supply chains can directly minimize the manufacturing waste of overproduction through; 1) sharing information within the entire supply chain, 2) providing the required visualization, 3) enabling automated JIT strategies; 4) more accurate demand forecasting; and 5) better materials management. Therefore, ***H1****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of Overproduction* was accepted.

Unnecessary waiting

The used survey construct for unnecessary waiting waste consists of five items (see Table 4) and the overall mean values ranged from 3.92 to 4.15. The top-rated potential impact was better inventory management and control (mean = 4.15). Unavailable inventory was perceived as one of the main causes of disruptions in supply chains and this normally leads to interruptions in production activities and disrupted materials movements and supply, workers productivity and performance, and machinery waiting time. Therefore, use of technologies, such as predictive analytics, within supply chains would directly minimize the risk of supply disruptions and this will prevent unnecessary (materials, machines, and workers) waiting. Better information

management (mean = 4.14) came second. Information management allows businesses to better forecast fluctuating demands (Tjahjono et al., 2017) make better capacity allocation and utilization decisions, and optimization of production planning (Farahani et al., 2017).
More visible supply chains that leads to reduced, or eliminated, flow disruptions and wastes times came third with (mean = 4.12). Providing supply chain visibility from adopting certain technologies made businesses' supply chains more 'digital' (Korpela et al., 2017; Li et al., 2018). Digitalizing supply chains prevents flow disruptions through establishing more visualized supply chains. Such disruptions can also be avoided through accurate provision of information about production levels, required materials, and the technical conditions of the machinery. This potential benefit was in the fourth place (mean = 4.07). Finally, early spotting of errors in supply chains allows managers to take proactive actions to avoid any potential disruptions in the production process came in the last position (mean = 3.92). We can infer from these results that the core impact of digitalizing supply chain on minimizing waiting waste is the identification of potential causes of supply chain flow disruptions.

Table 4 Impact of Digitalizing Supply Chains on unnecessary waiting waste.

Examined Impact	Mean (n=73)	Cronbach α	Factor loading	χ^2 test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Better inventory management and control	4.15	0.778	0.528	-14.6	-12.6	-5.6	23.4	20.4	Yes
2. Better information management regarding product or service demand, forecasts, capacity, and production (or service) plans	4.14	0.771	0.590	-14.6	-14.6	-6.6	32.4	3.4	Yes
3. More visible supply chains lead to reduced/eliminated flow disruptions and wasted times	4.12	0.764	0.592	-14.6	-12.6	-2.6	19.4	10.4	Yes
4. Providing accurate information about production levels, required materials, the technical conditions of the machinery	4.07	0.759	0.637	-14.6	-13.6	-11.6	34.4	0.4	Yes
5. Early spotting of errors in supply chains allows managers take proactive actions to avoid any potential disruptions in the production process	3.92	0.780	0.520	-14.6	-12.6	2.4	24.4	0.4	Yes

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 4, the highest residual values under each of the examined five potential impacts were under the 'Agree' option. Therefore, all of these impacts were supported, and digitalizing supply chains can directly minimize the manufacturing waste of unnecessary waiting; 1) better inventory management and control, 2) providing better information management, 3) more visible supply chains; 4) more accurate information; and 5) early spotting of errors in the supply chain. Therefore, ***H2****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of unnecessary waiting* was accepted.

Unnecessary transportation

The used survey construct for unnecessary transportation waste consists of five items (see Table 5). The five examined potential impacts had overall mean values ranged from 3.57 to 4.09. The highest-rated impact was better inventory scheduling (mean = 4.09). The participants felt that better scheduling, generation, and delivery of inventory using JIT approach is the top benefit businesses may gain from digitalizing supply chains that will minimize the unnecessary transportation waste. This result goes in line with similar reported benefits from digitalizing supply chains (Farahani et al., 2018). Better items traceability and network visibility came in the second place (mean = 4.08). Better item and material traceability and supply network visibility have already been highlighted as a key benefit that businesses gain from digitalizing supply chain. This goes in line with similar reported benefits by other authors (Korpela et al., 2017; Li et al., 2018). Accurate delivery of materials to the designated facility or location (mean = 4.05) was in the third place. This potential benefit also relates to better inventory management and thus digitalizing supply chains leads to less of unnecessary transportation due to improved inventory and material management.

Furthermore, fewer human errors in material handling to avoid any misplaced materials came in the fourth place (mean = 3.64). One possible reason for this low rating could be the currently adopted technologies may still lack the capabilities to minimize human errors. However, newer smart technologies may have a different impact on this area. For example, Perboli et al., (2018) state that having the visibility of the entire supply chain and better information management could help to reduce the chances of human errors. Thus, more digitalization of supply chains would ultimately reduce human errors. Therefore, unnecessary transportation waste can be minimized by reducing the amount of misplaced materials. Finally, promoting better teamwork and feedback practices that leads to early identification of errors came in the last place (mean = 3.57). Other studies have indicated that the use of collaborative and internal social platforms can yield business benefits (Haddud et al., 2016) but it is hard to relate yielded benefits to the adoption of such platforms to minimize the unnecessary transportation.

Table 5 Impact of Digitalizing Supply Chains on unnecessary transportation waste.

Examined Impact	Mean (n=73)	Cronbach α	Factor loading	χ^2 test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Better schedule the generation and delivery of inventory to downstream processes using JIT approaches	4.09	0.740	0.816	-14.6	-12.6	-7.6	32.4	3.4	Yes
2. Better items traceability and network visibility	4.08	0.805	0.610	-14.6	-13.6	-2.6	26.4	5.4	Yes
3. Accurate delivery of materials to the designated facility/location	4.05	0.738	0.819	-14.6	-13.6	-2.6	28.4	3.4	Yes
4. Fewer human errors in material handling to avoid any misplaced materials	3.64	0.753	0.782	-13.6	-10.6	15.4	10.4	-0.6	**No**

5.	Promoting better teamwork and feedback practices that leads to early identification of errors	3.57	0.773	0.727	-13.6	-9.6	14.4	14.4	-4.6	**No**

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 5, the highest residual values under the first three examined potential impacts were under the 'Agree' option and were under 'Neutral' for the remaining two impacts. Therefore, the first three impacts were supported only, and it can be inferred that digitalizing supply chains can directly minimize unnecessary transportation through; 1) better scheduling, generation, and delivery of inventory; 2) better items traceability and supply network visibility; and 3) accurate delivery of materials. Therefore, ***H3****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of unnecessary transportation* was partially accepted.

Inappropriate processing

The used survey construct for the waste of inappropriate processing consists of five items (see Table 6) and they had overall mean values ranging from 3.57 to 4.04 with only one item above 4. Better production planning and control came in the first place with (mean = 4.04). Optimization of production planning has already been identified earlier as one of the key potential benefits that help businesses reduce the waste of unnecessary waiting in the manufacturing processes and this was reported by Farahani et al., (2017) as well. More accurate information shared on collaborative systems was in the second place with (mean = 3.97). While providing more accurate information has already been identified as a key benefit that businesses would gain from digitalizing supply chains, the use of collaborative systems part of the item's wording may have led to this lower ranking. The third, fourth, and fifth ranked potential impacts were; less over-production (mean = 3.72), better quality control (mean = 3.66), and fewer defective goods (mean = 3.57). The participants did not perceive digitalizing supply chains to have a great impact on these three areas.

Table 6 Impact of Digitalizing Supply Chains on inappropriate processing waste.

Examined Impact		**Mean (n=73)**	**Cronbach α**	**Factor loading**	**χ^2 test (frequencies residual values)**					**Item is supported**
					SD (1)	**D (2)**	**N (3)**	**A (4)**	**SA (5)**	
1.	Better production planning and control	4.04	0.889	0.572	-13.6	-13.6	-4.6	29.4	3.4	Yes
2.	More accurate information shared on collaborative systems	3.97	0.894	0.539	-13.6	-12.6	-2.6	27.4	2.4	Yes
3.	Less over-processing (e.g. reworks, inspections, over-processing, and waiting time)	3.72	0.847	0.804	-13.6	-12.6	11.4	18.4	-2.6	Yes
4.	Better quality control practices and actions	3.66	0.845	0.814	-13.6	-9.6	8.4	19.4	-3.6	Yes
5.	Fewer defective goods and consequently greater resource efficiency is achieved	3.57	0.856	0.776	-12.6	-7.6	11.4	10.4	-0.6	**No**

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 6, the highest residual values under four of the examined five impacts came under the 'Agree' option. Therefore, these four impacts were supported, and it can be inferred that digitalizing supply chains can directly minimize inappropriate processing thorough; 1) better production planning; 2) more accurate information; 3) less over-processing; and 4) better-quality control. However, the fifth examined impact had the highest residual value under 'Neutral' and this impact was not supported. Therefore, ***H4****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of inappropriate processing* was partially accepted.

Unnecessary inventory

The used survey construct for unnecessary inventory waste consists of five items (see Table 7) and they had overall mean values ranging from 3.95 to 4.29 with four values were above 4.0. Sharing of more accurate information came in the first place with a mean value of 4.29. This potential impact has already been confirmed as important earlier. In the second place, better visualization and tractability had an overall mean value of 4.27. The importance of this impact was also highlighted earlier, and this indicates that digitalizing supply chains results in better supply chain visualization and item trackability and ultimately minimizes the unnecessary inventory waste. Furthermore, more accurate inventory forecasting (mean = 4.18) and better inventory management and control (4.01) came in the third and fourth places. These findings go in line with available literature that clearly indicate that the use of technologies within supply chains improves inventory forecasting and management (Tjahjono et al., 2017; Zsidisin and Henke, 2019). Finally, achieving a smooth and continues flow of inventory thorough the entire operations and processes came in the last place.

Table 7 Impact of Digitalizing Supply Chains on unnecessary inventory waste.

Examined Impact	Mean (n=73)	Cronbach α	Factor loading	χ^2 test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Sharing of more accurate information	4.29	0.869	0.559	-14.6	-14.6	-3.6	17.4	16.4	Yes
2. Better processes visualizations and item tractability	4.27	0.840	0.716	-14.6	-14.6	-4.6	20.4	14.4	Yes
3. More accurate inventory forecasting	4.18	0.843	0.692	-14.6	-12.6	-5.6	24.4	9.4	Yes
4. Better inventory management and control	4.10	0.844	0.695	-14.6	-12.6	-1.6	19.4	9.4	Yes
5. Achieving a smooth and continuous flow of inventory through the entire operations and processes	3.95	0.845	0.688	-14.6	-8.6	-5.6	29.4	0.4	Yes

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 7, the highest residual values for the five examined potential impacts were under the 'Agree' option. Therefore, these five impacts were supported, and it can be inferred that digitalizing supply chains can directly minimize unnecessary inventory through; 1) sharing of accurate information; 2) better visualization and tractability; 3) more accurate inventory forecasting; 4) better inventory management and control; and 5) achieving a smooth and continuous flow of inventory through the supply chain. Therefore, ***H5****: Digitalizing*

supply chains positively impacts lean operations by minimizing the waste of unnecessary inventory was accepted.

Unnecessary motion

The used survey construct for the waste of unnecessary motion consists of five items (see Table 8) and they had overall mean values ranging from 3.41 to 3.99. Better information-sharing came in the first place (mean = 3.99). This potential impact was perceived to have a direct impact on overproduction, unnecessary waiting, unnecessary transportation, and unnecessary inventory waste in this study already. Better trackability and visibility came in the third place (mean = 3.89). This potential impact was ranked higher earlier, but the participants did not see a clear link to its reduction of unnecessary motion waste. Creating smooth and continuous flow of inventory came in the third place (mean = 3.79). The participants did not perceive this potential impact to have a great impact on unnecessary inventory and unnecessary motion. In the fourth place, using collaborative systems to increase teamwork had a mean value of 3.74 and minimizing individuals' excess motion came in the last place (mean = 3.41).

Table 8 Impact of Digitalizing Supply Chains on unnecessary motion waste.

Examined Impact	Mean (n=73)	Cronbach α	Factor loading	χ^2 test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Better information-sharing about consistent and documented work methods	3.99	0.867	0.760	-13.6	-9.6	-4.6	22.4	6.4	Yes
2. Better items' traceability and network visibility ensuring work-in-progress stocks are placed within an easy access to every workstation.	3.89	0.862	0.792	-13.6	-12.6	2.4	22.4	1.4	Yes
3. Smooth and continuous flow of inventory through the entire operations and processes	3.79	0.882	0.678	-13.6	-11.6	0.4	32.4	-6.6	Yes
4. Using collaborative systems to increase teamwork and feedback practices to flag any errors in the flow of materials so that unnecessary movement is avoided.	3.74	0.865	0.752	-12.6	-9.6	3.4	20.4	-0.6	Yes
5. Minimizing individuals' excess motions	3.41	0.895	0.632	-8.6	-6.6	4.4	17.4	-5.6	Yes

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 8, the highest residual values of the examined five potential impacts were under the 'Agree' option. Therefore, these impacts were supported, and it can be inferred that digitalizing supply chains can minimize the unnecessary motion through: 1) better information sharing; 2) better traceability and visibility; 3) smooth and continuous flow of inventory; 4) using collaborative systems to increase teamwork; and 5) minimizing individuals' excess motions. Therefore, ***H6****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of unnecessary motion* was accepted.

Defects

The used survey under this construct consists of five items (see Table 9) and had overall mean values ranging from 3.94 to 4.17. Providing visibility of the entire supply chain came in the first place. This potential impact has already been confirmed in this study and such a benefit from digitalizing supply chains was mentioned in the literature (Korpela et al., 2017; Li et al., 2018). Providing faster feedback and timely identification of any errors in the system came in the second place (mean = 4.04). Faster identification of errors and reaction to abnormalities in the supply chain contributes to improving the first pillar of the Triple Constraint pyramid that includes: speed, quality, and cost. Making the defect-detecting process more accurate and easier to manage came in the third place (mean = 3.96). In the fourth and fifth places, using advanced predictive analytics to identify potential causes of defects (3.94) and Providing more effective technological collaboration environments (3.94). While these two potential impacts were ranked at the bottom of the list; they still have a relatively high mean value and therefore, we consider them important as well.

Table 9 Impact of Digitalizing Supply Chains on defects waste.

Examined Impact	Mean (n=73)	Cronbach α	Factor loading	χ^2 test (frequencies residual values)					Item is supported
				SD (1)	D (2)	N (3)	A (4)	SA (5)	
1. Visibility of the entire supply chain and the ability to track materials moving across the production process to identify any damaged work-in-progress parts or wrong delivered material quality	4.17	0.859	0.679	-13.6	-11.6	-6.6	20.4	12.4	Yes
2. Provide faster feedback and timely identification of any errors in the system	4.04	0.846	0.758	-13.6	-13.6	-2.6	26.4	3.4	Yes
3. Making the defect-detecting process more accurate and easier to manage	3.96	0.848	0.731	-13.6	-11.6	-2.6	26.4	2.4	Yes
4. Using advanced predictive analytics to identify potential causes of defects	3.94	0.867	0.640	-13.6	-9.6	-6.6	28.4	2.4	Yes
5. Providing more effective technological collaboration environments	3.94	0.870	0.619	-13.6	-11.6	-4.6	30.4	-0.6	Yes

Notes: SD= Strongly disagree; D= Disagree; N= Neutral; A= Agree; SA= Strongly agree

In summary, and as shown in Table 9, the highest residual values of the five examined potential impacts were under the 'Agree' option. Therefore, these impacts were supported, and it can be inferred that digitalizing supply chains can directly minimize defects through; 1) visibility of the entire supply chain; 2) provide faster feedback and timely identification of errors; 3) making the defect-detecting process more accurate; 4) better identification of potential causes of defects; and 5) providing more effective technological collaboration environments. Therefore, ***H7****: Digitalizing supply chains positively impacts lean operations by minimizing the waste of defects* was accepted.

CONCLUSIONS

Conclusions

This study aimed to assess the impact of digitalizing supply chains on the seven common types of manufacturing wastes. A total of 35 potential impacts were examined (5 impacts under each construct). As the results showed in section 4, 32 of the 35 examined potential impacts were supported and 5 out of the 7 examined hypotheses were accepted. These accepted hypotheses were related to; Overproduction, Unnecessary waiting, Unnecessary inventory, Unnecessary motion, and Defects. Whereas two hypotheses related to Unnecessary transportation and Inappropriate processing were partially accepted. Table 10 represents how each of the examined lean manufacturing waste is impacted using possible technologies within supply chains. While the impact of implementing different Industry 4.0 applications on different lean manufacturing has been explored (Chen and Lin, 2017; Karre et al., 2017; Wagner et al., 2017; Wang, 2016), businesses already implementing lean manufacturing need guidelines on how to react to the impacts of Industry 4.0 (Meudt et al., 2017). It is important to know how and where to integrate the new technologies from Industry 4.0 into their existing lean manufacturing systems (Wagner et al., 2017), but the knowledge of how and where this should be done is still not fully explored (Kolberg and Zühlke 2015; Wagner et al., 2017). The finding from this research closes this knowledge gap and provides useful insights.

Table 10 Impact of digitalizing supply chains on the seven manufacturing wastes.

Manufacturing waste	Impact of digiting supply chains	Possible adopted technologies
Overproduction	1) better sharing information within the entire supply chain 2) providing the required visualization 3) enabling automated JIT strategies 4) more accurate demand forecasting 5) better materials management	Mobile Technologies, Big Data, Cloud Computing, Predictive Analytics, and Internet of Things.
Unnecessary waiting	1) better inventory management and control 2) better information management 3) more visible supply chains 4) more accurate information 5) early spotting of errors in the supply chain	Mobile Technologies, Big Data, Cloud Computing, Predictive Analytics, and Internet of Things,
Unnecessary transportation	1) better scheduling, generation, and delivery of inventory 2) better items traceability and supply network visibility 3) accurate delivery of materials	Mobile Technologies, Predictive Analytics, and Internet of Things.
Inappropriate processing	1) better production planning 2) more accurate information 3) less over-processing 4) better-quality control.	Mobile Technologies, Big Data, Cloud Computing, Predictive Analytics, and Internet of Things.
Unnecessary inventory	1) sharing of accurate information 2) better visualization and tractability 3) more accurate inventory forecasting 4) better inventory management and control 5) achieving a smooth and continuous flow of inventory through the supply chain	Mobile Technologies, Big Data, Cloud Computing, Internal Social Media, Predictive Analytics, Internet of Things, and 3D Printing

Manufacturing waste	Impact of digiting supply chains	Possible adopted technologies
Unnecessary motion	1) better information sharing 2) better traceability and visibility 3) smooth and continuous flow of inventory 4) using collaborative systems to increase teamwork 5) minimizing individuals' excess motions	Mobile Technologies, Big Data, Cloud Computing, Predictive Analytics, Internet of Things.
Defects	1) visibility of the entire supply chain 2) provide faster feedback and timely identification of errors 3) making the defect-detecting process more accurate 4) better identification of potential causes of defects 5) providing more effective technological collaboration environments	Mobile Technologies, Big Data, Cloud Computing, Internal Social Media, Predictive Analytics, and Internet of Things.

Research Theoretical Contribution

The theoretical contribution of the paper is threefold. First, in response to the limited research about the impact of supply chain digitalization on lean manufacturing wastes, for example, Buer et al. (2018) indicate that how Industry 4.0 impacts lean manufacturing and vice versa has not been studied; this study provides an empirical evidence of such impact. Second, despite the growing level of supply chain digitalization, it is not clear how such practices impact the seven types of lean manufacturing wastes. This study offers a unique insight into different benefits business are likely to result from supply chain digitalization and what the impacted areas are categorizing them based on the seven manufacturing wastes. This fills several highlighted literature gaps about the need to better understand supply chain digitalization and its potential (e.g. Bär et al., 2018; Brinch et al., 2018; Iddris, 2018; Pflaum et al., 2018). Last, the findings provide a visualized map for businesses to use that help them determine potential benefits and impacted areas should implementing technological applications within supply chains are sought. Such visualization helps assess and evaluate better digital supply chain strategies. This goes in line with similar research endeavors that examined ways to provide waste-hunting environments for Industry 4.0 applications (Satoglu et al., 2018).

Managerial Implications

This study allows to draw conclusions relevant to supply chains digitalization for organizations and extended value chains. Organizations continue adopting new technologies in supply chains; however, the 'never-stopping' emergence of disruptive technologies made some organizations hesitant about allocating the necessary resources to embrace such technologies. Such hesitation stems from; doubts about promised benefits, no clear identification of businesses areas that would be directly impacted, and fear from unforeseen potential challenges. This study provides support to top management to better comprehend the need and potential impacts of digitalizing supply chains and its business implications. Table 10 can be used as a framework helping to understand current digital supply chain status. Such understanding provides more appreciation of these potential impacts which would help with making better strategic decisions about acquiring suitable technologies, resources, and required external support. Furthermore, the outcomes from this study can support managers to develop, promote and defend the

acceptance of digitalizing supply chains projects prior to embarking on adoption journeys, during the implementation phases, and later during the operational stage. Therefore, promoting the leverage of current initiatives and success in the deployment of digital transformation solutions. Finally, lean manufacturing continues to be an important management practices and businesses continue exploring better ways to implement such philosophy. This study provides an interesting framework for businesses to use to identify continuous improvement opportunities that would strengthen the implementation of lean thinking concepts.

Limitations and Future Research

This research study sought the perception of selected participants about several pre-determined potential impacts resulting from digitalizing supply chains. Future research may explore potential impacts within the same research theme, but by using open-ended or multiple-choice survey questions. This may result in reporting different potential impacts that were not included in this study. Furthermore, this study examined potential impacts digitalizing supply chains may have on the seven examined manufacturing wastes, but it did not explore how organizations should adopt technological trends within supply chains. Thus, further studies may investigate this theme. In addition, it is suggested that further research may consider surveying participants from industry professionals preferably working for different manufacturing sectors. Finally, every effort has been made to ensure data is collected from a larger group, but this study was based on the perceptions of 73 participants and it is strongly recommended that a larger sample size is used.

REFERENCES

Agrahari, R.S., Dangle, P.A. and Chandratre, K.V., 2015. Improvement of Process cycle Efficiency by Implementing a Lean Practice: A Case Study. *International Journal of Research in Aeronautical and Mechanical Engineering*, *3*(1), pp.142-148.

Aitken, J., Childerhouse, P., Deakins, E. and Towill, D., 2016. A comparative study of manufacturing and service sector supply chain integration via the uncertainty circle model. *The International Journal of Logistics Management*, *27*(1), pp.188-205.

Ansah, R.H., Sorooshian, S. and Mustafa, S.B., 2016. Lean construction: an effective approach for project management. *Journal of Engineering and Applied Sciences*, *11*(3), pp.1607-1612.

Bär, K., Herbert-Hansen, Z.N.L. and Khalid, W., 2018. Considering Industry 4.0 aspects in the supply chain for an SME. *Production Engineering*, *12*(6), pp.747-758.

Bienhaus, F. and Haddud, A., 2018. Procurement 4.0: factors influencing the digitisation of procurement and supply chains. *Business Process Management Journal*. *24*(4), pp.965-984

Brinch, M., Stentoft, J., Jensen, J.K. and Rajkumar, C., 2018. Practitioners understanding of big data and its applications in supply chain management. *International Journal of Logistics Management, The*, (just-accepted), pp.00-00.

Buer, S.V., Strandhagen, J.O. and Chan, F.T., 2018. The link between Industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda. *International Journal of Production Research*, *56*(8), pp.2924-2940.

Büyüközkan, G. and Göçer, F., 2018. Digital Supply Chain: Literature review and a proposed framework for future research. *Computers in Industry*, *97*, pp.157-177.

Cawley, O., Wang, X. and Richardson, I., 2013. Lean software development–what exactly are we talking about?. In *Lean Enterprise Software and Systems* (pp. 16-31). Springer, Berlin, Heidelberg.

Chaudhary, K., Singh, M., Tarar, S., Chauhan, D.K. and Srivastava, V.M., 2018, April. Machine Learning Based Adaptive Framework for Logistic Planning in Industry 4.0. In *International Conference on Advances in Computing and Data Sciences* (pp. 431-438). Springer, Singapore.

Chen, J.C., Cheng, C.H. and Huang, P.B., 2013. Supply chain management with lean production and RFID application: A case study. *Expert Systems with Applications*, *40*(9), pp.3389-3397.

Chen, T. and Lin, Y.C., 2017. Feasibility evaluation and optimization of a smart manufacturing system based on 3D printing: a review. *International Journal of Intelligent Systems*, *32*(4), pp.394-413.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S., 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*.3rd ed., Lawrence Erlbaum Associates, London.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), pp.297-334.

DeVellis, R.F., 2016. *Scale development: Theory and applications* (Vol. 26). Sage Publications, Newbury Park, CA.

Eljazzar, M.M., Amr, M.A., Kassem, S.S. and Ezzat, M., 2018. Merging supply chain and blockchain technologies. *arXiv preprint arXiv:1804.04149*.

Farahani, P., Meier, C. and Wilke, J., 2017. Digital supply chain management agenda for the automotive supplier industry. In *Shaping the digital enterprise* (pp. 157-172). Springer, Cham.

Gîfu, D., Teodorescu, M. and Ionescu, D., 2014. Design of a stable system by lean manufacturing. *International Letters of Social and Humanistic Sciences*, *17*(2), pp.61-69.

Haddud, A., DeSouza, A., Khare, A. and Lee, H., 2017. Examining potential benefits and challenges associated with the Internet of Things integration in supply chains. *Journal of Manufacturing Technology Management*, *28*(8), pp.1055-1085.

Haddud, A., Dugger, J.C. and Gill, P., 2016. Exploring the Impact of Internal Social Media Usage on Employee Engagement. *Journal Social Media for Organizations*, *3*(1), pp.1-17.

Haddud, A., Dugger, J.C. and Lee, H., 2015. Lee Manufacturing Control, Asset Tracking, and Asset Maintenance: Assessing the Impact of RFID Technology Adoption. *Journal of International Technology and Information Management*, *24*(2), pp.35-54.

Haddud, A., and Khare, A. (2018). The impact of digitizing supply chains on lean operations. In Khare, A., Kessler, D., and Wirsam, J. *Market-oriented product and production management in digital environments* (pp.27-46). Springer Gabler Verlag, Heidelberg

Haghirian, P., 2016. Japanese management in the twentieth century. In *Routledge Handbook of Japanese Business and Management* (pp. 37-48). Routledge.

Hair, Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. 2006. *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Handfield, R., 2016. Preparing for the Era of the Digitally Transparent Supply Chain: A Call to Research in a New Kind of Journal.

Harris, C., Harris, R. and Streeter, C., 2016. *Lean supplier development: Establishing partnerships and true costs throughout the supply chain*. CRC Press.

Hart, M., Taraba, P. and Konečný, J., 2016. Sustainable manufacturing systems based on demand forecasting—supply chain sustainable growth. In *Sustainable Design and Manufacturing 2016* (pp. 191-202). Springer International Publishing.

Hassan, M.K., 2013. Applying lean six sigma for waste reduction in a manufacturing environment. *American Journal of Industrial Engineering*, *1*(2), pp.28-35.

Iddris, F., 2018. Digital Supply Chain: Survey of the Literature. *International Journal of Business Research and Management*, *9*(1), pp.47-61.

Ivanov, D., Dolgui, A. and Sokolov, B., 2018. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, pp.1-18.

Kang, P.S., Duffy, A., Shires, N., Smith, T. and Novels, M., 2014, November. Lean Cables–A Step towards Competitive, Sustainable and Profitable Processes. In *Proceedings of the 63rd International Wire and Cable Symposium, Providence, RI, USA* (pp. 9-12).

Karre, H., Hammer, M., Kleindienst, M. and Ramsauer, C., 2017. Transition towards an Industry 4.0 state of the LeanLab at Graz University of Technology. *Procedia manufacturing*, *9*, pp.206-213.

Kolberg, D. and Zühlke, D., 2015. Lean automation enabled by industry 4.0 technologies. *IFAC-PapersOnLine*, *48*(3), pp.1870-1875.

Korpela, K., Hallikas, J. and Dahlberg, T., 2017, January. Digital supply chain transformation toward blockchain integration. In *proceedings of the 50th Hawaii international conference on system sciences*.

Kulkarni, P.P., Kshire, S.S. and Chandratre, K.V., 2014. Productivity improvement through lean deployment & work study methods. *International Journal of Research in Engineering and Technology*, *3*(2), pp.429-434.

Kumar, A., Kumar, R.G., Narahari, N.S. and Mithilesh, K.M., 2015. Case study on identification and elimination of waste through lean implementation in an automotive part manufacturing industry. *International Journal of Lean Enterprise Research*, *1*(3), pp.253-265.

Li, Z., Wu, H., King, B., Miled, Z.B., Wassick, J. and Tazelaar, J., 2018, June. A Hybrid Blockchain Ledger for Supply Chain Visibility. In IEEE *2018 17th International Symposium on Parallel and Distributed Computing (ISPDC)* (pp. 118-125).

McDonald, R.P., 2013. *Test theory: A unified treatment*. Psychology Press. Mahwah, NJ: Erlbaum.

Meier, C., 2016. Digital supply chain management. *Digital Enterprise Transformation: A Business-Driven Approach to Leveraging Innovative IT*, p.231.

Merlino, M. and Sproģe, I., 2017. The Augmented Supply Chain. *Procedia Engineering*, *178*, pp.308-318.

Mrugalska, B. and Wyrwicka, M.K., 2017. Towards lean production in industry 4.0. *Procedia Engineering*, *182*, pp.466-473.

Nabhani, F., Uhl, C., Kauf, F. and Shokri, A., 2018. Supply chain process optimisation via the management of variance. *Journal of Management Analytics*, *5*(2), pp.136-153.

O'Leary-Kelly, S.W. and Vokurka, R.J., 1998. The empirical assessment of construct validity. *Journal of operations management*, *16*(4), pp.387-405.
Ohno, T., 1988. Toyota production system: beyond large-scale production. Portland: Productivity Press.

Okpala, C.C., 2014. Tackling muda–the inherent wastes in manufacturing processes. *International Journal of Advanced Engineering Technology*, *5*(4), pp.6-11.

Perboli, G., Musso, S. and Rosano, M., 2018. Blockchain in Logistics and Supply Chain: A Lean Approach for Designing Real-World Use Cases. *IEEE Access*, *6*, pp.62018-62028.

Pflaum, A., Prockl, G., Bodendorf, F. and Chen, H., 2018. The Digital Supply Chain of the Future: From Drivers to Technologies and Applications. Minitrack Introduction. In

The 51st Hawaii International Conference on System Sciences. HICSS 2018Hawaii International Conference on System Sciences (pp. 3924-3925). Hawaii International Conference on System Sciences (HICSS).

Rivard, S., & Huff, L. (1988). Factors of success for end-user computing. Communication of the ACM, *31*(5), pp. 552-561.

Santos, J.R.A., 1999. Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension*, *37*(2), pp.1-5.

Satoglu, S., Ustundag, A., Cevikcan, E. and Durmusoglu, M.B., 2018. Lean production systems for industry 4.0. In *Industry 4.0: Managing The Digital Transformation* (pp. 43-59). Springer, Cham.

Slack, N., Brandon-Jones, A., & Johnston, R. (2016). *Operations management* (8th ed.). Essex, UK: Pearson Education.

Suguna, S.K. and Kumar, S.N., 2019. Application of Cloud Computing and Internet of Things to Improve Supply Chain Processes. In *Edge Computing* (pp. 145-170). Springer, Cham.

Sultana, M. and Islam, M.N., 2011. Scope of Value Stream Mapping to Initiate Lean Manufacturing: An Analysis in the Apparel Industry of Bangladesh. *International Journal of Lean Thinking*, *4*(1), pp.20-30.

Tjahjono, B., Esplugues, C., Ares, E. and Pelaez, G., 2017. What does industry 4.0 mean to supply chain?. *Procedia Manufacturing*, *13*, pp.1175-1182.

Wahab, A.N.A., Mukhtar, M. and Sulaiman, R., 2013. A conceptual model of lean manufacturing dimensions. *Procedia Technology*, *11*, pp.1292-1298.

Wagner, T., Herrmann, C. and Thiede, S., 2017. Industry 4.0 impacts on lean production systems. *Procedia CIRP*, *63*, pp.125-131.

Wang, K., 2016. Intelligent predictive maintenance (IPdM) system–Industry 4.0 scenario. *WIT Transactions on Engineering Sciences*, *113*, pp.259-268.

Waller, J.L., 2012. How to Perform and Interpret Chi-Square and T-Tests. *SAS Institute Inc*.

Womack, J.P. and Jones, D.T., 2010. *Lean thinking: banish waste and create wealth in your corporation*. Simon and Schuster.

Yearling, M., 2018. Digital inventory powered by drones and robots. *Journal of Supply Chain Management, Logistics and Procurement*, *1*(1), pp.41-47.

Zsidisin, G.A. and Henke, M., 2019. Research in Supply Chain Risk: Historical Roots and Future Perspectives. In *Revisiting Supply Chain Risk* (pp. 1-12). Springer, Cham.

DECISION SCIENCES INSTITUTE

Inventory safety stock levels with uncertain supply using a service level objective

Ronald S. Lembke
University of Nevada, Reno, NV 89557
ronlembke@unr.edu

ABSTRACT

We calculate safety stock levels for a distribution center (DC) when the depot that supplies it is subject to potential outages. We assume both the DC and the depot follow a Type I (stockout probability) service level. Under this objective, outages will inevitably occur at the depot. In this case, the lead time to the DC will be unexpectedly increased. Under these conditions, the probability function of demand over the lead time at the DC becomes bimodal. An algorithm for the optimal reorder point is presented.

KEYWORDS: Inventory Management, Inventory Coordination, Retail Inventory, Supply Chain Management, Omni-channel Retail

INTRODUCTION

Timely, accurate fulfillment of customer orders is mandatory, and accomplishing this requires effective inventory management. To satisfy customers, firms must have the right inventory, in the right place, at the right time. Profitability also requires efficiency, minimizing costs.

In this paper, we discuss ways firms can determine optimal inventory levels to meet customer demand for consumer products, in an omni-channel environment, where supplier deliveries are uncertain.

In the omni-channel context, supply chain coordination is of utmost importance. The costs of service failures have never been greater. Disgruntled reviewers can leave negative reviews on social media and retail sites that can influence other shoppers' decisions. Most customers are engaged in social media and complain about service failures to their friends and their followers.

Not only can every customer complaint be potentially widely disseminated, but in this global marketplace, every other competitor is just one click away. Customers can easily leave one retail platform, or one vendor on a platform, for another one.

Specifically, we study the optimal amount of safety stock to plan for, when the lead time has a probability P_1 of being L_1, and probability of $(1 - P_1)$ of being $L_1 + L_2$. We present results for the general case of demand probability distribution, and more specific results for Normally distributed demands.

CHANNEL STRUCTURE

Our focus is the retail distribution channel. The number of levels is not the only relevant information about a distribution channel. The structure is critically important. In Figure 1, some common distribution network configurations are shown.

Figure 1a represents the simplest channel structure, where the retailer sells directly to the consumer. A distribution channel of a single retailer has this structure, and customers travel to the retail store. A channel that consists of a single online fulfillment center (FC) has the same

structure, but products are shipped directly from the fulfillment center to the customers. In either case, the retailer's challenge is to ensure sufficient inventory levels at a single location to meet demands.

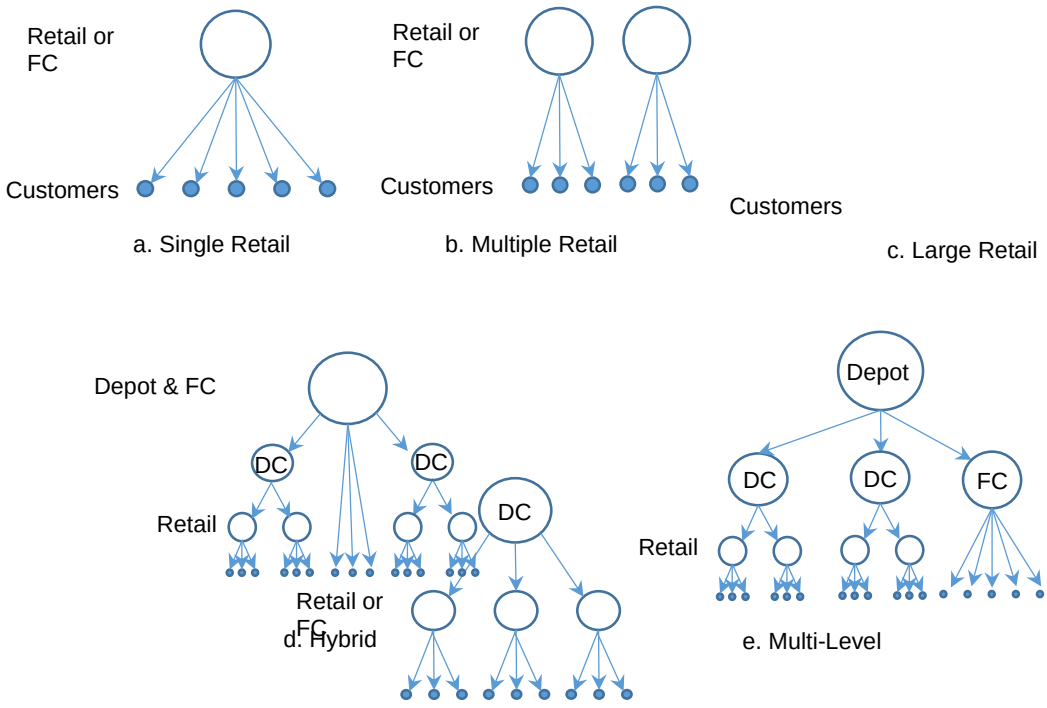
If a retailer operates multiple retail stores, or multiple fulfillment centers, the structure is shown in Figure 1b. If the firm has multiple retail locations or fulfillment centers, and does not have a distribution center (DC), it may allow the locations to share inventory, to pool the risk of a shortage.

A retailer with a DC that serves its retail or fulfillment center locations is shown in Figure 1c. The retailer must decide on the levels to hold at each retail location, as well as the amount to be kept in the DC.

Some retailers may attempt to ship orders directly to customers from the retail distribution center, using the DC also as an FC, as in Figure 1d, so the Depot must to both retail stores and individual customers. This was often not been considered a best practice, because the two activities are so different. Shipments to retail stores involve full truckloads, split among multiple stores, often with pallets of a single product. Fulfillment center activities involve picking multiple single items, and packing and labeling them for delivery, which is a very different. A more common approach is to use separate Fulfillment Centers, as in Figure 1e.

If a firm has multiple DCs, they may be fed, in turn, from a central Depot, as in Figure 1e. The Depot may be located by a manufacturing facility, or be a warehouse that serves multiple regional DCs and/or FCs. The complex situation of 1e is the focus of this paper.

Figure 1. Distribution Channel Structures



Most of the retailers operating in the omni-channel fall into the multi-level category, shown in Figure 1e. Instead of only having DCs that feed retail locations, they also have FCs that ship

product directly to the customers, when customers place an order online and have it sent directly to them. This has required retailers to create these new FCs and learn how to manage them.

If customers request in-store pick-up, store employees pick the items and set them aside for the customers, which is a challenge, figuring out how to incorporate these new responsibilities inside of an existing retail facility. From the perspective of ensuring sufficient supply of product to meet demand, fulfillment from the store may increase both the amount and variability in demand.

Discrete Time, Continuous Review

In this paper, we consider the case of discrete time, with continuous review.

An important distinction must be made between the two ways that time may be considered in inventory management: continuously, or discretely. With continuous time models, demands arrive randomly throughout time, and an order may be placed at any moment, when needed. With discrete time models, time is divided into periods, like days, and all of the demands that arrive during a given day are treated as the demand of that day.

Most companies would seem to operate using a discrete time frame. Although customer orders arrive continuously throughout the day, the realities of shipping enforce a discrete nature: either an order is received early enough in the day to be packed and shipped before the cutoff for the carrier, or not. A replenishment order to the supplier has the same character: either the order is placed before the supplier's cutoff for that day, or it is not. If two orders were placed to one supplier on the same day, they would be sent by the supplier on the same day, and most likely received by the company on the same day, possibly in the same delivery. Orders placed in continuous time end up being discretized.

As Lee and Nahmias (1993) observed, "From our experience, practitioners seem to favor the continuous review model" (p. 33), and we will focus on continuous review models.

The second distinction is in the flexibility in the timing of ordering. *Reorder point* methods assume an order can be placed on any day, while *Fixed time-period* methods assume that orders must be placed according to a fixed time schedule. With reorder point methods, the inventory level must be continually monitored, and once it falls to the reorder point, a replenishment order is placed. With fixed time-period methods, at same point in every period, for example every Wednesday, an order is placed, and the on-hand inventory plus the order quantity will be equal to the "order up to level."

LITERATURE REVIEW

Inventory reordering policies in the case of uncertain demand have long been a central area of study for inventory control (e.g., Arrow, et al., 1951; Arrow, et al., 1958; Morse, 1958; Wagner, 1962). Hadley and Whitin (1963) considered continuously arriving demand following a Poisson distribution. In their treatment of a random lead time, they assumed the lead time can be fitted to a gamma distribution.

For continuous review methods, Dvoretzky, et al. (1953) demonstrated that an (s, S) policy is optimal under a number of conditions related to ordering, holding, and stockout costs. Under such a policy, when on-hand inventory falls below s, an order of $S - s$ units is placed.

The best-known result for random lead time assumes the random demand and lead time are independent, and presents a simple formula for estimating the standard deviation of demand over the lead time (Silver, et al., 1998), when the lead time is Normally distributed. We are interested in the case where lead times are not Normally distributed, but discrete.

When lead times are random, there is the possibility that orders may cross in the process, and an order placed later may arrive before placed earlier. We assume that orders are filled in the order they are placed, a frequent assumption (e.g. Zipkin, 1986).

If variability of demand per day follows a Normal distribution, and the lead time can only take on a small number of possible values, assuming that the total demand over the lead time fits a Normal distribution can be a very poor approximation (Eppen & Martin, 1988). They show how to combine the demand distributions from multiple lead times with different probabilities of occurrence. Our result goes further, explicitly formulating a computationally more direct method for determining the safety stock levels.

Tyworth (1992) follows Eppen & Martin, but assumes the average and standard deviation of demand over a given number of periods can be expressed, as a different Normal distribution for each number of periods.

Chopra, Reinhardt & Dada (2004) follow Eppen & Martin, and study the errors generated by using the Normal approximation for total demand over the lead time, assuming lead time has a uniform or gamma distribution.

Cobb (2013) makes similar assumptions to ours, modeling demand as a mixture distribution, and derives an approximation for total lead time demand, while we formulate an exact formulation. Cobb, *et al.* (2015) extend this line of inquiry with by approximating demand over the lead time with mixtures of polynomials.

Bagchi & Hayya (1984) assume Normally distributed demand, but Erlang lead times. Wang, Zinn, and Croxton (2010) consider correlated demand and lead time. Bookbinder and Çakanyildirim (1999) consider expediting as a decision variable under random lead times. Bagchi, Jack, and Chu (1986) study the impact of using an inappropriate approximation of a random lead time.

Following Eppen & Martin (1988), we present mathematical results for the solution of safety stock levels for general demand functions, and a direct mathematical solution for Normally distributed demand. Normally distributed demands have long been used to represent high-volume products. Normally distributed demands also have the added benefit of being known and understood by practitioners, reducing barriers to implementation of the results.

Reorder Points

To calculate safety stock levels, we will focus on reorder point levels. The reorder point calculation begins with the expected demand during the lead time between placing the order and its arrival, which we will call D_L.

If demand had no variability, the demand over the lead time would be $D_L = dL,$ where d is the demand per day, and L is the lead time in days, and the reorder level, R, would be given by $R = D_L = dL$. If demand is variable, but known in advance, R can be calculated by summing the demands over the next L days $D_L = \sum_{t=1}^{L} d_t$, where d_t is the demand on day t.

Unfortunately, few companies have the luxury of knowing the demand in advance each day. To be useful, an inventory calculation must incorporate randomness.

Safety Stock Calculations
When demand is random, the reorder level must be increased by some amount of safety stock, so the reorder level becomes $R = D_L + SS_L$, where the amount of safety stock, SS_L, is chosen to meet the performance target specified by the firm.

Some firms set a service level by minimizing the total of ordering, holding and stockout costs, but stockout costs are problematic to estimate. The most commonly used objective is the probability of not stocking out during the replenishment period (stockout probability), known as a Type I service level.

Type I Service Levels – Stockout Probability
With a Type I service level, the probability of not running out during the reorder interval must be at least as high as the target, for example, 95%. If the desired service level is P_1, and $F(x)$ is the cumulative distribution function of demand, then R is set such that

$$F(R) = P_1 \quad (1)$$

Therefore, $R = F^{-1}(P_1)$. If the randomness of total demand over the lead time can be approximated by the Normal distribution, the reorder level is

$$R = \mu_L + z\sigma_L \quad (2)$$

where μ_L is the average demand over the lead time, σ_L is the standard deviation of demand over the lead time, and the number of standard deviations of safety stock, z, is given by $z = \Phi^{-1}(P_1)$, using the inverse Normal distribution.

If the randomness is stable and the average and standard deviation are unchanged from day to day, $R = \mu L + z\sqrt{L}\sigma$, where μ is the average per day, L is the lead time in days, σ is the standard deviation per day.

REORDER POINTS IN MULTI-LEVEL NETWORKS
In a large network, DCs directly satisfy retailer demands, as in Figure 1c. In many networks, a Depot provides inventory for the DCs and FCs, as in Figure 1e.

Such networks can be referred to as multi-echelon networks. In 1c, the retail stores are the first echelon, and then the first echelon plus the DCs comprise second echelon, etc. One primary finding from this area of research is that the optimal amount of inventory in an echelon is dependent on the total demand and variability within that echelon.

Put another way, this result says that the amount of inventory in each location should be whatever is required to meet its needs and those of its customers. Looking at the DC, the amount in the stores plus the DC is enough to satisfy the total demand across all locations.

The details of such networks can quickly become non-intuitive. Some years ago, the author spent six months as a visiting professor in the corporate offices of a Fortune 30 manufacturing company. My colleagues had decades of experience in manufacturing and distribution. One of

the projects I worked on involved this exact issue, setting inventory levels and safety stocks at DCs and a Depot that was subject to stockouts.

Although many sophisticated models have been developed for multi-echelon systems, the reality is that many inventory managers do not have the luxury of sophisticated systems, and must manage inventories at each location independently. Most of the research on random lead times has focused on lead times that may take on a large number of values, such as when the lead time is Normally distributed (Silver, Pyke and Peterson 1998, 280-284).

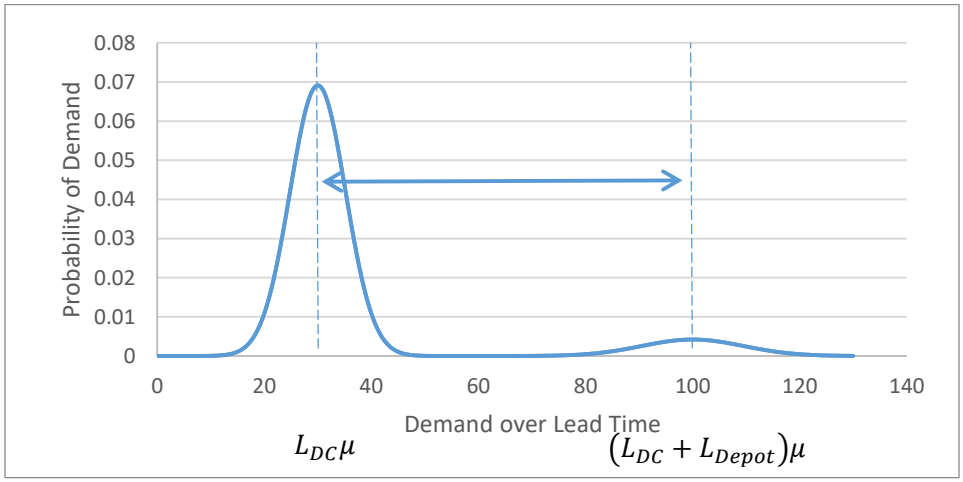
REORDER POINT WITH UNCERTAIN DEPOT SUPPLY

Consider Figure 1e. Suppose the DCs have a service level of $P_{DC} = 99\%$. What should the service level for the Depot be? The answer given by my experienced colleagues when I was a visiting professor was that the Depot can have a lower service level, because it is farther from the customer, and not involved in meeting customer demand on a daily basis. For example, say $P_{Depot} = 90\%$.

The lead time from the Depot to the DCs was short, 3 days. But the Depot received shipments from the production plant, and its lead time was a week. Thus, the lead time to the DCs was usually 3 days, but if the Depot was out of inventory, the order would have to wait for the replenishment, which could be as long as an additional 7 days. So 90% of the time, $L = 3$, but there is a 10% chance that $L = 10$.

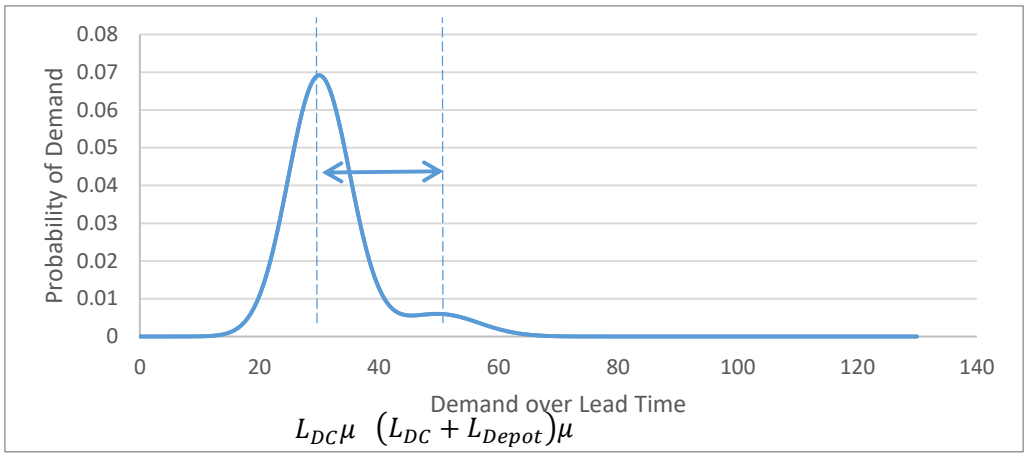
Assuming Normally distributed daily demand, the demand over the lead time is then a mixture of two Normal distributions, and generally bi-modal. If the lead time has a 90% chance of being 3 days, and a 10% chance of being 10 days, and the average per day is $\mu = 10$, with standard deviation $\sigma = 3$, the probability of lead time demand is shown in Figure 2a.

Figure 2a. Probability of Lead Time Demand, LT_{DC} = 3 days, LT_{Depot} = 7 days, Depot Service Level is 90%, μ = 10, σ = 3.



If the lead time at the Depot is only 2 days, the bi-modal nature is less obvious, as in Figure 2b.

Figure 2b. Probability of Lead Time Demand, when LT_{DC} = 3 days, LT_{Depot} = 2 days, Depot Service Level is 90%, μ = 10, σ = 3



The demand probability in 2a is clearly bimodal. The area under the left peak is 0.9, and under the right is 0.1. If the DC wants to achieve a stockout probability of say, $P_{DC} = 0.99, R$ would need to be set in the right half of the distribution. The distance between the peaks is determined by L_{Depot}, and the amount of overlap depends on the dispersion of each peak, which depends on L_{DC}, L_{Depot}, and σ.

Result 1: For a DC service level of P_{DC}, with lead time L_{DC}, the reorder point at the DC, R, is chosen to satisfy the following equation, given that demand is Normally distributed with average per day μ and standard deviation σ, and the Depot has reorder time of L_{Depot} and service level of P_{Depot}:

$$F_{DC}(R) = P_{Depot} * \Phi\left(\frac{R - \mu L_{DC}}{\sqrt{L_{DC}}\sigma}\right) + \left(1 - P_{Depot}\right) * \Phi\left(\frac{R - \mu\left(L_{DC} + L_{Depot}\right)}{\sqrt{L_{DC} + L_{Depot}}\sigma}\right) = P_{DC} \qquad (3)$$

■

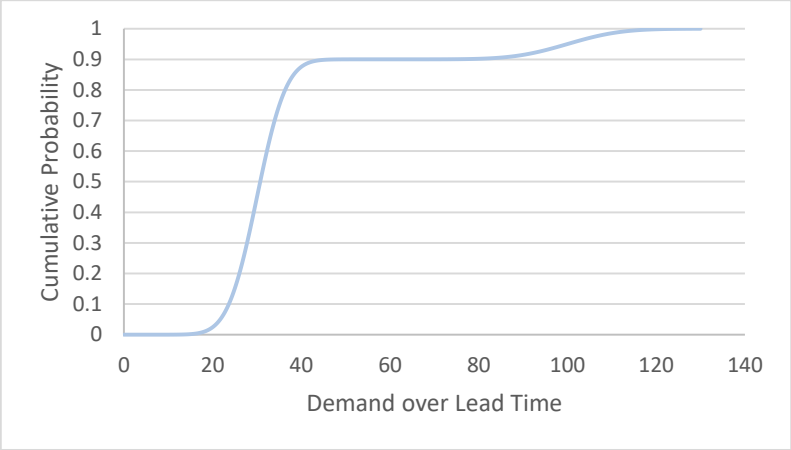
Proof: For service level objectives, following (1) above, R is set such that the probability of demand being less than or equal to R is P_1, in this case, P_{DC}. With uncertain service by the Depot, the probability of demand during the lead time being less than or equal to x is a mixture distribution given by

$$F(x) = P_{Depot}F_{DC}(x) + \left(1 - P_{Depot}\right)F_{Depot}(x) \qquad (4)$$

where P_{Depot} is the Type I service level at the Depot, $F_{DC}(x)$ is the probability of demand being less than or equal to x over the lead time when the shipment comes directly from the Depot, and $F_{Depot}(x)$ is the probability of demand less than or equal to x when the shipment has to wait for the Depot to be replenished. Because demand is Normally distributed, $F_{DC}(x)$ and $F_{Depot}(x)$ can be written in terms of the standard Normal distribution with means of μL_{DC} and $\mu\left(L_{DC} + L_{Depot}\right)$, and standard deviations of $\sqrt{L_{DC}}\sigma$ and $\sqrt{L_{DC} + L_{Depot}}\sigma$, which leads to the result. ■

Unfortunately, in Result 1, R cannot be solved for directly. However, as a cumulative distribution function (cdf), $F_{DC}(R)$ in non-decreasing in R, as shown in Figure 3, and the process for searching or the optimal R is to search on the function until the desired value is found.

Figure 3. Cumulative Probability of Lead Time Demand, when DC LT is 3 days, Depot LT is 2 days, and Depot Service Level is 90%, μ = 10, σ = 3



ITERATIVE SOLUTION PROCESS

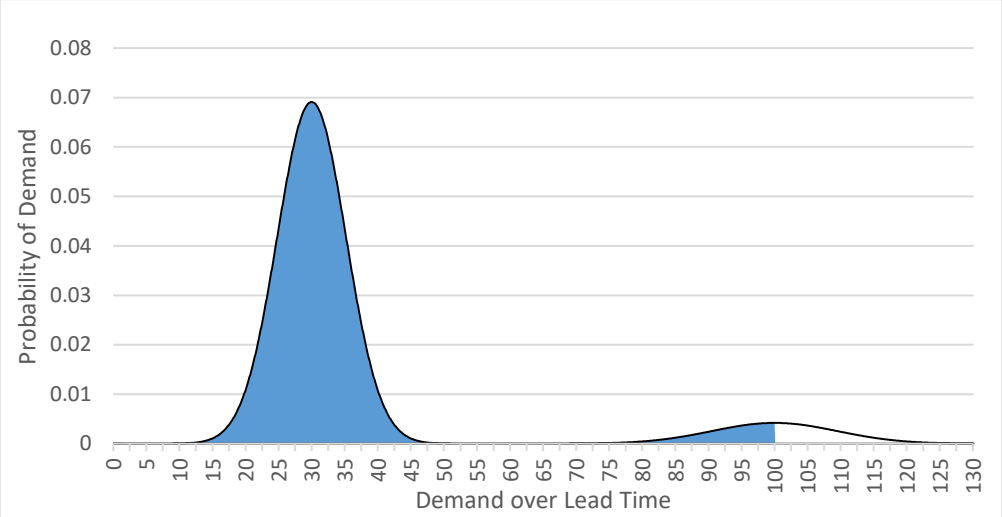
To present an iterative process, we must separate the analysis into three separate cases, depending on the relative sizes of P_{DC} and P_{Depot}.

Case 1: $P_{DC} > P_{Depot}$

We begin with the case where the desired service level at the DC is higher than the service level at the Depot that serves it. For example, in Figure 4, below, the Depot provides a 90% service level, but the DC must achieve a 95% service level.

In this case, the DC cannot rely solely on the service provided by its own inventory. To provide the service, is it is dependent on the arrival of inventory from the depot.

Figure 4. 95% Service Level at DC, when DC LT is 3 days, Depot LT is 7 days, and Depot Service Level is 90%, μ = 10, σ = 3



Result 2a: If $P_{DC} > P_{Depot}$, for a DC service level of P_{DC}, the Reorder point at the DC, R, may be found via the following algorithm, given that demand is Normally distributed with average per day μ and standard deviation σ, with lead time L_{DC}, and the Depot has reorder time of L_{Depot} and service level of P_{Depot}:

1. Set $R = \mu(L_{DC} + L_{Depot}) + \Phi^{-1}\left(\frac{P_{DC}-P_{Depot}}{1-P_{Depot}}\right)\sqrt{L_{DC} + L_{Depot}}\sigma$,
2. Calculate $\delta_{DC}(R) = 1 - \Phi\left(\frac{R-\mu L_{DC}}{\sqrt{L_{DC}}\sigma}\right)$
3. Revise R to reflect the amount by which step 1 over-estimates the probability contribution from the DC: $R = \mu(L_{DC} + L_{Depot}) + \Phi^{-1}\left(\frac{P_{DC}-P_{Depot}(1-\delta_{DC}(R))}{1-P_{Depot}}\right) * \sqrt{L_{DC} + L_{Depot}}\sigma$
4. Repeat 2 and 3 until R converges to within desired tolerance.

Proof: From (4) above, R is to be chosen such that:

$$F(R) = P_{Depot}F_{DC}(R) + (1 - P_{Depot})F_{Depot}(R) = P_{DC}. \quad (5)$$

If $P_{DC} > P_{Depot}$, and the two modes of the mixture distribution $F_{DC}(x)$, are far enough apart (as in Figure 4), then likely values of R may be so far above the average demand over the DC lead time (μL_{DC}, the mode of the left half of the distribution), that $F_{DC}(R) \approx 1.0$. In this case,

$$P_{Depot} + (1 - P_{Depot})F_{Depot}(R) = P_{DC},$$

which can be solved for $F_{Depot}(R)$:

$$F_{Depot}(R) = \frac{P_{DC} - P_{Depot}}{1 - P_{Depot}},$$

which we can directly solve for R:

$$R = F_{Depot}^{-1}\left(\frac{P_{DC} - P_{Depot}}{1 - P_{Depot}}\right).$$

For Normally distributed demands, we may then write:

$$R = \mu\left(L_{DC} + L_{Depot}\right) + \Phi^{-1}\left(\frac{P_{DC} - P_{Depot}}{1 - P_{Depot}}\right) * \sqrt{L_{DC} + L_{Depot}}\sigma \quad (6)$$

This approximation requires $F_{DC}(R) \approx 1.0$. How far apart the modes of the mixture distribution must be for this to hold depends on L_{DC}, L_{Depot}, μ, and σ, and how closely one wants $F_{DC}(R)$ to approximate 1.0. Even for very large values of R, any approximation of $F_{DC}(R) = 1$ will be off by

$$\delta_{DC}(R) = 1 - F_{DC}(R) = 1 - \Phi\left(\frac{R - \mu L_{DC}}{\sqrt{L_{DC}}\sigma}\right).$$

Because $\delta_{DC}(R)$ represents the amount by which the assumption of $F_{DC}(R) \approx 1.0$ over-estimates $F_{DC}(R)$, then the actual service provided by R in (5) is given by

$$F(R) = P_{Depot} * \left(1 - \delta_{DC}(R)\right) + \left(1 - P_{Depot}\right) * F_{Depot}(R).$$

Setting this equal to P_{DC}, and solving for R as before, we obtain

$$P_{Depot} * \left(1 - \delta_{DC}(R)\right) + \left(1 - P_{Depot}\right) * F_{Depot}(R) = P_{DC}$$

Solving for $F_{Depot}(R)$,

$$F_{Depot}(R) = \frac{P_{DC} - P_{Depot} * \left(1 - \delta_{DC}(R)\right)}{\left(1 - P_{Depot}\right)}$$

This gives the following value for R:

$$R = \mu\left(L_{DC} + L_{Depot}\right) + \Phi^{-1}\left(\frac{P_{DC} - P_{Depot}\left(1 - \delta_{DC}(R)\right)}{1 - P_{Depot}}\right) * \sqrt{L_{DC} + L_{Depot}}\sigma \quad (7)$$

■

Step 1 over-estimates, by $\delta_{DC}(R)$, the service provided by the DC, which results in an R which is too small. $\delta_{DC}(R)$ is the additional service that must be provided by the Depot in step 3, which leads to a larger R. But with this larger value of R, the DC is also providing slightly more service, which means that the total service provided is greater than it needs to be, so in the next iteration, R will be lowered, but not as much, only to be raised, until the optimal R value is reached. This process of alternately raising and lower continues until the optimal value is found.

Experimentally, the algorithm generally converges in a very small number of iterations, and broadly speaking, the convergence is proportional to the bimodality of the conditions. Using the conditions from Figure 4, the solution is found on the first pass, and $\delta_{DC}(R) = 0$. If $L_{Depot} = 1$, with all other parameters unchanged, the pdf clearly looks unimodal, as shown in Figure 5a, and 26 iterations are required for the subsequent R values to converge to within 0.1, as shown in Figure 5b.

Figure 5a. 95% Service Level at DC, when DC LT is 3 days, Depot LT is 1 day, Depot Service Level is 90%, μ = 10, σ = 3

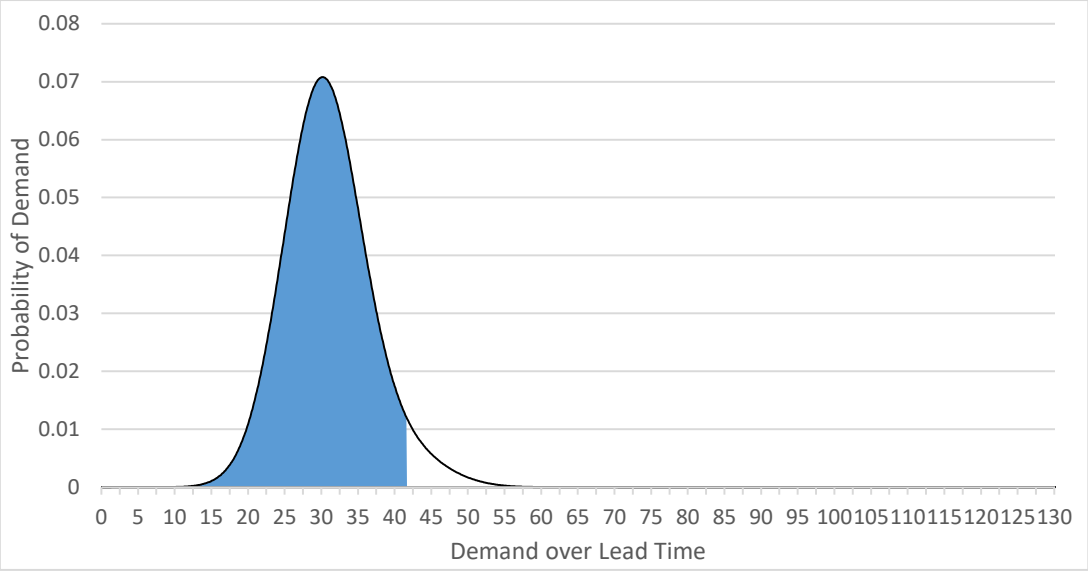
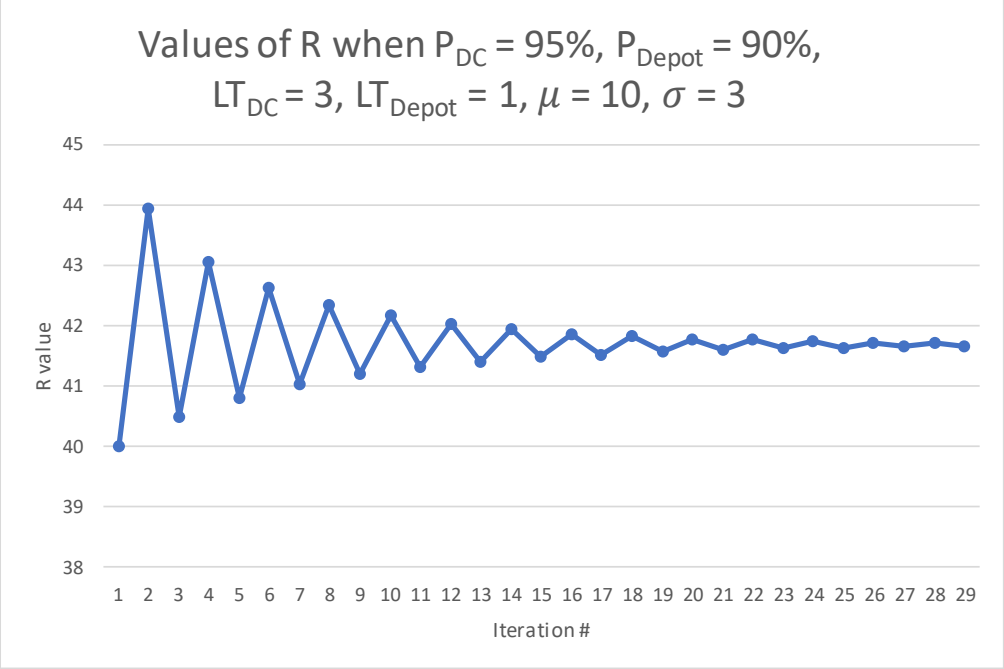


Figure 5b. Values of R, by iteration, DC Service Level 95%, Depot Service Level 90%, DC LT is 3 days, Depot LT is 1 days, μ = 10, σ = 3.



Case 2: $P_{DC} < P_{Depot}$

In this case, the target service level at the DC is low enough that the DC can always rely on its own inventories to be high enough to provide the expected service level.

Result 2b: If $P_{DC} < P_{Depot}$, for a DC service level of P_{DC}, the Reorder point at the DC, R, may be found via the following algorithm, given that demand is Normally distributed with average per day μ and standard deviation σ, with lead time L_{DC}, and the Depot has reorder time of L_{Depot} and service level of P_{Depot}. If $P^{DC} < P^{Depot}$

1. Set $R = \mu L_{DC} + \Phi^{-1}\left(\frac{P_{DC}}{P_{Depot}}\right) * \sqrt{L_{DC}}\sigma$,
2. Calculate $\delta_{Depot}(R) = \Phi\left(\frac{R-\mu(L_{DC}+L_{Depot})}{\sqrt{L_{DC}+L_{Depot}}\sigma}\right)$
3. Revise R to reflect the amount by which 1 under-estimated the probability contribution from the Depot: $R = \mu L_{DC} + \Phi^{-1}\left(\frac{P_{DC}-(1-P_{Depot})\delta_{Depot}(R)}{P_{Depot}}\right) * \sqrt{L_{DC}}\sigma$
4. Repeat 2 and 3 until R converges to within desired tolerance.

Proof: As above, the target service level of R is:

$$F(R) = P_{DC} = P_{Depot}F_{DC}(R) + (1 - P_{Depot})F_{Depot}(R).$$

If $P^{DC} < P^{Depot}$, then values of R may be so far below the average demand over the Depot lead time $\mu(L_{DC} + L_{Depot})$, that $F_{Depot}(R) \approx 0$. In this case,

$$F(R) = P_{DC} = P_{Depot}F_{DC}(R),$$

which can be solved for $F_{DC}(R)$:

$$F_{DC}(R) = \frac{P_{DC}}{P_{Depot}}.$$

In this case, we can set $R = F_{DC}^{-1}\left(\frac{P_{DC}}{P_{Depot}}\right) = \mu L_{DC} + \Phi^{-1}\left(\frac{P_{DC}}{P_{Depot}}\right) * \sqrt{L_{DC}}\sigma.$
This approximation requires $F_{Depot}(R) \approx 0$, and any assumption of $F_{Depot}(R) = 0$ will be off by

$$\delta_{Depot}(R) = F_{Depot}(R) = \Phi\left(\frac{R - \mu(L_{DC} + L_{Depot})}{\sqrt{L_{DC} + L_{Depot}}\sigma}\right).$$

Because $\delta_{Depot}(R)$ represents the amount by which the assumption $F_{Depot}(R) \approx 0$ under-estimates $F_{Depot}(R)$, then the actual service provided by R in (5) is given by

$$F(R) = P_{Depot}F_{DC}(R) + (1 - P_{Depot})\delta_{Depot}(R).$$

Setting this equal to P_{DC}, and solving for $F_{DC}(R)$, we obtain

$$F_{DC}(R) = \frac{P_{DC} - (1 - P_{Depot})\delta_{Depot}(R)}{P_{Depot}}$$

Which yields:

$$R = \mu * L_{DC} + \Phi^{-1}\left(\frac{P_{DC} - (1 - P_{Depot})\delta_{Depot}(R)}{P_{Depot}}\right) * \sqrt{L_{DC}}\sigma \quad (8)$$

■

The same convergence properties are observed for Result 4b as were shown above for Result 2a.

Case 3: $P_{DC} = P_{Depot}$

Finally, we consider the case where the two probabilities are equal to each other. This might seem at first, to imply that the DC could rely on only its own inventories, but because of the overlapping of the distributions, to do so would be inaccurate.

Result 2c: If $P_{DC} = P_{Depot} = P$, for a DC service level of P_{DC}, the Reorder point at the DC, R, may be found via the following algorithm, given that demand is Normally distributed with average per day μ and standard deviation σ, with lead time L_{DC}, and the Depot has reorder time of L_{Depot} and service level of P_{Depot}. If $P^{DC} = P^{Depot} = P$,

1. Set $R = \mu L_{DC} + \Phi^{-1}(P) * \sqrt{L_{DC}}\sigma$,
2. Calculate $\delta_{Depot}(R) = \Phi\left(\frac{R-\mu(L_{DC}+L_{Depot})}{\sqrt{L_{DC}+L_{Depot}}\sigma}\right)$
3. Revise R to reflect the amount by which 1 under-estimated the probability contribution from the Depot: $R = \mu L_{DC} + \Phi^{-1}\left(1 - \frac{(1-P)}{P}\delta_{Depot}(R)\right) * \sqrt{L_{DC}}\sigma$
4. Repeat 2 and 3 until R converges to within desired tolerance.

Proof: We will begin with the same assumption as in the case of $P_{DC} < P_{Depot}$, by assuming that R is small enough that $F_{Depot}(R) \approx 0$. In this case, $P^{DC} = P^{Depot} = P$,

$$F(R) = P = PF_{DC}(R) + (1-P)F_{Depot}(R)$$

Solving for $F_{DC}(R)$, we obtain $F_{DC}(R) = 1$, which requires $R = \infty$. It can be approximated by setting R extremely high, but without considering the difference in means (determined by L_{Depot}) and the relative size of σ, it is difficult to ensure that the R chosen this way is truly an overestimate.

To guarantee that our initial estimate of R is an underestimate, we will assume all of the service level is provided by the DC. In this case, we may then set $F_{DC}(R) = P$, giving

$$R = \mu L_{DC} + \Phi^{-1}(P) * \sqrt{L_{DC}}\sigma.$$

This also assumes that $F_{Depot}(R) \approx 0$, which clearly underestimates $F_{Depot}(R)$. Given the assumptions above, the service provided by the Depot, $F_{Depot}(R)$, is not 0, but rather the following amount:

$$\delta_{Depot}(R) = F_{Depot}(R) = \Phi\left(\frac{R-\mu(L_{DC}+L_{Depot})}{\sqrt{L_{DC}+L_{Depot}}\sigma}\right)$$

Because we assumed $F_{Depot}(R) = 0$, our error is $\delta_{Depot}(R) = F_{Depot}(R) - 0 = F_{Depot}(R)$. The actual service provided by R is given by

$$F(R) = PF_{DC}(R) + (1-P)\delta_{Depot}(R).$$

Because our goal is to find $F(R) = P$, we set this equal to P and solve for $F_{DC}(R)$,

$$F_{DC}(R) = 1 - \frac{(1-P)}{P} * \delta_{Depot}(R).$$

This gives us a new value of R:

$$R = \mu L_{DC} + \Phi^{-1}\left(1 - \frac{(1-P)}{P}\delta_{Depot}(R)\right) * \sqrt{L_{DC}}\sigma$$

■

Combined Solution to All Relative Values of P_{DC}, P_{Depot}

Having separately considered the cases of $P_{DC} > P_{Depot}$, $P_{DC} < P_{Depot}$, and $P_{DC} = P_{Depot}$, we can now combine the results into one expression.

Result 3 For a DC service level of P_{DC}, the Reorder point at the DC, R, may be found via the following algorithm, given that demand is Normally distributed with average demand per day μ and standard deviation σ, with lead time L_{DC}, and the Depot has reorder time of L_{Depot} and service level of P_{Depot}:

1. Set $R = \begin{cases} \mu(L_{DC} + L_{Depot}) + \Phi^{-1}\left(\frac{P_{DC} - P_{Depot}}{1 - P_{Depot}}\right) * \sqrt{L_{DC} + L_{Depot}}\sigma & if \quad P_{DC} > P_{Depot} \\ \mu L_{DC} + \Phi^{-1}(P) * \sqrt{L_{DC}}\sigma & if \quad P_{DC} = P_{Depot} = P \\ \mu L_{DC} + \Phi^{-1}\left(\frac{P_{DC}}{P_{Depot}}\right) * \sqrt{L_{DC}}\sigma & if \quad P_{DC} < P_{Depot} \end{cases}$

2. Calculate $\delta_{Depot}(R) = \begin{cases} 1 - \Phi\left(\frac{R - \mu L_{DC}}{\sqrt{L_{DC}}\sigma}\right) & if \quad P_{DC} > P_{Depot} \\ \Phi\left(\frac{R - \mu(L_{DC} + L_{Depot})}{\sqrt{L_{DC} + L_{Depot}}\sigma}\right) & if \quad P_{DC} \leq P_{Depot} \end{cases}$

3. Revise R to reflect the amount by which previous step over or under-estimated the probability contribution from the Depot: $R = \begin{cases} \mu(L_{DC} + L_{Depot}) + \Phi^{-1}\left(\frac{P_{DC} - P_{Depot}(1 - \delta_{DC}(R))}{1 - P_{Depot}}\right) * \sqrt{L_{DC} + L_{Depot}}\sigma & if \quad P_{DC} > P_{Depot} \\ \mu L_{DC} + \Phi^{-1}\left(1 - \frac{(1-P)}{P} * \delta_{Depot}(R)\right) * \sqrt{L_{DC}}\sigma & if \quad P_{DC} = P_{Depot} = P \\ \mu L_{DC} + \Phi^{-1}\left(\frac{P_{DC} - (1 - P_{Depot})}{P_{Depot}} * \delta_{Depot}(R)\right) * \sqrt{L_{DC}}\sigma & if \quad P_{DC} < P_{Depot} \end{cases}$

4. Repeat steps 2 and 3 until values of R converge.

Proof: Follows from combining Results 2a-2c.

■

Note that when $P^{DC} = P^{Depot}$, the term $\Phi^{-1}\left(\frac{P_{DC}-P_{Depot}}{1-P_{Depot}}\right)$ cannot be evaluated, meaning the result from 4a cannot be used in this situation. Similarly, the term $\Phi^{-1}\left(\frac{P_{DC}}{P_{Depot}}\right)$ also cannot be evaluated, which means the result from 4b cannot be used, either, which means a separate result is required for $P^{DC} = P^{Depot}$.

Computational Convergence
As shown above, experimentally, the algorithm generally converges in a very small number of iterations, and broadly speaking, the convergence is proportional to the bimodality of the conditions.

For example, using the parameters from Figure 5, and setting $L_{DC} = 3$, in order to generate an error, we must set $L_{Depot} < 0.9$ days. If we set $L_{Depot} = 7$, then we must set $\sigma > 25$ to generate an error. In other words, the model only fails in highly unlikely situations, such as extremely fast Depot replenishment or extremely uncertain demand, and behaves well in any range of likely demand parameters.

Under some conditions, algorithms for Results 2a and 2b will not converge and Result 1 must be used. Specifically, in step 3 of Result 2a, $\Phi^{-1}\left(\frac{P_{DC}-P_{Depot}(1-\delta_{DC}(R))}{1-P_{Depot}}\right)$ cannot be evaluated if the fraction is greater than 1, which happens when $\delta_{DC}(R)$, the error term, is large.

A similar issue arises in step 3 of Result 4b. A complete discussion is beyond our scope, but the issue depends on a complex inter-relationship between $L_{DC}, L_{Depot}, \mu,\ P_{DC,}P_{Deopt,}$ and σ.

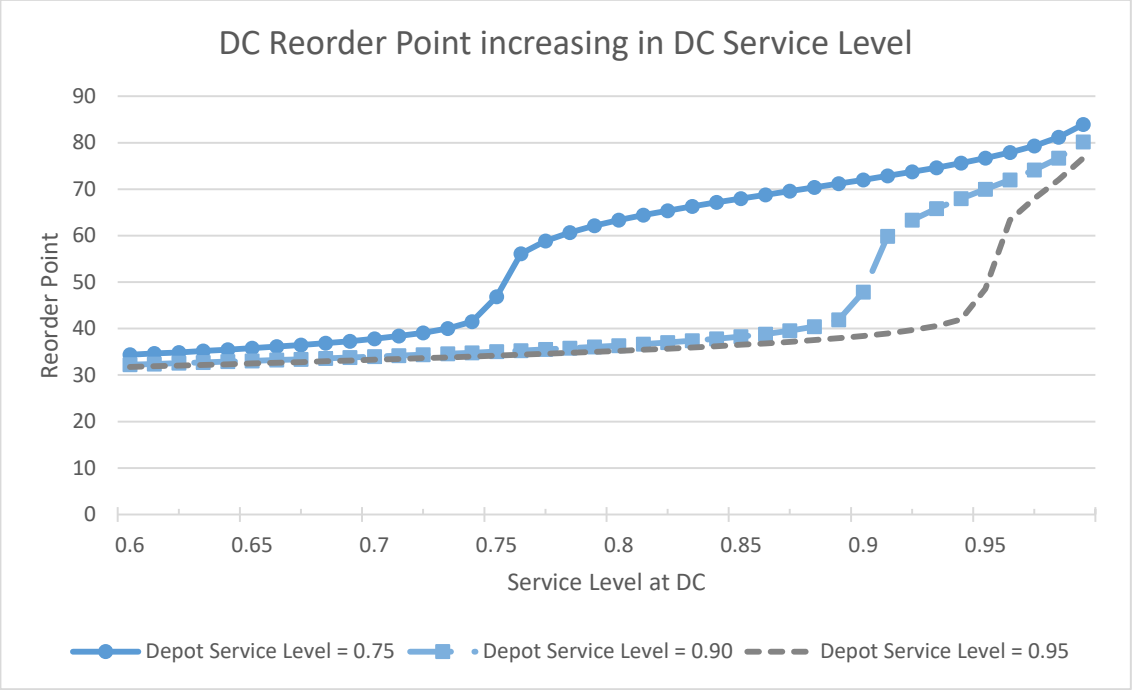
In practice, it only arises when L_{Depot} is small, σ is large, and $|P_{DC} - P_{Depot}|$ is small. For example, in computational experiments, if $\mu = 10, \sigma = 3, P_{DC} = 0.95, P_{Depot} = 0.9$, and $L_{DC} = 3$, in order to generate an error, we must set $L_{Depot} < 0.9$. If we set $L_{Depot} = 7$, then we must set $\sigma > 25$ to generate an error. In other words, the model only fails in highly unlikely situations, such as extremely fast Depot replenishment or extremely uncertain demand, and behaves well in any range of likely demand parameters.

IMPACT OF TARGET SERVICE LEVELS

In Figure 6, we can see how DC inventories increase as the desired service level increases at the DC. In the top line of the graph, the service level at the Depot is 0.75. As the desired service level at the DC increases, the reorder point at the DC also increases, as we would expect.

If the desired service at the DC is below the level at the Depot, inventories can be relatively low. But when the desired service level increases beyond the service level of the Depot, the inventory required increases quickly. When the service level of the DC is below that of the Depot, the inventory level can be close to $L_{DC} * \mu$, as represented in Figure 2a. But when $P_{DC} > P_{Depot}$, as shown in Figure 4, the inventory level must increase significantly, towards $(L_{DC} + L_{Depot})\mu$.

Figure 6. Reorder Point levels at DC, for varying levels P_{DC}, for P_{Depot} levels of 0.75, 0.90, and 0.95, DC LT is 3 days, Depot LT is 4 days, μ = 10, σ = 3.



As the desired service level is raised beyond that, inventory levels must increase, and more significantly for each percentage point of desired service increase. As seen in Figure 4, the tails of the probability distribution are very thin. As the desired service level increases closer to 1.00, the reorder point must go farther and farther out along the tail, requiring higher and higher levels of inventory, as seen in Figure 6.

With these results, it is now easy to see the importance of channel inventory location, and that my Fortune 30 colleagues' intuition was not correct. The service level at the Depot makes a dramatic difference on the service level. For example, looking at the reorder point = 60 horizontal line, if $P_{Depot} = 0.75$, the service level is 0.78. However, if $P_{Depot} = 0.95$, that same DC inventory provides the DC a service level of 0.98.

Another way to see this importance is to look a DC service level of 0.85. If the Depot provides a service level of 0.75, the DC needs a reorder point of nearly 70. If the Depot provides a service level of 0.90, the DC needs a reorder point below 40, a reduction of almost half. The amount and location of channel inventory strongly impacts customer service, and inventory decisions throughout the channel must be coordinated for maximum customer satisfaction.

SUMMARY OF THEORETICAL RESULTS

In this paper, in Result 1, we have determined the optimal reorder point for the Distribution Center, when the probability that the Depot may be out of stock is known. However, reorder level *R* cannot be solved for directly. In Result 3, we present an iterative algorithm that converges to the optimal value. Computational experiments have shown that it converges quickly.

SUMMARY OF PRACTICAL RESULTS

In today's omni-channel environment, the structure of distribution channels has been come more and more complex, with more varied structures. The service (or failure) at one level can have cascading effects on the performance of the facilities downstream from it.

Many supply chain professionals depend on reorder points based on a target service level. We consider the case where both the distribution center under consideration and the depot that feeds it use a target service level. When the depot is out of stock, the lead time to the DC increases dramatically, and suddenly. This is a situation likely experienced by the vast majority of practitioners.

Although the structure of the results in Result 3 are complex, they are simple to implement in a spreadsheet model for practitioners to perform what-if analysis. Previous results in the literature do not permit of such direct solution of safety stock calculations.

Additionally, because we assume the Normal distribution, our results easily can be understood and appreciated by practitioners.

FUTURE RESEARCH

In future research, we will expand the model to include the case where the stockout at the Depot does not always increase the lead time by the same number of days. To approximate the worst-case situation, we have assumed that when the Depot is out of stock, and has a lead time of L_{Depot}, the full lead time experienced by the Depot. However, it is also quite likely that the lead time remaining for the order at the Depot may be anywhere between 1 day and L_{Depot}.

We must consider the case where the lead time from the Depot may have a probability of P_{Depot} of being L_{DC}, but the lead time may be $L_{DC}+1, \ldots, L_{DC}+L_{Depot}$, each with a probability of $(1-P_{Depot})/\,L_{DC}$.

We must also consider the implications for larger numbers of DCs and Depots.

REFERENCES

Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal Inventory Policy. *Econometrica.* 19(3) 250-72.

Arrow, K. J., Karlin, S., & Scarf, H. E. (1958). *Studies in the Mathematical Theory of Inventory and Production*. Stanford: Stanford University Press.

Bagchi, U., & Hayya, J. C., (1984) Demand During Lead Time for Normal Unit Demand and Erlang Lead Time. *Journal of the Operational Research Society,* 35(2) 131-135.

Bagchi, U., Jack, C. H., & Chu, C.H. (1986). The Effect of Lead-Time Variability: The Case of Independent Demand. *Journal of Operations Management,* 6(2).

Bookbinder, J. H., & Çakanyildirim, M. (1999). Random lead times and expedited orders in (Q,r) inventory systems. *European Journal of Operational Research,* 115(2).

Chopra, S., Reinhart, G., & Dada, M. (2004). The Effect of Lead Time Uncertainty on Safety Stocks. *Decision Sciences,* 35(1), 1-24.

Cobb, B. R. (2013). Mixture distributions for modelling demand during lead time. *Journal of the Operational Research Society,* 64(2) 217-228.

Cobb, B. R., Johnson, A. W., Rumí, R., & Salmerón, A. (2015). Accurate lead time demand modeling and optimal inventory policies in continuous review systems. *International Journal of Production Economics,* 163, 124-136.

Croxton, K. L., Lambert, D. M., Garcia-Dastugue, S. J., & Rogers, D. S. (2008). The Demand Management Process. In *Supply Chain Management: Processes, Partnerships, and Performance* 3rd edition. Edited by Lambert, D. M. Sarasota FL: Supply Chain Management Institute.

Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1953). On the optimal character of the (s,S) policy in inventory theory. *Econometrica.* 21(4), 586-596.

Eppen, G. D., & Martin, R. K. (1988). Determining safety stock in the presence of stochastic lead time and demand. *Management Science,* 34(11) 1380-1390.

Hadley, G., Whitin, T. M. (1963). *Analysis of Inventory Systems.* Englewood Cliffs, N.J: Prentice Hall.

Jackson, P., Maxwell, W., & Muckstadt, J. (1985). The Joint Replenishment Problem with a Powers-of-Two Restriction. *IIE Transactions,* 17(1).

Lee, H., & Nahmias, S. (1993). Single-Product, Single Location Models. In *Handbooks in Operations Research and Management Science, Logistics of Production and Inventory,* edited by Graves, S. C., Rinnoy Kan, A. H. G., & Zipkin, P. H. Amsterdam: North-Holland.

Morse, P. M. (1958). *Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply*. New York: John Wiley & Sons.

Silver, E. A., Pyke, D. F., & Peterson, R. (1988). *Inventory Management and Production Planning and Scheduling.* New York: John Wiley & Sons.

Tyworth, J. E. (1992). Modeling Transportation-Inventory Trade-Offs in a Stochastic Setting. *Journal of Business Logsitics.* 13(2) 97-124.

Wagner, H. M. (1962). *Statistical Management of Inventory Systems.* New York: John Wiley & Sons.

Wang, P., Zinn, W., & Croxton. K.L. (2010). Sizing Inventory When Lead Time and Demand are Correlated. *Production and Operations Management,* 19(4), July/August.

Zipkin, P. (1986). Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly*, 33(4), 763-774.

DECISION SCIENCES INSTITUTE

How Does the Rule to Promote Campaigns by A Crowdfunding Platform Affect Target Setting and Sharing of Campaign's Revenues?

Joyaditya Laik
University of Pittsburgh
Email: jol76@pitt.edu

Esther Gal-Or
University of Pittsbugh
Email: esther@katz.pitt.edu

Prakash Mirchandani
University of Pittsburgh
Email: pmirchan@katz.pitt.edu

ABSTRACT

Considering an entrepreneur who relies on crowdfunding as the only source of funding, we analyse how the characteristics of a crowdfunding campaign and the platform's share of campaign revenues affect an entrepreneur's choice of funding target and pledge level. We find that when backers derive some altruistic benefits from participating in the campaign in addition to consumption benefits, the entrepreneur sets a funding target that exposes backers to the risk of production failure. We also demonstrate a potential misalignment between the platform's desired level of funding target and that chosen by the entrepreneur to maximize her profits. We investigate whether the platform can alleviate this misalignment by setting an appropriate campaign promotion rule and sharing rule of campaign revenues. Kickstarter, for instance, seems to promote campaigns that succeed in reaching a significant portion of their declared funding target soon after their launch. We compare this approach to a rule that identifies campaigns for promotion randomly. We find, surprisingly, that either rule leads to the same expected profits of the platform unless tying promotion to success in reaching a certain threshold of the declared target, attracts a larger number of new uninformed backers than random selection. We also investigate how the sharing of campaign revenues between the platform and entrepreneur depends on the characteristics of the campaign. The platform's share is lower when backers derive lower consumption or altruistic benefits, when there are fewer informed backers who can evaluate the characteristics of the product, when entrepreneurs incur significant reputational losses if unable to deliver the promised product, and when development cost of the product is high. In addition, if backers are altruistic, the revenue share set by the platform does not prevent participation of some entrepreneurs who are projected to never deliver the promised product.

KEYWORDS: Crowdfunding, Start-up Finance, New Product Introduction

DECISION SCIENCES INSTITUTE

Synthetic Average Neighborhood Sampling Algorithm (SANSA): A Neighborhood Informed Synthetic Sample Placement Approach to Improve Learning from Imbalanced Data

Murtaza Nasir
University of Massachusetts Lowell
Email: academic@murtaza.cc

Ali Dag
Creighton University
Email: alidag@creighton.edu

Serhat Simsek
Montclair State University
Email: simseks@montclair.edu

Asil Oztekin
University of Massachusetts Lowell
Email: asil_oztekin@uml.edu

ABSTRACT

Machine-learning classification models are increasingly being used in both real-world applications as well as in academic literature. However, many real-world phenomena happen much less often, and thus are more interesting and in many cases much more high-stakes to predict. In this work, we propose a new synthetic data generation algorithm that uses a novel "placement" parameter that can be tuned to adapt to each dataset's unique manifestation of the imbalance. SANSA also defines a novel modular framework to rank, generate and scale the new samples, which can be used in the future with other functions to propose better methods.

DECISION SCIENCES INSTITUTE
Current Status of Commercial Drone Operations and Predictions for Future

Nima Molavi
Department of Information, Operations and Technology Management,
College of Business and Innovation, University of Toledo, Toledo, Ohio, USA
Nima.Molavi@utoledo.edu

Marcelo J. Alvarado Vargas
Department of Management,
College of Business and Innovation, University of Toledo, Toledo, Ohio, USA
Marcelo.AlvaradoVargas@utoledo.edu

Yue Zhang
Department of Information, Operations and Technology Management,
College of Business and Innovation, University of Toledo, Toledo, Ohio, USA
Yue.Zhang@utoledo.edu

ABSTRACT

This paper aims to determine whether commercial drones have crossed the chasm of innovation and identify the influential factors in the rate of adoption. The Diffusion of Innovation theory and TOE framework are used to synthesize high-quality publications of top journals. It is found that the commercial drones industry are before the chasm. Despite considerable knowledge in technological context, organizational and environmental contexts are neglected. Social concerns, regulations, and the characteristics of firms that can impact the adoption of drones are the factors that should be studied thoroughly in future researches.

KEYWORDS: Commercial Drones, Technology Adoption, Diffusion of Innovation, TOE Framework, Advanced Technologies

INTRODUCTION

Drones are among technologies that are transforming the foundations of businesses (Cascio and Montealegre, 2016; McDonald and Eisenhardt, 2019; Pedraza-Martinez and Van Wassenhove, 2016). Substantial technological advancements of small drones made them a promising technological innovation that can disrupt many business models. Their potential applications run the gamut from monitoring and surveillance to package delivery in different domains such as construction, retailing, healthcare, agriculture, search and rescue, and transportation.

Despite numerous technological advancement, there are not much evidence about the real applications of drones in commercial domain. In this regard, understanding the current status of commercial drones is important. The lag in the exponential growth of drones points out the danger of falling in the chasm of innovation. So, it is important to understand the location of commercial drones on the diffusion curve. Based on the current phase of diffusion, the factors that can enhance the adoptions should be identified and understood in order to speed up the growth and achieve the apex of adoption.

The theory of "Diffusion of Innovation" (Moore, 1999) is used to determine the current position of commercial drones. The diffusion of innovation theory explains how innovations are spreading in a population and states different groups of adopters as well as their characteristics (Robinson, 2009). After identifying the current status of commercial drones using diffusion of innovation theory, Technological-Organizational-Environmental (TOE) framework (Tornatzky and Fleischer, 1990) is used to identify the factors that can influence the adoption of drones. Based on the current location and important adoption factors, some suggestions for promoting commercial drones' adoptions are proposed.

Tang and Veelenturf (2019) studied the potential strategic role of advanced logistics in industry 4.0 and proposed some interesting research questions for future researches: "How would drones and robots take over the home deliveries? Will there be drones and delivery robots everywhere? How to organize and regulate these unmanned delivery services without crating chaos?"

By reviewing the literature published in top academic journals, this paper aims to answer the first two questions about drones in a more general domain rather than delivery alone: (1) How would drones take over the operations, and (2) Will there be drones everywhere.

The rest of this paper is organized as follows: First, in theoretical background section, a theoretical framework that is used to investigate the commercial drones industry is described. Then, the methodology is defined and literature review phases are mentioned. Results are described in the next section as well as some results explanations. Lastly, the findings are discussed as well as limitations and some future research directions are proposed.

LITERATURE REVIEW

In this sections, first of all a brief description of drone industry is proposed. Afterwards, two fundamental theories of technology adoption are reviewed, namely diffusion of innovation and TOE framework. Finally, theoretical contributions of this paper are explained.

Drone Industry

In the past decade, the aviation industry has witnessed the rapid growth of new types of flying machines: Drones or UAVs (Unmanned Aerial Vehicles). Similar to many other technologies, drones have been born and developed from the military sectors. The idea of using autonomous flying vehicles is not new. It is first developed and used years ago for military purposes; however, recently, substantial technological advancement made it possible for manufacturing companies to produce small, lightweight drones for consumer use. These drones were initially used for hobby and fun as their users could capture beautiful footage from the sky or control them in a race. With the emergence of consumer drones, many large firms have realized the huge potential of drones and tried to utilize drones in their operations.

Nowadays, there are three types of drone industries: Military drones, Consumer drones, and Commercial drones. The first two industries are billion-dollar sectors that experienced exponential growth over the past years; however, the commercial drones have not been widely accepted and achieved exponential growth. Regarding commercial drones, there are two types of companies:

1. Users: companies that started to integrate drones within their operations
2. Providers: companies that provide drones solutions for other companies

The first group is usually large and famous companies who already started to experiment and implement drones within their operations. For instance, Amazon and DHL are started to deliver packages using drones in some regions. Publicly announcing the utilization of drones, these

companies have created hype around the commercial drones industry. The second stream of literature are mostly SMEs that are highly specialized firms, focusing on the application of drones and its technical features. Precision Hawk is a large company that focuses on drone solutions for businesses. They also benefit from the largest drone pilot network that can be hired by companies in order to operate their drones programs. Parallel Flight is another company that produces heavyweight hybrid drones that can carry up to 70 lbs. loads and fly up to 6.4 hours. These hybrid drones can deliver heavy packages safely, and also can fly and land in heavy winds. Another company that provides drones with various types of sensor and cameras is MicaSense. This company has focused on data collection using drones as well as data analytics to help firms generate desired information from collected data.
This paper focuses on commercial drones and tries to investigate whether commercial usage of drones will experience exponential growth and how the process can be accelerated. To conclude, this paper tries to answer three fundamental questions regarding commercial drones industry:

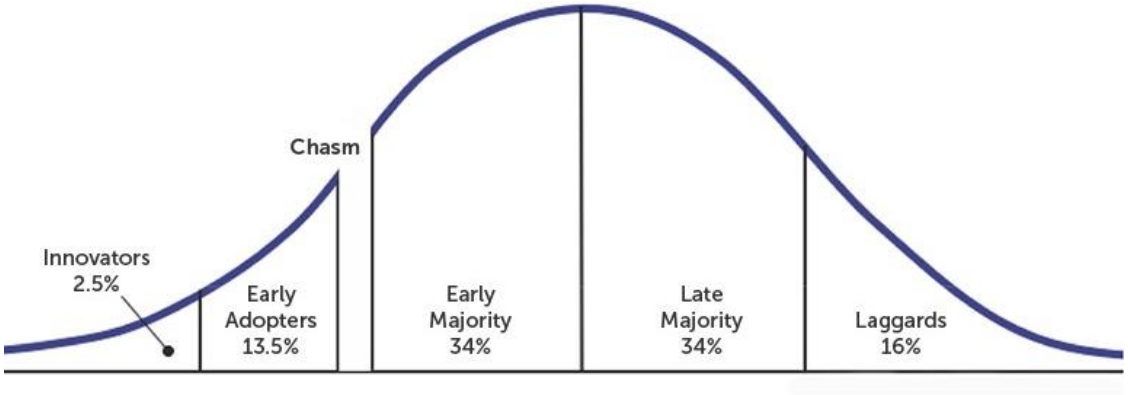
1. What is the current status of commercial drones?
2. What factors can speed up the adoption of commercial drones?
3. Will commercial drones achieve exponential adoption growth?

Diffusion of Innovation

In a seminal work, Rogers (1983) discussed the innovation characteristics that have high impacts on the diffusion of innovation. These characteristics are “Relative advantage, Compatibility with current practices, Complexity of use, Trialability, and Observable Results”. Since then, the Diffusion of Innovation (DOI) theory has applied to several contexts and the adoptions of many new technologies have been examined using the framework provided by this theory.
In order to diffuse and achieve their apex, innovations should evolve in order to match the needs of individuals or firms (Robinson 2009). Diffusion literature suggests that adopters can be categorized into 5 major groups: “Innovators, Early adopters, Early Majority, Late Majority, and Laggards”. Each group has a specific speed of adoption which is related to their propensity to adopt or resist. Innovation should evolve during its lifecycle to satisfy the requirements of each group. It would be valuable to first review the characteristics of each group of adopters:
Innovators – This groups mostly consists of firms who created something new or find a way to use an old thing in new ways. These firms are visionary and imaginative and their initial purposes are usually not financial incentives. These are the firms or individuals that introduce a new technology or new application of old technology to the world.
Early Adopters – Second groups of firms will start to use the technology as soon as the benefits of it became apparent. These firms are smart in identifying new technologies that can meet their needs and building opportunities from it. These are usually big companies that have sufficient money to invest in the new innovation and adopt it in their own operations.
Early Majority – After the success of early adopters, early majority firms will adopt the technology. These firms are pragmatists and need to observe the actual benefit of innovation in order to implement that in their operations. However, they are more open-minded about new ideas.
Late Majority – Late majority are completely conservative and highly pragmatist. These firms are highly risk-averse and not comfortable with progressive ideas. By observing the wide adoption of new innovation by previous groups, the late majority do not want to fall behind. Consequently, they adopt the innovation to not lose to the firms who already adopted the innovation.

Laggards – Laggards are firms that have a high level of resistance to change. They see high levels of risk in the adoption of new innovation and do not want to endanger their current status. These firms are the last group that will adopt the technology and some of them may never do. Each of these groups has a notional percentage, 2.5% innovators, 13.5% early adopters, 34% early majority, 34% late majority, and 16% laggards. If we can roughly determine the percentages of firms that are currently using a specific technological innovation, we can understand what is the current status of that field. Figure 1 illustrates the position of each group of adopters, the technology adoption curve and the chasm.

Figure 1- Technology Adoption Lifecycle (Moore 1999)

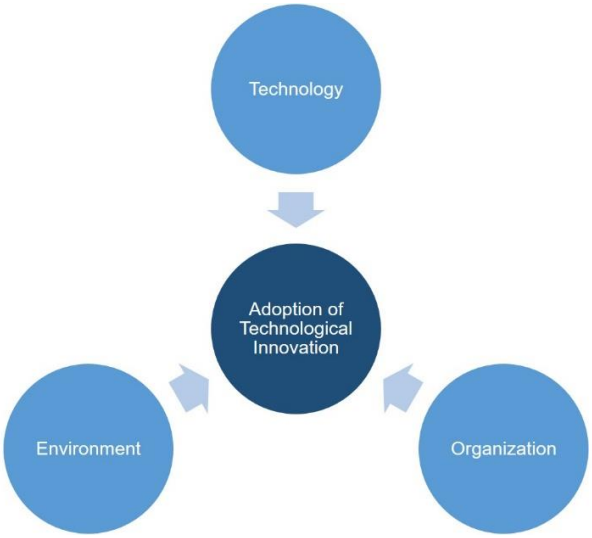


Moore (1999) proposed that there is a chasm in the diffusion of technological innovation. The chasm is somewhere between the early adopters and early majority phases. The location and width of the chasm depend on the technology itself and the context. After Moore's (1999) classic work, many research papers applied their theories into different technological contexts. For instance, Cho et al. (2008) investigated the adoption of telehealth innovation and stated that the chasm is crossed in this field. Elgort (2005) suggested several ways to bridge the chasm in e-learning adoption.
Following the concepts of diffusion of innovation theory by Rogers (1983) and theoretical view of Moore (1999) on how to cross the chasm, this paper tries to investigate the current status of drone applications and determine whether drones already crossed the chasm and experienced exponential growth or they need to face it in the future.

Technology-Organization-Environment Framework

Trying to pursue the second research question of this paper, factors that may help the adoption of commercial drones should be investigated. In order to identify such factors, Technology-Organization-Environment (TOE) framework is used. TOE framework is initially proposed by Tornatzky and Fleischer (1990). They suggested that there are three aspects of context that may influence the process by which a business adopts a technological innovation.
Technological context – Technological context consists of the characteristics of a specific technology that can enhance or hinder the adoption process. Many papers have mentioned "Technology readiness" or "Technology competence" as one of the most important factors in adoption.

Figure 2 - TOE Framework for Adoption of Technological Innovation (Tornatzky and Fleischer 1990)



For instance, Schaupp and Belanger (2014) suggested that technology readiness is an antecedent of social media adoption by firms, and Oliveira et al. (2014) argued the same in the adoption of cloud computing. Technology readiness simply means that to what extent the technology is mature and available for firms to adopt. In this paper, technology readiness can be considered as the maturity of drone technology itself and the maturity of its operations management systems.

Organizational context – Organizational factors are related to the structure and characteristics of the firm. These are mostly the resources or features that can support or push back the adoption of technology. Two of the most mentioned factors in organizational context are "Top management support" and "Firm size" (Low et al., 2011; Oliveira et al., 2014).

Environmental context – Environmental context refers to the external situations and conditions that can affect technology adoption. For instance government regulations (Oliveira et al., 2014), competitive pressure (Schaupp and Belanger, 2014; Low et al., 2011; Oliveira et al., 2014), and customers' expectations are all external factors that have a significant impact on the adoption of technological innovation. Currently, in the adoption of commercial drones, maybe the most boundary antecedents are related to environmental context. Having very tight regulations coupled with substantial concerns in society has created a severe obstacle for the adoption of commercial drones.

In this paper, the technological innovation is commercial drones. Based on TOE framework the adoption process of commercial drones will be affected by technological context, organizational context, and environmental context. Accordingly, the factors related to each context should be identified initially and then be studied in order to realize that whether the current status of TOE contexts can ameliorate the adoption of commercial drones or hinder its growth.

One of the biggest challenges for drone operations is social concerns and acceptance (De Boeck et al. 2019).

Theoretical Contributions

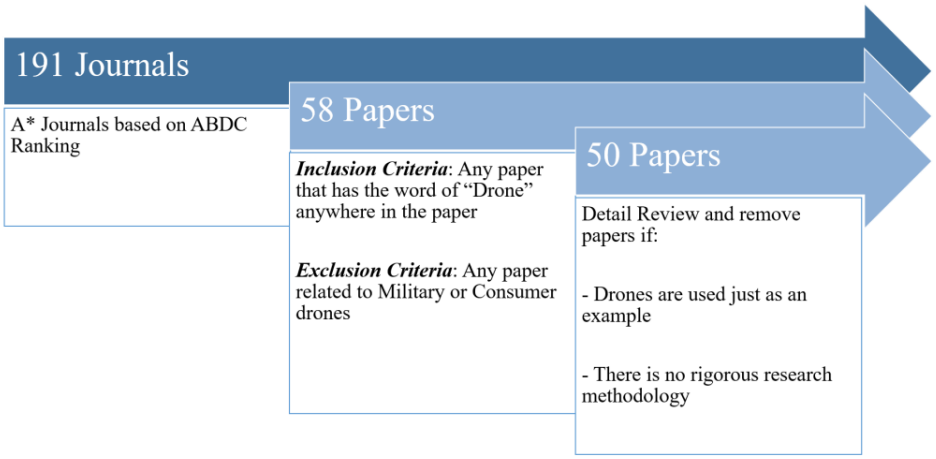
As mentioned earlier, this paper is focusing on the commercial drones industry. Relying on the diffusion of Innovation theory and TOE framework, this paper studies the current status of

commercial drones in the adoption process and identifies the important factors that can influence the future growth of drones' applications for commercial purposes. To sum it up, the theoretical contributions of this paper can be summarized as follows:
Considering the diffusion of innovation theory, this paper determines the current status of commercial drones' development and adoption
Applying TOE framework to commercial drones industry, this paper identifies factors that can affect the future growth of commercial drones' usage

METHODOLOGY

In order to understand the current status of commercial drones industry, this paper uses a comprehensive literature review considering published articles in top business journals. The presence of academic knowledge is one of the most important aspects of each industry and can be used to determine the extent of that industry's progress. In this regard, the literature review can help to understand the current status of academic knowledge about commercial drones. Choosing the scope of literature review, one should only consider researches with rigorous methodologies and avoid noisy or unscientific papers. One way to guarantee the quality of reviewed literature is to limit the literature review to quality journals. In order to achieve this goal, this paper uses the 2019 Australian Business Deans Council (ABDC) Journal Quality List, which is one of the most well-known rankings of business journals. Based on this list, business journals are ranked into four classifications: A*, A, B, and C.
This paper uses a comprehensive search through A* journals in the list. In total, 191 journals have been reviewed. In the first phase, the keyword "drone" is used to search and identify the related paper. The word "drone" could be in any part of the papers. Since this paper only focuses on commercial drones industry any paper related to consumer drones or military drones is excluded. Additionally, all papers related to engineering and technical knowledge of drones are excluded. 58 papers were found in the first phase.
In the second step, the topic of the papers was reviewed in order to realize whether drone is the main topic of the paper or it is only used as an example or illustration. Any paper that has not had drone as the core idea or have not utilized a rigorous research method were eliminated from the sample. 8 papers were eliminated and 50 papers remained at the end of this step.

Figure 3- Literature review framework and procedure



Finally, in the last step, all of the remaining papers are reviewed in detail in order to be classified based on context, domain, application, and research methodology. Based on the TOE framework, three contexts are considered, namely, Technological, Organizational, and Environmental. Domain, application, and research methodology categories are also determined based on the collected papers and will be discussed in the results section. Figure 3 illustrates the literature review framework and procedure.

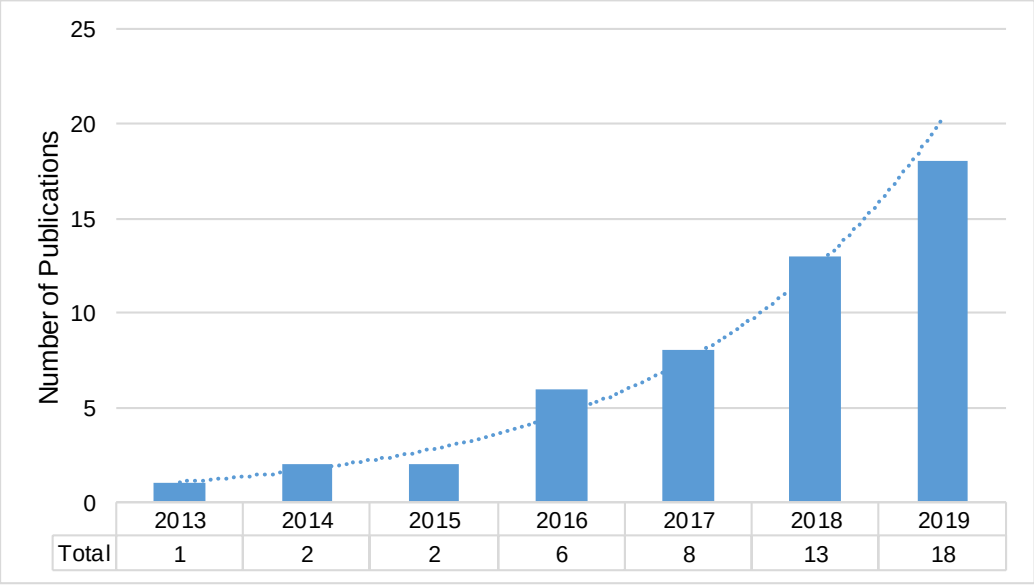
RESULTS

The results of literature review are discussed in two separate parts. In the first part the number of publications in each year, number of publications in each journal, and the research methodologies are examined, in order to determine the current status of literature in commercial drone area. In the second part, by reviewing each paper, the studied context, investigated drone application, and the fields in which the drone application is applied are indicated, in order to find the gap in knowledge accumulation and suggest some remedies to fill the gap.

Current Status

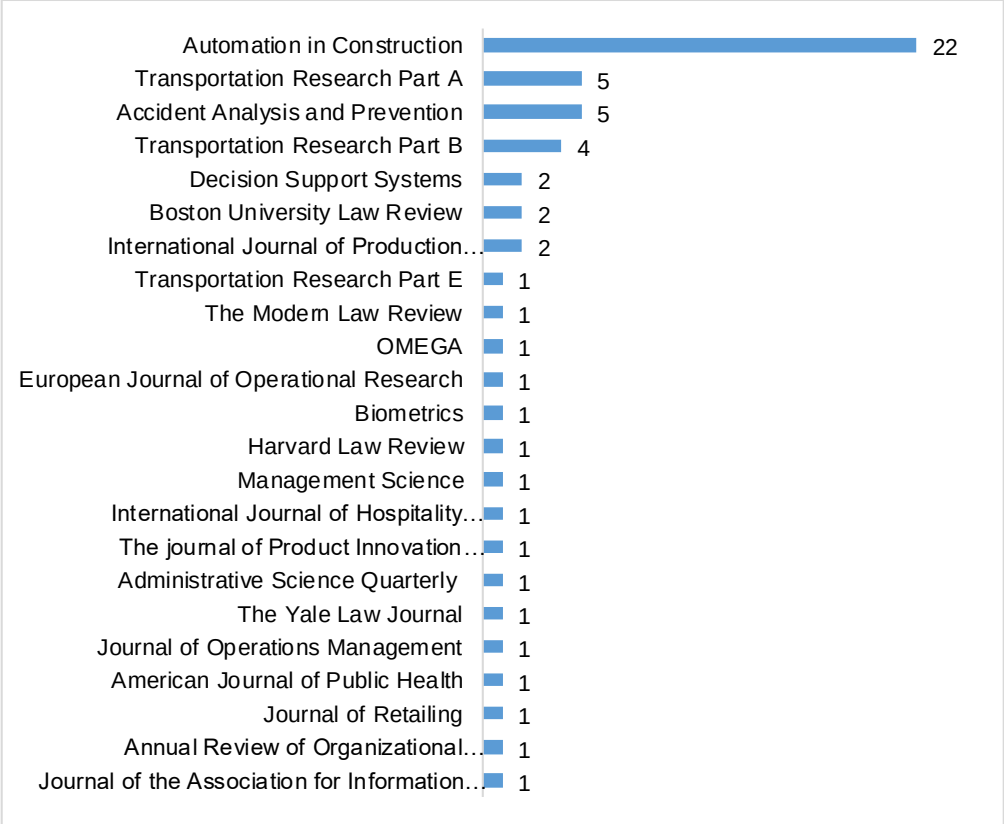
In order to examine the current status of drone-related knowledge, it would be useful to analyze the number of publications each year. Drone-related articles have started to being published in "A*" journals, from 2013. The first paper is written by Roca et al. (2013) suggesting the application of drones to inspect building facades. This paper is published in "Automation in Construction" journal. Afterward, in 2014 and 2015 only four papers were published in aggregate. Starting in 2016, drone-related publications got momentum. In 2016, six publications were published and followed by eight papers in 2017. In 2018 the number increased to thirteen. Finally, in 2019, 18 papers were published related to commercial drones. This study is conducted in December 2019, hence it is unlikely that any paper published in 2019 is missed in. Figure 4 shows the number of publications in each year.

Figure 4- Number of Publications in each year



As it can be observed in figure 4, it seems that publications have been accelerated since 2016. The number of publications can be best estimated by an exponential function and the R-square will be 0.9671. This simply indicates that the number of commercial drones' publications has experienced an exponential growth and we expect to observe more publications in 2020.
The next discussion is related to the number of publications in each journal. Figure 5 shows the number of publications in each journal. During the reviewed period (2013-2019), only 24 out of 191 journals have published papers related to drones. "Automation in Construction" journal has published 22 papers which is the highest number of publications in "A*" journals. This journal is related to constructions and civil engineering and examines how the operations can be automated in this domain. Publishing 5 papers each, "Transportation Research Part A" and "Accident Analysis and Prevention" journals are ranked as second journals with highest number of publications. "Transportation Research Part B" has also published 4 papers and ranked next. All of these three journals are related to logistics and transportation.

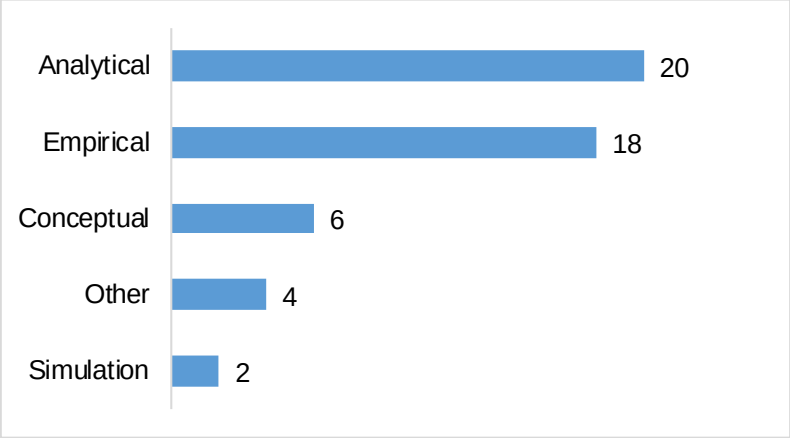
Figure 5- Number of Drone Publications in each Journal



The interesting point is the presence of "Boston University Law Review", "The Modern Law Review", "Harvard Law Review", and "The Yale Law Journal". In aggregate, these law journals have published 5 paper related to drones which indicate that at least there are some attentions to drones regulations and laws. Although there are many more papers in law journals related to drones, most of them have focused on military drone sections and did not pay enough attention to commercial drone sector and its regulations.
The last discussion in this section is related to the research methodologies that have used in this paper. As it is shown in figure 6, analytical methods are used most, as 20 out of 50 papers (40 percent) of the papers have used analytical methods. Empirical methods are very close to

analytical methods and have been used in 18 papers. Finally, 6 papers were conceptual and 2 papers have used simulation, while 4 papers have pursued other research methodologies that not be classified under current paper categories.

Figure 6- Number of publications using each methodology

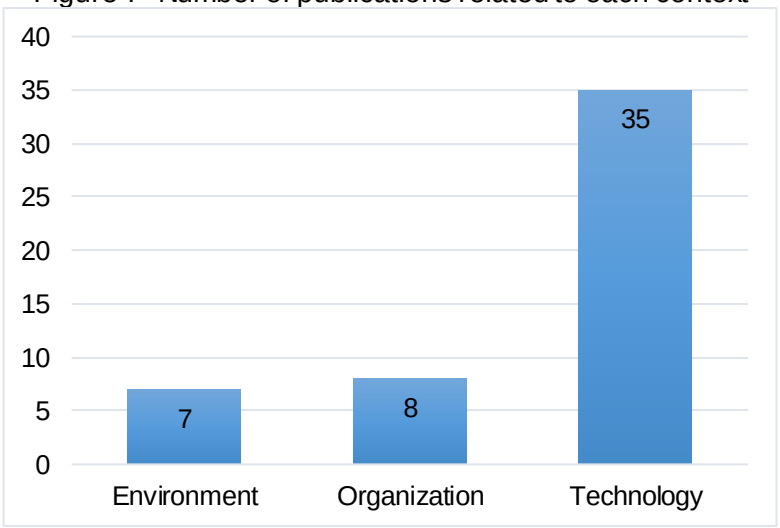


Prediction for Future

As it is mentioned, in order to predict the future of commercial drones' adoption, it is fruitful to investigate the level of knowledge accumulation regarding each context of the TOE framework. 50 collected papers were reviewed and categorized under one context. Figure 7 shows the number of commercial drones' publications related to environmental, organizational, and technological contexts.

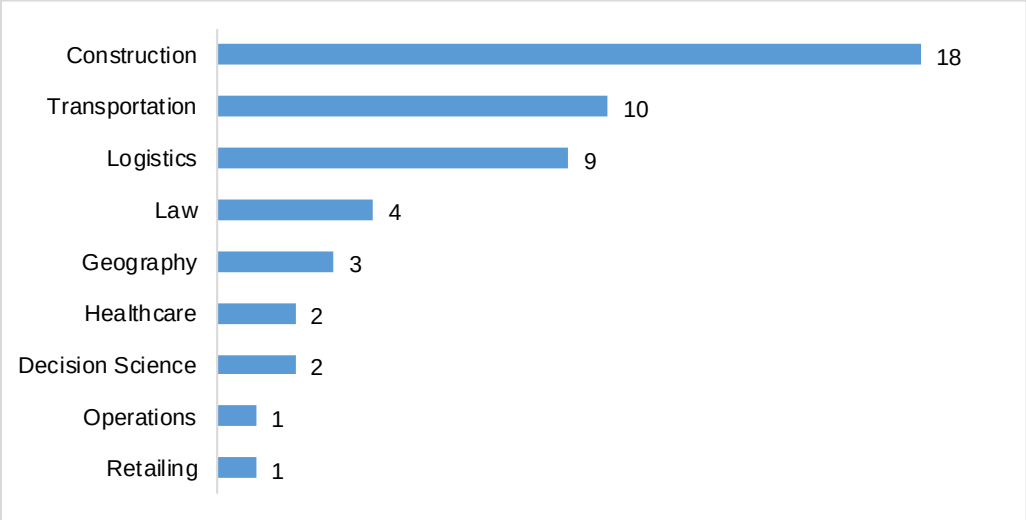
Most of the papers (35 out of 50) are related to the technological context. This category of papers has consisted of two streams (1) the technical features of drones and how to operate them, and (2) scheduling and optimization of drones routes or station locations. The first stream is more related to the engineering perspective and the second stream is more related to operations research. One of the most frequently used models in the second stream is the traveling salesperson problem.

Figure 7- Number of publications related to each context



In contrast, two other contexts of the TOE framework have attracted much less attention. Only 8 papers were related to the organizational context and 7 papers were related to environmental context. To achieve the high rate of commercial drones' adoption, knowledge should be generated and accumulated in all three contexts of the TOE framework. Based on the current results, it can be concluded that the organizational and environmental contexts are neglected; however, it seems that the rate of publication is getting momentum and we will see more papers on environmental and organizational contexts.
The organizational context is more related to the benefits of drones for firms and the characteristics of firms that can mediate the adoption decision or adoption speed. The environmental context of the drones should focus on social concerns and regulations as these two are the two most important environmental factors that can impact the adoption. Social concerns are not investigated. It is good to understand the remedies for social obstacles.
It also would be useful to examine the number of publications in each scientific field. In this regard, papers were categorized under fields to which it is mostly related. The results of this classification can be observed in figure 8.

Figure 8- Number of publications on each field



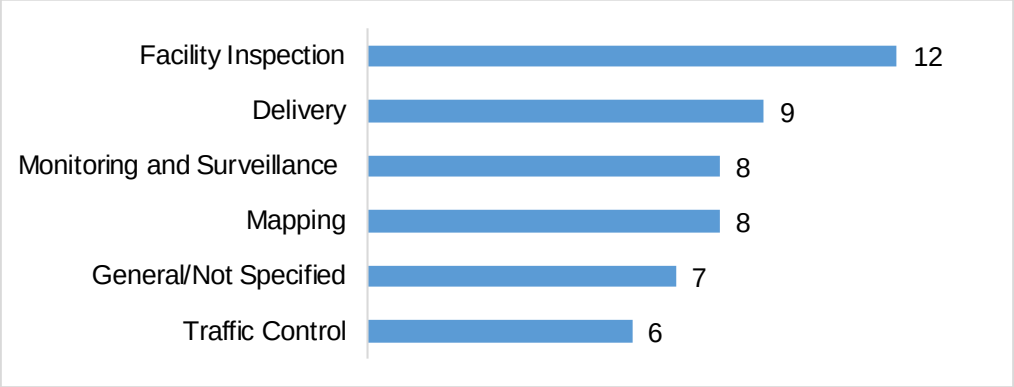
As can be observed, Construction field has 18 papers which is the highest number of publications. It is not surprising since the "Automation in Construction" journal had the highest number of publications related to commercial drones. Additionally, Cassetta et al. (2017) and Tang and Veelenturf (2019) stated that drones will be the new technological innovation in transportation and logistics. As it could be observed in Figure 8, transportation and logistics are ranked second and third in terms of the number of commercial drones' publications, and in aggregate, transportation and logistics has consisted of 19 papers, which is even higher than construction field. Accordingly, it can be predicted that drones will have a huge impact on transportation and logistics fields in the future.
Finally, the last aspect of reviewed paper is the applications of commercial drones. In other words, what is the purpose of commercial drones' adoption? As one can imagine, drones can be used for many different purposes. Based on reviewed literature, the application of drones can be categorized into three groups:

1. Data generation and collection
2. Delivery

3. Communication

Data generation and collection is the most investigated application of drones, followed by delivery application. The communication application of drones has not attracted much attention yet. To be more specific, reviewed papers were assigned to 6 different categories: Facility inspection, Delivery, Monitoring and surveillance, Mapping, Traffic control, and the papers that did not consider any specific application. Figure 9 shows the application of drones based on mentioned classification.

Figure 9 - Number of publications related to each application



The results are consistent with the Agnisarman et al. (2019) that mentioned drones will be one of the most important tools of infrastructure visual inspection. Moreover, the high number of publications studying delivery application of drones is consistent with the hype created by large companies like Amazon and DHL.

DISCUSSION AND CONCLUSION

Evidence suggests that drones will be a significant technology for future operations. In light of this fact, this paper tries to investigate the current status of commercial drones from the academia perspective and suggests some practices to enhance the rate of adoption. This paper contributes to the technology adoption literature by applying the diffusion of Innovation theory and TOE framework. Additionally, this paper will contribute to the growing literature in commercial drones by summarizing the previous works using different classifications.
Based on the literature review results, the level of knowledge for technological context of commercial drones are satisfactorily developed and still increasing. So, it seems that there will not be any barrier related to drone technology itself. In contrast, the organizational and environmental contexts of commercial drones are less investigated. The reason for this unbalanced development of commercial drones industry is that drone technology is not a new innovation, but its application in commercial domain is a new phenomenon. Hence, its technological context is well developed as it is already used in military and consumer domain; however, commercial domain requires different types of organizational and environmental considerations. In this regard, more researches should focus on the organizational and environmental contexts of drones.
In the organizational context, researchers should focus more on the economic benefits of commercial drones for firms, characteristics of firms that can enhance or hinder the adoption. More importantly, environmental factors should be investigated that can act as a tough barrier for commercial drones adoption. Two environmental factors that are highly important are: (1) Social perceptions, and (2) Regulations. If people think that drones can be dangerous or

useless, then there will be a serious pressure not to adopt the commercial drones. Consequently, social concerns should be address and the practices that can increase the awareness of society about the benefits of commercial drones should be investigated. On the other hand, regulations about commercial drones are so severe in the United States. The commercial drones operations are so limited due to these regulations and companies perform under drastic restrictions. So, it seems necessary to study the regulations and their impact on commercial drones' adoption.

Finally, it is highly important to mention the limitations of this paper. This paper has three major limitation. First, reviewing literature, this paper investigates the current status of commercial drones from the academia perspective; however, it is highly important to examine the practitioner side as well. In this paper no data from the industries and practitioners are analyzed. Therefore, one fruitful future research can examine practitioners' perspectives and determine the current status of each technological, organization, and environmental context.

Second, this paper only considers "A*" journals based on the ABDC ranking. Although reviewing top journals is highly crucial as it can show the current status of best researches in academia, it would be more comprehensive to review at least "A" journals to gather a more complete picture of the academia side. So, future researches can extend this study by studying a wider scope of literature. It would be highly interesting to see whether the publications of other journals also follow the same pattern or not.

To put in a nutshell, drones will be part of our future. So, it is better to fully understand their characteristics, their limitation, and their adoption processes in order to make a smooth transition to the drone era.

REFERENCES

Agnisarman, S., Lopes, S., Madathil, K. C., Piratla, K., & Gramopadhye, A. (2019). A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection. Automation in Construction, 97, 52-76.

Babel, L. (2017). Curvature-constrained traveling salesman tours for aerial surveillance in scenarios with obstacles. European Journal of Operational Research, 262(1), 335-346.

Bang, S., Kim, H., & Kim, H. (2017). UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching. Automation in Construction, 84, 70-80.

Carlsson, J. G., & Song, S. (2017). Coordinated logistics with a truck and a drone. Management Science, 64(9), 4052-4069.

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. Annual Review of Organizational Psychology and Organizational Behavior, 3, 349-375.

Cassetta, E., Marra, A., Pozzi, C., & Antonelli, P. (2017). Emerging technological trajectories and new mobility solutions. A large-scale investigation on transport-related innovative start-ups and implications for policy. Transportation Research Part A: Policy and Practice, 106, 1-11.

Chen, P., Zeng, W., & Yu, G. (2019). Assessing right-turning vehicle-pedestrian conflicts at intersections using an integrated microscopic simulation model. Accident Analysis & Prevention, 129, 211-224.

Cho, S., Mathiassen, L., & Gallivan, M. (2008, October). Crossing the chasm: From adoption to diffusion of a telehealth innovation. In IFIP Working Conference on Open IT-Based Innovation: Moving Towards Cooperative IT Transfer and Knowledge Diffusion (pp. 361-378). Springer, Boston, MA.

Chowdhury, S., Emelogu, A., Marufuzzaman, M., Nurre, S. G., & Bian, L. (2017). Drones for disaster response and relief operations: A continuous approximation model. International Journal of Production Economics, 188, 167-184.

De Boeck, K., Decouttere, C., & Vandaele, N. (2019). Vaccine distribution chains in low-and middle-income countries: A literature review. Omega.

Dodgson, M., Gann, D., MacAulay, S., & Davies, A. (2015). Innovation strategy in new transportation systems: The case of Crossrail. Transportation Research Part A: Policy and Practice, 77, 261-275.

Dulebenets, M. A., Ozguven, E. E., & Moses, R. (2018). The highway beautification act: towards improving efficiency of the federal outdoor advertising control program. Transportation research part A: policy and practice, 110, 88-106.

Elgort, I. (2005, December). E-learning adoption: Bridging the chasm. Ascilite, 155-165.

Ellenberg, A., Kontsos, A., Moon, F., & Bartoli, I. (2016). Bridge Deck delamination identification from unmanned aerial vehicle infrared imagery. Automation in Construction, 72, 155-165.

Freimuth, H., & König, M. (2018). Planning and executing construction inspections with unmanned aerial vehicles. Automation in Construction, 96, 540-553.

Furini, F., Persiani, C. A., & Toth, P. (2016). The time dependent traveling salesman planning problem in controlled airspace. Transportation Research Part B: Methodological, 90, 38-55.

Goessens, S., Mueller, C., & Latteur, P. (2018). Feasibility study for drone-based masonry construction of real-scale structures. Automation in Construction, 94, 458-480.

Gray, D. (2016). Fourth Amendment Remedies as Rights: The Warrant Requirement. BUL Rev., 96, 425.

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The future of retailing. Journal of Retailing, 93(1), 1-6.

Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., & Yuan, J. (2019). Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. Accident Analysis & Prevention, 123, 159-169.

Ham, Y., & Kamari, M. (2019). Automated content-based filtering for enhanced vision-based documentation in construction toward exploiting big visual data from drones. Automation in Construction, 105, 102831.

Hwang, J., Lee, J. S., & Kim, H. (2019). Perceived innovativeness of drone food delivery services and its impacts on attitude and behavioral intentions: The moderating role of gender and age. International Journal of Hospitality Management, 81, 94-103.

Inzerillo, L., Di Mino, G., & Roberts, R. (2018). Image-based 3D reconstruction using traditional and UAV datasets for analysis of road pavement distress. Automation in Construction, 96, 457-469.

Jeong, H. Y., Song, B. D., & Lee, S. (2019). Truck-drone hybrid delivery routing: Payload-energy dependency and No-Fly zones. International Journal of Production Economics, 214, 220-233.

Keenan, P. B., & Jankowski, P. (2019). Spatial decision support systems: three decades on. Decision Support Systems, 116, 64-76.

Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Automation in Construction, 99, 168-182.

Kim, P., Park, J., Cho, Y. K., & Kang, J. (2019). UAV-assisted autonomous mobile robot navigation for as-is 3D data collection and registration in cluttered environments. Automation in Construction, 106, 102918.

Li, F., Zlatanova, S., Koopman, M., Bai, X., & Diakité, A. (2018). Universal path planning for an indoor drone. Automation in Construction, 95, 275-283.

Low, C., Chen, Y., & Wu, M. (2011). Understanding the determinants of cloud computing adoption. Industrial management & data systems, 111(7), 1006-1023.

Masutti, A., & Tomasello, F. (Eds.). (2018). International Regulation of Non-military Drones. Edward Elgar Publishing.

Moore, G. A. (1999). Crossing the chasm. Harper Business.

Morgenthal, G., Hallermann, N., Kersten, J., Taraben, J., Debus, P., Helmrich, M., & Rodehorst, V. (2019). Framework for automated UAS-based structural condition assessment of bridges. Automation in Construction, 97, 77-95.

Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. Information & Management, 51(5), 497-510.

Omar, T., & Nehdi, M. L. (2017). Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography. Automation in Construction, 83, 360-371.

Oruc, B. E., & Kara, B. Y. (2018). Post-disaster assessment routing problem. Transportation research part B: methodological, 116, 76-102.

Park, J., Kim, P., Cho, Y. K., & Kang, J. (2019). Framework for automated registration of UAV and UGV point clouds using local features in images. Automation in Construction, 98, 175-182.

Park, H., Oh, C., Moon, J., & Kim, S. (2018). Development of a lane change risk index using vehicle trajectory data. Accident Analysis & Prevention, 110, 1-8.

Pick, J. B., Turetken, O., Deokar, A., & Sarkar, A. (2017). Location Analytics and Decision Support: Reflections on Recent Avancementa, a Research Framework and the Path Ahead. Decision Support Systems, 99.

Rakha, T., & Gorodetsky, A. (2018). Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones. Automation in Construction, 93, 252-264.

Richard, A., Bempong, NE, Flahault, A. (2019). Geneva Health Forum: The First international conference on precision global health International Conference on Precision Global Health. American journal of public health, 109(6), 863-865.

Robinson, L. (2009). A summary of diffusion of innovations. Enabling change, 5(10).

Roca, D., Lagüela, S., Díaz-Vilariño, L., Armesto, J., & Arias, P. (2013). Low-cost aerial unit for outdoor inspection of building façades. Automation in Construction, 36, 128-135.

Rodriguez-Gonzalvez, P., Gonzalez-Aguilera, D., Lopez-Jimenez, G., & Picon-Cabrera, I. (2014). Image-based modeling of built environment from an unmanned aerial system. Automation in Construction, 48, 44-52.

Rosenfeld, A. (2019). Are drivers ready for traffic enforcement drones?. Accident Analysis & Prevention, 122, 199-206.

Rule, T. A. (2015). Airspace in an Age of Drones. BUL Rev., 95, 155.

Schaupp, L. C., & Bélanger, F. (2013). The value of social media for small businesses. Journal of Information Systems, 28(1), 187-207.

Seo, J., Duque, L., & Wacker, J. (2018). Drone-enabled bridge inspection methodology and application. Automation in Construction, 94, 112-126.

Siebert, S., & Teizer, J. (2014). Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. Automation in construction, 41, 1-14.

Singer v. City of Newton No. 17-CV-10071 (D. Mass. Sept. 21, 2017); Harvard Law Review. 2018, Vol. 131, Issue 7, pages. 8; Harvard Law Review Association, 2018.

Stevenson, B. C., Borchers, D. L., & Fewster, R. M. (2019). Cluster capture-recapture to account for identification uncertainty on aerial surveys of animal populations. Biometrics, 75(1), 326-336.

Tang, C. S., & Veelenturf, L. P. (2019). The Strategic Role of Logistics in the Industry 4.0 Era. Forthcoming in: Transportation Research Part E: Logistics and Transportation Review.

Tornatzky, L. G., & Fleischer M. (1990). The Process of Technological Innovation. Lexington, MA: Lexington Books.

Wang, L., Chen, F., & Yin, H. (2016). Detecting and tracking vehicles in traffic by unmanned aerial vehicles. Automation in construction, 72, 294-308.

Wang, Z., Sheu, JB. (2019). Vehicle routing problem with drones. Transportation research part b-methodological, 122, 350-364.

Wang, C., Xu, C., & Dai, Y. (2019). A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. Accident Analysis & Prevention, 123, 365-373.

Xia, J., Wang, K., & Wang, S. (2019). Drone scheduling to monitor vessels in emission control areas. Transportation Research Part B: Methodological, 119, 174-196.

Zakeri, H., Nejad, F. M., & Fahimifar, A. (2016). Rahbin: A quadcopter unmanned aerial vehicle based on a systematic image processing approach toward an automated asphalt pavement inspection. Automation in Construction, 72, 211-235.

Zhong, X., Peng, X., Yan, S., Shen, M., & Zhai, Y. (2018). Assessment of the feasibility of detecting concrete cracks in images acquired by unmanned aerial vehicles. Automation in Construction, 89, 49-57.

DECISION SCIENCES INSTITUTE

Intentional Structural Homogeneity (ISH): Improving post-merger integration by reducing impact on HR processes and practices

Vivek Jirnal
XLRI School of Management
Email: fh14002@astra.xlri.ac.in

ABSTRACT

Structural inferences in mergers and acquisitions have been an underutilized dimension in understanding the success of integration. Previous studies have focused predominantly only on post-merger integration. With negligible emphasis on pre-merger strategic processes. The result is an unendurable experience for the human resources of firms involved in the event. A theoretical assessment is initiated in synthesizing a possible solution for reducing the zero or negative sum contributing activities in M&A's. Intentional Structural Homogeneity is discussed as a possible solution for reducing the disruptions to structure of the firms and therefore reducing the impact caused on the HR processes and practices.

KEYWORDS: Intentional Structural Homogeneity, Isomorphism, Absorptive Capacity, Post merger integration, Organizational inertia

INTRODUCTION

Strategic management's overall objective is to understand and create the conditions under which superior performance can be garnered and derived (Barney, 2002). Further, Mintzberg (1987) states that strategy can be a plan. A dominant utilization of strategy as a plan by firms is to create conditions for superior performance. One such category of strategy as a plan to create such conditions are Mergers and acquisitions (M&A's). As they are at the apex of strategic choice when it comes to inorganic growth (Schweizer, 2005). M&A's have been considered as a superior method of utilizing corporate resources for growth (Hitt et al, 1998). Although Hitt et al (1991) did not consider M&A's to be strategy in themselves, they considered them as vehicles. Vehicles which are utilized by management to realize its objectives and outcomes in a broader context (Jarillo, 2003). Despite the inherent goal of achieving performance improvement, the results of M&A's have often been disappointing and there is a considerable body of evidence that performance of firms formed as a result of Mergers and Acquisitions (M&A's) consistently fail to live up to these expectations (Rees and Edwards, 2009). The impact of such failures of M&A's is felt most intensely by the human resources of both the firms. This paper intends to discuss a possible way of challenging the conventional practice followed during M&A's. In the form of Intentional structural homogeneity (ISH). The firms under this method target organizations which are similar to their model in terms of structure. Ideally, targeting isomorphs as explained and understood through the contributions on this topic by Dimaggio and Powell (1983).

Background

Hitt et al (1998) attribute the failure of M&A's to the lack of understanding around this strategy as a plan in great theoretical and practical depths. In spite of a significant amount of research on M&A's there is very little or no research done to understand how M&A's can be successful (*ibid*). Davis and Stout (1992), have also concluded similarly in regards to organizational theories. Stating that, organizational theories have not been useful in explaining profoundly the success or behavior of M&A's. Studies on M&A's from Dewings (1921) to the contributions by Schweizer (2005), Puranam, Singh, & Zollo (2006), Vaara & Monin (2010), and finally Monin et. al. (2013) are in support of this theoretical inability. These

studies have highlighted the unpredictability and ambiguity involved with respect to the end results of M&A's. The mixed results are indicative of the fact that M&A's success is still an unsolved puzzle (King, et al. (2003:2004); Stahl, and Voigt (2008).

However, attempts have been made to understand the reasons for its success or failure and some studies have been able to expound some explanation. These studies have mostly been involving the unidimensional assessments of M&A's. To highlight some significant contributions such as Trautwein (1990), focussed on the motives or objectives behind the acquisitions. Sirower, (2000) used stock return measures to conclude that 65% of acquisitions fail to benefit the acquiring company. Child, Faulkner, and Pitkethly (2001) based their study on national culture and management style to state that, as many as 50% of acquisition's fail due to these two reasons. Furthermore, Puranam et al (2006), studied the structural integration alone and Shaver, (2006) discussed the synergy realization aspects of M&A's. Similarly, Weber and Tarba (2010), studied the post-acquisition integration. On the other hand, Monin et al (2013), studied the significance of justice in post-merger integration to explain the "Rules of the Game". Finally, a Meta analytical study which went beyond unidimensional approach to a more comprehensive nature, was from Straub et al (2013). They provided a three-prong approach to study M&A's success and failure, which involved the study of variables of strategic logic, organizational behavior, and financial perspective to predict the M&A's performance.

Research which has been carried out so far about M&A's can be broadly categorized as falling under three broad categories of resources-based view, content-based approach, and the process-based approach. However, though these studies have been pivotal in contributing to the existing literature on M&A performance. They suffer from some flaws. Firstly, these studies have all have been done with an emphasis on the post-merger scenario of M&A's only and not the pre-merger analysis. Secondly, they all have been a content-based approach to demystify the M&A's performance with a few exceptions, which focussed on the processual analysis but with only from the view of strategy. Thirdly, the structural integration phenomena have been studied only from the organizational behavior perspective and not around the implications of other structural characteristics. This paper thus aims to fill this gap in literature and create a continuum between the pre-merger and post-merger process. By focusing on the process as well as the content to unpacking the M&A's performance behavior. The paper makes a theoretically abductive attempt to identify a new dimension in reducing this impact on the human resources through appropriate planning in pre-merger stages. Through creating an understanding around how the structure of acquiring and merging firms impact the processes and practices of a firm, especially the HR practices and processes.

LITERATURE REVIEW

Factors impacting M&A's performance

Larsson and Fickelstein (1999), have stated that the post-merger integration is the single most important predictor of synergy realization in M&A's. Similarly, many researchers have added to this literature on integration (Birkinshaw, et al, 2000; Heimeriks, Schijven, and Gates, 2012; Pablo, 1998). However, the integration literature has also been criticized for not providing adequate theoretical frameworks for linking the explanatory variables to acquisition performance (Datta and Grant, 1990; Hitt et al, 1998; Hoskisson, Hitt, Johnson, and Moesel, 1993).

Pablo (1998), defines integration as the making of changes in the functional activity arrangements, organizational structures and systems, and cultures of combining organizations to facilitate their consolidation into a functioning whole. Later Puranam and Singh (2006) have discussed the post-acquisition integration in the light of structural integration alone. And discussed how organizations with technology orientation could allow initial autonomy to the target firms before "Grafting" them. However, there is an inherent cause for conflict in this phenomenon, which is due to the existence of organizational inertia.

Hannan and Freeman, (1984) & Heimeriks, Schijven, and Gates (2012) place a classic argument around post-merger integration based on organizational inertia and the role of structure leading to structural inertia. The authors argue that the larger established organizations develop organizational inertia over a period of time and hence, such organizations are not adaptable to changes in any of the above discussed variables. This is also predominantly because of the embedding nature of practices and processes which have provided positive results over a period of time into the structure of the organizations. Therefore, whether a firm is an acquirer or target in an M&A, higher inertia will prevent complete structural integration of the two firms thereby preventing realization of synergies.

Therefore, the crux of this discussion was to highlight that post-merger integration is dependent on the extent to which the practices, procedures and processes embedded in the structure of the firms are compatible to integration without causing a disruption (Puranam, and Singh, 2006, 2008; Cording et al, 2008; Nahavandi and Malekzadeh, 1988). Further analysis of this structure argument leads to revelation of the human component as playing the key role in prevention of integration. And subsequently causing hurdles in realization of synergies in M&A's. Thus, the study of human aspect becomes crucial for understanding M&A's performance.

Human side of M&A's

The mergers and acquisitions are very uncertain events, characterized by bounded rationality (Straub et al, 2013). Cartwright and Shoenberg (2006), have stated that this uncertainty itself has a dysfunctional effect on the organizational members, leading to most of the M&A's failing. Similarly, Hitt et al (1998), studied 12 pairs of M&A's to conclude that a careful pre-merger study about the target firm was a common characteristic of successful M&A's. Whereas inadequate evaluation of the target firm leads to failure in most cases. Suggesting similarly are Haspeslagh and Jemisons (1991a) who posit the importance of quality pre acquisition decision process. Boxall and Purcell (2003), discussed this failure only at the implementation stage, but Hitt et al. (1998), discussed failure at the negotiation and purchase stage. And discussed further, that failure during the implementation period is due to creation of several hurdles attributed to limited or incomplete pre-merger analysis. Which according to Straub et al. (2013) further leads to failure in attaining strategic synergy sought after from M&A's. This incomplete evaluation of the target firm is also the basis for problems to emanate. Particularly from HR and overall, from organizational culture of merging firms (Rees and Edwards, 2009).

Therefore, Nahavandi and Malekzadeh (1988), emphasize that during M&A's most organizations should not limit their attention to the strategic and the financial perspectives but extend it to the human side of M&A's. They discuss the prospects of higher success in related acquisitions than unrelated acquisition. This they believe is due to the degree of knowledge about the target firm existing among the decision makers (employees) of the acquiring firm. Furthermore, as already discussed the human element in M&A's is of great significance, since every other factor can be integrated over a shorter period of time to overcome the organizational inertia (Hannan, Pollos and Carroll, 2002; Helfat et al., 2007). It is also very well established that majority of the M&A's fail due to the improper attention provided to the way employment and HR issues are handled (Bouno and Bowditch, 1989; Cartwright and Cooper, 1996; Marks and Mirvis, 2001). In sum, these contributions emphasize that most M&A's fail due to the human element, mostly at the implementation stage when the two companies come together. Which is a result of the lack of a comprehensive strategic planning in the pre-merger stages. Particularly around the implications of the human side of M&A's. So, in order to minimize or eradicate hurdles emanating from the human element, let's take the above argument forward and in a different but related tangent.

Role of Structure in M&A's

The human resources of every organization are not freely moving, autonomous beings with absolute access to resources, but are restricted, controlled and characterized by the structure of the firms. Organization structure may be metaphorically considered as the anatomy of the organization, providing a foundation within which the organization functions (Dalton et al, 1980). Further, they also add that organization structure affects the behavior of members (ibid). The study conducted by Dalton et al (1980), is of a comprehensive nature and helps in formulating the premise for understanding the role of structure in organizations. They have concisely presented all the contributions until 1980 around organization structure.

Some important studies have highlighted the significance of structure. For example, Williamson (1975), has shown that structure could be responsible for inter divisional competition. Rajan et al (2000), have studied the effect on internal power struggles in internal capital markets. Structure is the root cause for power asymmetry in organizations, although it is designed to facilitate uniform power distribution (Foucalt, 1982). Also, Structure of an organization has implications for incentives and wage allocation (Marino and Zabojnik, 2004). Some authors like Baligh (1994), have gone the length to even state that organizational culture and organizational structure are two sides of the same coin and inseparable.

Moving ahead, structure is always stated to follow strategy (Chandler, 1962) and according to Hall (1977) it has two major functions. Firstly, it is created to minimize or at least regulate the influence of individual variations on the organizations and secondly to provide a setting in which power can be exercised, decisions can be made, and the organizations activities can be carried out. Campbell et al (1974), consider it to be the physical boundary of organization within which the structuring of the individuals takes place.

Moreover, the contributions of Boxall and Purcell (2000), and Rees and Edward (2009), highlight the significance of structure of organizations, indicating how structure has to be adaptable to change. These studies state that structure of the organization is an end result of strategy that is adopted by the firm. And that structure is what guides the further designing and adoption of various processes and practices. Whatever the strategy, firms are always in need to alter, modify, or design the processes and practices frequently. A major trigger to changing processes and practices by modifying the structure, are M&A's (Hitt et al., 2001). So, in order to understand how the human dimension affects the performance of M&A's a careful diagnosis of the role structure plays in M&A's is quintessential (Birkinshaw and Bresman, 2000).

Hitt et al. (1998), have proposed in their paper that structural changes in M&A's is an attribute of unsuccessful results. Hence, in order to minimize or eliminate negative impact on the performance of M&A's structural changes should be avoided post-merger. Hence, to achieve this an organization should engage in a profound and detailed pre-merger evaluation and analysis of the target firm (Haspeslagh and Jemisons, 1991a). And chose a firm which is similar or identical to the acquiring firm in regards to structure, as this would facilitate lesser changes to the target firms structure post-merger leading to higher order integration and subsequent realization of superior synergy and performance on other dimensions such as financial or behavioral. Thus, structure would be the fourth metric for pre-merger evaluation of M&A's, along with strategic logic, behavioral aspects, and financial aspects (Straub et al, 2013).

Organizational processes and practices and their impact on HR

Pugh et al (1968) operationalized five dimensions of structure in their study of fifty-two different organizations using a sixty-component scale. These dimensions even today hold good to explain most of the structure's organizations follow, whether U form, M form, H form or more recent additions the matrix form and network form (Williamson, 1975; Milgrom and Roberts, 1990,1992; Roberts, 2004). The organizations, as stated earlier would adopt a certain form based on the dominant strategy, they employ during a certain period which is

based on the market forces in which they function. Although, there are inherent issues involved with the adoption of any form of structure it is not the subject matter of this paper. The emphasis here is to understand that the structures adopted by the organizations are based on certain selection criteria employed by the decision makers. And that the structural composition of the organizations dictates the overall processes and practices within a firm. Within the overall processes and practices, the central argument is around HR processes and practices.

Based on the contributions of Rees and Edward (2009) and Eriksson and Ortega (2014), it can be posited that the HR practices and processes are intertwined with and emanating from the choice of structure a firm adopts due to the external and the internal forces in play during its life cycle. Hence, when a strategy is affected by the market forces or due to some internal forces it will impact the overall processes and practices of a firm. Thereby, affecting the HR processes and practices of the firm (Rees and Edward, 2009). Leading to changes in the HR strategy adopted by the firm frequently (Legnick Hall and Legnick Hall, 1988). Hence, whenever the overall strategy of the firm is impacted it could percolate down to the HR strategy changes, as in the case of M&A's.

In M&A's whether the reason is corporate diversity or growth, the overall strategy of the firm is revisited to accommodate the decision and hence this also impacts the HR strategy (Blake and Mouton, 1964). In case of M&A's as discussed by Puranam and Singh (2006), if the target firm is allowed to maintain autonomy then there will be no or very minimal changes to the HR practices and processes in both firms. On the contrary if the target firm is sooner or later merged with the acquirer then there will be changes in HR practices and processes in a varying degree (Eriksson and Ortega, 2014). It is needless to say that these changes will impact the structure of the firm or vice versa. It is necessary to state at the onset that this paper is emphasizing on the changes to the HR practices and processes, but it is not to say that the other practices and processes of the firms are not getting affected.

Ergo, we have discussed so far that when a firm decides to grow inorganically through M&A. Research states that there will be resistance presented by the acquiring firm as well as the target firm to changes in the structure of the two firms. As the structural changes would result in disruption of the processes and practices of the firms. This results in changing the HR practices and processes. And hence, the people of the firm avoid and resist integration. Which in turn leads to unrealized synergies and subsequent failure of M&A's. To avoid this failure, it can be argued that, in their pre-merger stage organizations can focus their search and evaluation around finding structurally similar firms as targets (Yang and Hyland, 2012; Villadsen, 2013). These structurally similar organizations can be searched for in the event of related or unrelated M&A's. One simple way to find such organizations is to focus on isomorphs (DiMaggio and Powell, 1983). Hence, at this juncture it is imperative to introduce 'Intentional Structural Homogeneity' as a strategy for improving the success rate of M&A's. And the way it is associated with the above discussion.

THEORETICAL DEVELOPMENT/MODEL

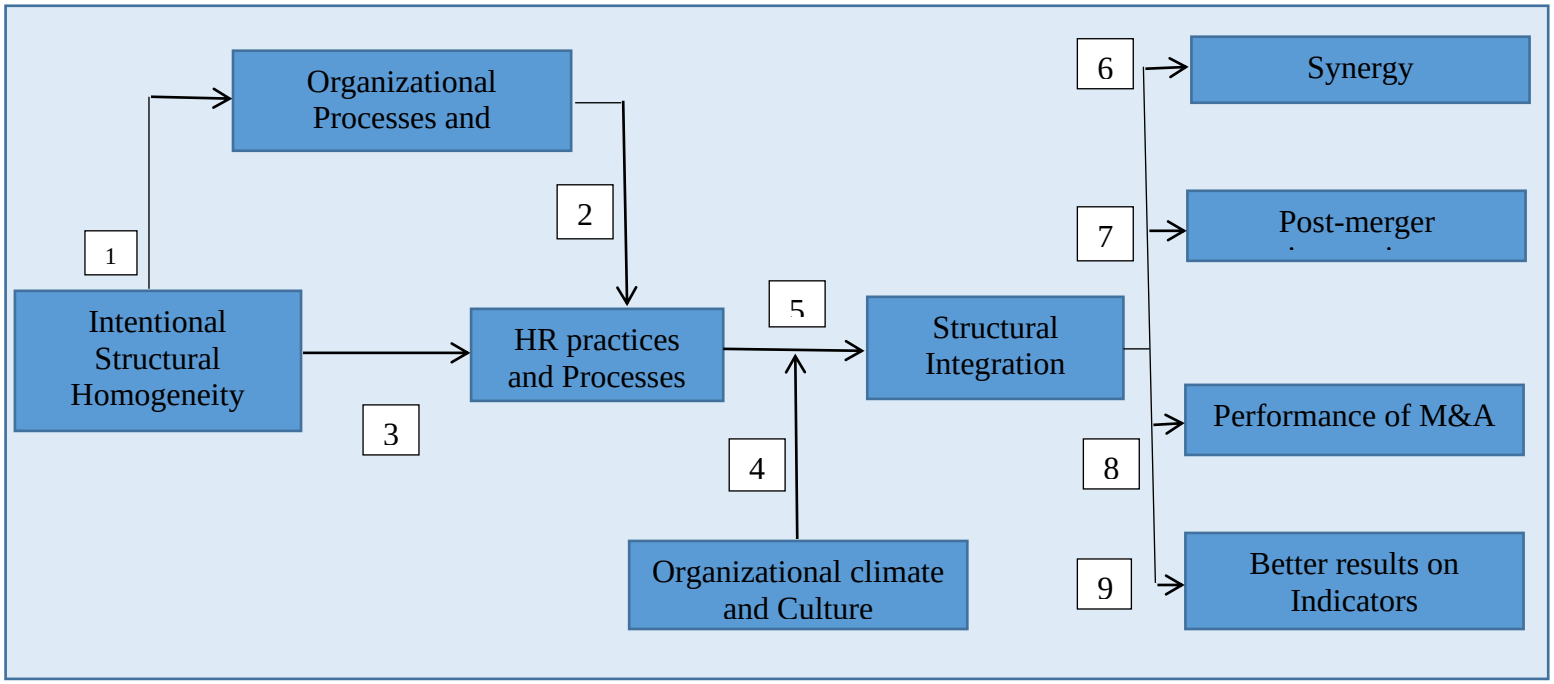
Intentional Structural Homogeneity (ISH)

It is evident that all the firms in a certain market compete for market share, labor, finance, and other such resources (DiMaggio and Powell, 1983). Hence, over a period of time in their life cycle they tend to adopt and foster certain traits required to build effective frameworks for survival against varied market pressures. Hence, in a similar or an analogous environment there are players who have developed similar practices and processes for survival or excellence also referred to as isomorphs (DiMaggio and Powell, 1983). Contrary to this, there are players who have adopted different strategies for survival and excellence (Straub et al, 2013).

As already established, the human capital is the most crucial element for M&A's success or failure. Therefore, Intentional Structural Homogeneity is a strategy under which the acquiring firm searches, identifies, and merges with target firms which are structurally

similar to or isomorphs of the acquiring firms. With the intention of minimizing the possible changes or disruptions to the HR practices and processes within the realm of overall practices and processes as a result of M&A's. Thereby, reducing or negating the impact felt directly by the human resources of both the firms. The term is a noun as well as verb as it describes the action taken with an intention to achieve better structural integration as compared to normal M&A's. This is significant mainly because of the possible implications (as discussed later) on M&A pedagogy and practical applications by the managers. ISH falls under strategy as a primary plan adopted by either the firms which are conducting their maiden M&A or the ones who have a history of acquisition. However, it would be more beneficial for the organizations which have accumulated experience around M&A's (Haleblian and Finkelstein, 1999). Adding, to this let us consider the following propositions which could be theorized (Please refer to the model in diagram 1.1).

Diagram 1.1



Explanation of the Model

Based on the model shown in diagram 1.1. The following are the variables
Independent variable: Intentional Structural Homogeneity
Dependent Variable: The HR processes and practices of the merging firms
Intervening/Mediator variable: The overall processes of the Merging firms
Moderator: The organizational climate and culture *second level*
Dependent variable: Structural integration
Consequences: Synergy, Performance of M&A, Post-merger integration, Indicators such as but not limited to: economies of scale, economies of production (Straub et al, 2013).

The relationship depicted in 1, 2, and 3 are the arguments of this paper and the remaining relationships have been already established by precious research and are ingratiated in the discussion of this paper.

Propositions and relative Implications for M&A decisions

Nahavandi and Malekzadeh (1988) state that, during M&A's the effort of the acquirer to integrate with the target firms is often hampered by the inertia of the target firm and also in some cases the outright resistance on the part of acquired firm. The Intentional Structural Homogeneity, in most cases would help the acquiring firms to overcome this resistance and inertia as in these cases mostly the inertia or resistance by the members of the firm is based on the possibility of radical changes and disruption to the target firms' structure and therefore the routines and in general way of doing business.

Utilizing ISH will help in reducing the overall impact on the processes and practices adopted and developed by the target firm over a period of time. As, these target firms are isomorphs or structurally similar firms to acquiring firm, the need for radical change to the structure of the firms within which the processes and practices are embedded, will be limited. Barkema and Schijven (2008), have demonstrated through their research on "related acquisitions", that there is higher synergistic potential and hence the possibility of successful mergers is higher when firms have similar structure. Also, this could also lead to higher levels of strategic similarity (Capron and Mitchell, 2009) between the two merging entities and thus higher levels of relative absorptive capacity (Zahra and George, 2002) would facilitate their integration furthermore. Hence, Intentional Structural Homogeneity takes this a step further to posit that whether related or unrelated, (Horizontal or vertical) acquisitions, if the organizations are structurally similar or isomorphs the chances of success could be relative higher;

Proposition 1: Intentional Structural Homogeneity leads to reduced changes/disruption in overall processes and practices embedded in the organizational structure.

Proposition 2: Organizational processes and practices (Overall) act as mediators in the relationship between Intentional Structural Homogeneity and HR processes and practices.

Also, ISH impacts in reducing disruptions to HR processes and practices directly. The implications of this proposition in managerial parlance would be that during M&A's the managers will have to face less resistance and inertia. As, the two firms would be having similar processes & practices the requirement of changes in routines will be minimized. And hence the relative avoidance of business disruptions (Hannan and Freeman, 1984). HR processes and practices will not get disrupted due to ISH adopted during M&A's. As a result, it will help in overcoming what Creasy, Stull and Peck (2009) state as 'Culture clash'. And hence better acculturation will result. Overall, it will help in better HR integration of the two firms and also the rationalization of the post-merger choices will be enhanced (Rees and Edward, 2009). The relationship between task integration and human integration has already been demonstrated by Birkinshaw, Bresman, and Hakanson (2000). Intentional Structural Homogeneity will also help in higher task integration due to effective human integration as a result of proposition two.

Intentional Structural Homogeneity will also help in reducing what Heimericks, Schijven, and Gates (2012) state as causal ambiguity due to the heterogeneity in sub activities having multiple dimensions, which subsequently creates better integration of the human resource processes and practices. Another implication will be the reconciliation of what Monin et al (2010) posit as the sense making around justice in post-merger integration between, what they term "Symbiotic Mergers". Intentional Structural Homogeneity will help in bringing about more clarity around the mergers as the intentions will be revealed in premerger stage. Moreover, due to lack of changes made to the processes and practices in the target firm the procedural justice and distributive justice will be evident. The end result will be the possibility of improved post-merger integration. Hence the following propositions are adept to explain the relationship between minimized changes to structure and holistic structural integration.

Proposition 3: Higher level structural integration is achieved by an organization which adopts Intentional structural homogeneity during pre-merger stages.

Proposition 4: Organizational processes and practices and HR processes and practices act as mediators in the relation between Intentional Structural Homogeneity and structural integration.

Adding to this, ISH will help in creating positive synergies by allowing for better resource complementarity between the merging firms. As the two firms will be structurally similar there will be similar resource requirements of both firm under the resource-based view (Barney, 1991).

Proposition 5: ISH results in positive synergy realization through resource complementarity.
The homogeneity maintained under ISH would help in reducing the time required in realization of performance objectives of M&A's. As there will be lesser steps to be followed in integration of the firms.

Proposition 6: ISH will help in reducing the time taken for realization of performance objectives of M&A.

ISH will help in more efficient adoption and application of intermural organizational best practices between the two firms. As the two firms will be structurally similar the best practices would not require multiple amendments for adoption and the application would also be easier due to familiarity in features of the best practices.

Proposition 7: ISH facilitates intermural transfer and adoption of organizational best practice.

Implications for future research

The concept of Intentional structural homogeneity (ISH) is an unconventional and novel concept. This concept is presented as a possible strategy for future mergers and acquisition activities. The ISH has to be studied empirically, a case-based approach will be very apt for the purpose. As there have been previous studies on mergers and acquisitions which have mostly followed case study approach and provided acceptable contributions in this field. The model designed in this paper is tentative and could definitely benefit by contributions by experts in this field. The model is intended to establish theoretically only points 1 to 3 and the remainder points 4 across 9 have been established from previous research in this field. The operationalization of the same will be a complicated, yet beneficial process.

Conclusion

In summary, so far, the paper has established through most studies done so far that; firstly, a key element in the failure of M&A's has been the human capital of the firms involved in the M&A's. Secondly, the structure of the merging entities plays a significant role in the process of integration. Thirdly, the processes and practices of the firms are designed and integrated around the structure that is adopted by the firm (Van de Ven et al., 1976). And these processes and practices are responsible for the development of routine rigidity (Tushman and Anderson, 1986; Teece, Pisano, and Shuen, 1997), which is evident in a firm's developing organizational inertia preventing successful integration. These already established findings from previous research were utilized to establish the basis for the genesis of the principal argument of the paper "Intentional Structural Homogeneity" (ISH) after establishing the backdrop in more detail.
ISH is placed as a strategic plan available to firms interested in M&A activity to reduce the negative impact on the human resources of the firms involved. As stated by Haspeslagh and Jemisons (1991a) and Hitt et al., (1998) a more in-depth analysis and evaluation in the pre-

merger stage helps in purchasing an acceptable target without adverse effects on both the firms. ISH will facilitate this position further. Possible propositions have been theorized and made accessible for future in depth exploration. These propositions if established empirically will be beneficial in bridging the identified gap in literature around successful performance of mergers and acquisition.

References

- Baligh, H. H. (1994). Components of culture: Nature, interconnections, and relevance to the decisions on the organization structure. *Management Science. 40*(1), 14-27.
- Barkema, H. G., & Schijven, M. (2008). Toward unlocking the full potential of acquisitions: The role of organizational restructuring. *Academy of management journal*, *51*(4), 696-722.
- Barney, J (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management.* 17. 99-120
- Barney, J. B. (2002). Strategic Management: From informed conversation to academic discipline. *Academy of Management Executive.* 16(2), 53-57.
- Birkinshaw, J., Bresman, H., & Håkanson, L. (2000). Managing the post-acquisition integration process: How the human integration and task integration processes interact to foster value creation. *Journal of management studies*, *37*(3), 395-425.
- Blake, R.; Mouton, J. (1964). *The Managerial Grid: The Key to Leadership Excellence*. Houston: Gulf Publishing Co.
- Boxall, P., & Purcell, J. (2000). Strategic human resource management: where have we come from and where should we be going? *International Journal of Management Reviews*, (2), 183-203.
- Buono, A. F., Weiss, J. W., & Bowditch, J. L. (1989). Paradoxes in acquisition and merger consulting: Thoughts and recommendations. *Consultation: An International Journal*.
- Capron, L., & Mitchell, W. (2009). Selection capability: How capability gaps and internal social frictions affect internal and external strategic renewal. *Organization Science*, 20(2), 294-312.
- Cartwright, S., & Cooper, C. L. (1990). The impact of mergers and acquisitions on people at work: Existing research and issues. *British Journal of Management.* 1(2), 65–76.
- Cartwright, S., & Schoenberg, R. (2006). Thirty years of mergers and acquisitions research: Recent advances and future opportunities. *British journal of management*, *17*(S1), S1-S5.
- Chandler, A.D. Jr. (1962). *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge, MA: MIT Press
- Child, J., Faulkner, D., & Pitkethly, R. (2001). *The management of international acquisitions*. Oxford University Press.
- Creasy, T., Stull, M., & Peck, S. (2009). Understanding employee-level dynamics within the merger and acquisition process. *Journal of general management*, *35*(2), 21.
- Cording, M., Christmann, P., & King, D. R. (2008). Reducing Causal Ambiguity in Acquisition Integration: Intermediate Goals as Mediators of Integration Decisions and Acquisition Performance. *Academy of Management Journal*, *51*(4), 744-767.
- Dalton, D.R., Todor, W.D., Spendolini, M.J., Fielding, G.J., and Porter, L.W. (1980). Organizational structure and performance: A critical review. *Academy of Management Journal*. 8, 49-64.
- Datta, D. K., & Grant, J. H. (1990). Relationships between type of acquisition, the autonomy given to the acquired firm, and acquisition success: An empirical analysis. *Journal of Management*, *16*(1), 29-44.
- Davis, G. F., & Stout, S. K. (1992). Organization theory and the market for corporate control: A dynamic analysis of the characteristics of large takeover targets, 1980-1990. *Administrative Science Quarterly*, 605-633.

- Dewing, A. S. (1921). *The financial policy of corporations: Volume I, corporate securities.* New York, NY: The Ronald Press Company.
- DiMaggio, P. J., & Powell, W. W. (2000). The iron cage revisited-Institutional isomorphism and collective rationality in organizational fields (Reprinted from the American Sociological Association vol 48, pg 147-160, 1983). *ADVANCES IN STRATEGIC MANAGEMENT, VOL 17, 2000*, *17*, 143-166.
- Eriksson, T., & Ortega, J. (2014). *Organizational Structure and Firms' Demand for HRM Practices* (No. 2013-09). Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.
- Foucault, M. (1982). The subject and power. *Critical inquiry*, 777-795
- Haleblian, J., & Finkelstein, S. (1999). The influence of organizational acquisition experience on acquisition performance: A behavioural learning perspective. *Administrative Science Quarterly*, *44*(1), 29-56.
- Hall, E. T. (1977). *Beyond culture*. Garden City, N.Y: Anchor Books.
- Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American sociological review*, 149-164.
- Haspeslagh, P., & Jemison, D. B. (1991). *Managing acquisitions: Creating value through corporate renewal*. New York, NY: Free Press.
- Heimeriks, K. H., Schijven, M., & Gates, S. (2012). Manifestations of higher-order routines: The underlying mechanisms of deliberate learning in the context of post-acquisition integration. *Academy of Management Journal*, *55*(3), 703-726.
- Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M. A., Singh, H., Teece, D. J., & Winter, S. G. (2007). Dynamic capabilities. *Understanding strategic change in organizations*.
- Hitt, M., Harrison, J., Ireland, R. D., & Best, A. (1998). Attributes of successful and unsuccessful acquisitions of US firms. *British Journal of Management*, *9*(2), 91-114.
- Hitt, M. A., Hoskisson, R. E., Ireland, R. D., & Harrison, J. S. (1991). Effects of acquisitions on R&D inputs and outputs. *Academy of Management Journal*, *34*(3), 693-706.
- Hoskisson, R. E., Hitt, M. A., Johnson, R. A., & Moesel, D. D. (1993). Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management Journal*, *14*(3), 215-235.
- Jarillo, J. C. (2003). *Strategic logic*. Palgrave Macmillan.
- Jemison, D. B., & Sitkin, S. B. (1986). Corporate acquisitions: A process perspective. *Academy of Management Review*, *11*(1), 145-163.
- King, D. R., Covin, J. G., & Hegarty, W. H. (2003). Complementary resources and the exploitation of technological innovations. *Journal of Management*. 29(4), 589.
- King, D., Dalton, D., Daily, C., & Covin, J. (2004). Meta-analysis of post-acquisition performance: Identifications of unidentified moderators. *Strategic Management Journal*. (February), 187–201.
- Larsson, R., & Finkelstein, S. (1999). Integrating strategic, organizational, and human resource perspectives on mergers and acquisitions: A case survey of synergy realization. *Organization Science*, *10*(1), 1-26.
- Lengnick-Hall, C. A., & Lengnick-Hall, M. L. (1988). Strategic human resources management: A review of the literature and a proposed typology. *Academy of Management Review.* 13(3), 454−470.
- Marino, A. M., & Zabojnik, J. (2004). Internal Competition for Corporate Resources and Incentives in Teams. *RAND Journal of Economics.* 35, 710-727

- Marks, M. L., & Mirvis, P. H. (2001). Making mergers and acquisitions work: Strategic and psychological preparation. Academy of Management Executive, 15, 80–92.
- Milgrom, P. and J. Roberts (1992). *Economics, Organization & Management*. Prentice Hall
- Mintzberg, H. (1987). The strategy concept 1: five p's for strategy.
- Monin, P., Noorderhaven, N., Vaara, E., & Kroon, D. (2013). Giving sense to and making sense of justice in post-merger integration. *Academy of Management Journal*, *56*(1), 256-284.
- Nahavandi, A., & Malekzadeh, A. R. (1988). Acculturation in mergers and acquisitions. *Academy of management review*, *13*(1), 79-90.
- Pablo, A. L. (1994). Determinants of acquisition integration level: A decision-making perspective. *Academy of management Journal*, *37*(4), 803-836.
- Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1968). Dimensions of organization structure. *Administrative science quarterly*, 65-105.
- Puranam, P., Singh, H., & Zollo, M. (2006). Organizing for innovation: Managing the coordination-autonomy dilemma in technology acquisitions.*Academy of Management Journal*, *49*(2), 263-280.
- Puranam, P., Powell, B. C., & Singh, H. (2006). Due diligence failure as a signal detection problem. *Strategic Organization*, *4*(4), 319-348.
- Puranam, P., Singh, H., & Chaudhuri, S. (2009). Integrating acquired capabilities: When structural integration is (UN) Necessary. *Organization Science*, *20*(2), 313-328.
- Puranam, P., Singh, H., & Zollo, M. (2006). Organizing for innovation: Managing the coordination-autonomy dilemma in technology acquisitions.*Academy of Management Journal*, *49*(2), 263-280.
- Rajan, R., Servaes, H., Zingales, L., 2000. The cost of diversity: the diversification discount and inefficient investment. Journal of Finance 55, 35–80.
- Rees, C., & Edwards, T. (2009). Management strategy and HR in international mergers: choice, constraint and pragmatism. *Human Resource Management Journal*, *19*(1), 24-39.
- Roberts, J. (2004). T*he Modern Firm: Organizational Design for Performance and Growth*. Oxford University Press, Oxford
- Schweizer, L. (2005). Organizational integration of acquired biotechnology companies into pharmaceutical companies: The need for a hybrid approach. Academy of Management Journal. 48(6). 1051-1074.
- Shaver, J. M. (2006). A paradox of synergy: Contagation and capacity effects in mergers and acquisitions. Academy of Management Review, 31(4), 962–976.
- Sirower, M. L. (2000). The synergy trap: How companies lose the acquisition game (Asia-pacific edition). Australia: Simon & Schuster.
- Stahl, G. K., & Voigt, A. (2008). Do cultural differences matter in mergers and acquisitions? A tentative model and examination. *Organization Science*, *19*(1), 160-176.
- Straub, T., Borzillo, S., & Probst, G. (2012). A Decision-Making Framework to Analyze Important Dimensions of M&A Performance. *Advances in Mergers and Acquisitions*, (11), 199-235.
- Tan, C. L., & Nasurdin, A. M. (2011). Human resource management practices and organizational innovation: assessing the mediating role of knowledge management effectiveness. *Electronic Journal of Knowledge Management*, *9*(2), 155-167.

- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.
- Trautwein, F. (1990). Merger motives and merger prescriptions. *Strategic Management Journal*. 11, 283–295.
- Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative science quarterly*, 439-465.Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *American sociological review*, 322-338.
- Vaara, E., & Monin, P. (2010). A recursive perspective on discursive legitimation and organizational action in mergers and acquisitions. *Organization Science*. 21(1), 3–22.
- Vermeulen, F., & Barkema, H. (2001). Learning through acquisitions. *Academy of Management journal*, *44*(3), 457-476.
- Villadsen, A. R. (2013). Similarity or difference? The relation between structure and strategy isomorphism in public organizations. *British Journal of Management*, *24*(S1), S62-S75.
- Weber, Y., & Tarba, S. Y. (2010). Human resource practices and performance of mergers and acquisitions in Israel. *Human resource Management review*, *20*(3), 203-211.
- Williamson, O. E. (1975). Markets and hierarchies: Analysis and antitrust implications. New York, NY: Free Press.
- Yang, M., & Hyland, M. (2012). Similarity in cross-border mergers and acquisitions: Imitation, uncertainty and experience among Chinese firms, 1985–2006. *Journal of International Management*, *18*(4), 352-365.
- Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of management review*, 27(2), 185-203.

DECISION SCIENCES INSTITUTE

An Assessment of Two-sided Matching Method of Strategic Partner Selection in the Business Model Context

Santosh Nandi, Ph.D.
Assistant Professor of Management
University of South Carolina Sumter
200 Miller Rd.
Sumter, SC 29150
Email: nandisan@uscsumter.edu; santoshn@usca.edu

ABSTRACT

In this study, partner selection method is explored. The theoretical and empirical aspects of two-sided matching models are discussed and compared with one-sided discrete choice models. Its applicability in "who partners with whom" and "whom to partner with" is explored using the context of business models. The study asserts that two-sided matching model approach in a "market of partners" enables two-sided decision making, and thus, helps in achieving greater complementary resources between partners. The study also elaborates the limiting conditions of its applicability towards the end of the paper.

KEYWORDS: Two-sided Matching Model, Partner selection method, Strategic partnership, Business Model

INTRODUCTION

Organizations face considerable challenge in developing fruitful strategic partnerships in their respective business model. Partnerships become burdensome, rather than beneficial, when objectives and goals are misaligned, or lack mutuality to realign. As a result, synergies in attributes between partners are given considerable importance in partner selection (Mitsuhashi & Greve, 2009). This notion of synergy narrows strategic partnership research outlook at least in two ways. First, it observes partnership formulation either from a focal firm or a dyad perspective, rather than a partner market perspective (i.e., a market of partners). Second, it ignores to account that a firm may be highly constrained in seeking a preferred partner since its competitors may also be seeking the same partner and vice versa, thus, creating competition to gain greater complementary resources and to realize higher value (MacDonald & Ryall, 2004). To add more confusion, scholars have proposed various one-side models of partner selection, such as single-agent discrete empirical method which intrinsically assumes that firms do not face rivalry from their rivals or have any constraints in choosing their preferred partners (Train, 2009). The discrete choice model approach of partner selection using focal or dyad firms perspective seems seriously flawed as it fails to capture the logic of assortment for symbiotic partners and value. Likewise, the random utility terms also fail to satisfy the structural assumptions of such discrete choice models. At the very least, this argument reveals there exists a gap in both theoretical and empirical understanding of strategic partner selection.

To bridge this gap, we discuss about a "matching model" partnering approach in the context of business models to look for "who partners with whom" and suggests "whom to partner with" (Manski, 1975; Fox, 2016). Commonly applied by sociologists and economists, this

partnering approach is capable in comparing with and competing among available partners in the market based on their utility parameters. By asserting matching model approach in a "market of partners", the resulting strategic partnerships are seen as an outcome of two-sided decision making to achieve greater complementary resources between partners (Becker, 1973; Mortensen, 1988; Mindruta et al., 2016). An underlined assumption of this matching model approach is existence of complementary resources in partners as a necessary, but insufficient condition about how partners are selected. In other words, the model approaches the competition in the "market of firms" in a way that each partner can create its symbiotic choices through assortment of partners that can maximize their levels of joint value creation (Fox, 2016). Matching model approaches allow identification of complementary resources between partnering firms based on the maximization of joint value creation after analyzing all possible options for value creation between all realized and potential matched partners in the market.

In this paper, the maximum score estimator (Manski, 1975; Fox, 2016) is discussed to understand its applicability in identifying the complementary (or substitute) resource combinations in strategic partnership formation, and assess its strengths and weaknesses in partner selection in business model context. Accordingly, the following sections in this study explains the theoretical concepts of this matching model approach by comparing and contrasting with the traditional focal or dyad firm discrete or binary choice models. Also, the paper demonstrates the maximum score estimator that enables top-down sorting of partnerships for identification of the combination that achieves maximum joint value creation in a business model. Also, the paper makes recommendations towards "partner-selection" methodology in business model context. The contribution of the paper lies in its representation of how matching models can be applied in selecting partners to achieve successful business models.

THEORETICAL CONCEPTS

Concept of Business Model Construct

In the last one and half decade, scholars have shown growing interest in business models and their innovations (collectively, BMI) (e.g., Chesbrough, 2007, 2010; Teece, 2010; Zott & Amit, 2010). Apart from its mediating role between a firm's technical potential and its economic values (Chesbrough, 2007), business models also allow firms to create and maintain sustainable competitive advantages (Pohle & Chapman, 2006). Demonstrating examples from Xerox's spin-off businesses, Chesbrough & Rosenbloom (2002) state that "*Business models are perhaps the most discussed and least understood aspect....*" (p533). Uber, LinkedIn, WhatsApp, Amazon and Android OS are just a few successful business models that justifies the contextual relevance of business model research.

Strategic Partnerships and Partner Selection Methods

While business models hold contextual relevance for conducting empirical studies, strategic partnership is an important element of business models. For example, strategic partnerships enable firms to reach to valuable resources beyond their boundaries. Literature reveals that TCE and RBV theories can collectively explain some of the criteria for such strategic partnerships from efficiency and novelty perspective. However, a major challenge for strategic decision makers is about whom to partner with for mutuality in successful outcomes. Although, considerable research has been conducted in strategic alliance area, the understanding of research methods for partner selection appears to be limited. For example, sociology, operations, and economics researchers have used different matching approaches to

examine dyad and multi-partner selection situations, such as marriages (Becker, 1973), buyer–supplier relationships (Chatain, 2011), firm–university relationships (Mindruta, 2013), sponsorship of sport teams (Yang et. al, 2009). Some of the known partner-selection methods in strategic alliance literature include discrete and binary choice models (e.g., logit, conditional logit, probit, rare-event probit, selectivity-corrected probit). But these methods are focal-firm centric and rigid in prediction. On the contrary, maximum score estimator method (Manksi, 1975) is considered to be a multinomial and flexible partner selection method, since it has the capability to include partner characteristics to be continuous variable (e.g., firm size, firm age) or discrete variable (e.g., total number of patents owned), and even, match-specific (e.g., how far are the partners located or how similar are their resources/capabilities).

Logic of "Markets of Partners"

At its core, strategic partnerships are voluntary arrangements between two or more firms to access complementary resources outside each other boundaries for value creation (Ahuja, 2000). However, it is a more complex phenomenon than it appears to on the surface. For instance, when firms start looking for partners in their respective markets, it becomes a two-sided choice of "whom - among several - to partner with" based on their combined capabilities to realize maximum value (Ahuja et al., 2009). This two-sided voluntary conduct suggests that both partners must approve in partnering with each other based on their internal evaluations of value maximization. Moreover, research show that strategic partnership is not just about seeking a partner that possesses the resources or capabilities lacked by a focal firm lack – an arrangement referred to as additive fit. Rather it is about achieving complementary fit with a partner where in the combination of each other resources or capabilities also leads to an incremental change in the total output. It occurs due to increase in the focal firm's capabilities caused due to its partner's higher levels of the other capabilities (Milgrom & Roberts, 1995). For example, a smaller firm's brand image significantly increases when it partners with a bigger firm. Strategic partnership literature refers this phenomenon as resource complementarity or asset complementarity.

Another important aspect of the "markets of partners" results from the scope to partner that leads firms to compete for partners who are believed to have greater complementary resources than others. For example, when firm possesses scarce resources, all other firms compete to secure it ahead of others. However, that particular firm also considers several strategic factors in choosing its ultimate partner or number of partners. Likewise, Stuart (1998) examined status homophily and noted that low-status firms may be constrained in their access to high-status firms. In nutshell, this indicates that two firms choose to partner when they jointly anticipate resource complementarities, or either firm cannot create any greater value with other available firms in the market. Therefore, this study argues that an appropriate theoretical and empirical approach using both-sided matching logic can fill the gap central to partner selection in the context of business models.

Theoretical Roots of Matching Models

Matching Models enables economists, sociologists and strategists to study relationships in general, i.e., "who is in a relationship with whom" (Fox, 2009). Scholarly empirical examples include a range of situation, such as which all businesses merged and with whom, which all couples (which all men with which all women) got married, and which all bidding firms earned which region-specific spectrums in US telecom spectrum auction. As such, it provides a theoretical framework to mathematically search and match for an equilibrium structure (i.e., formation of mutually most beneficial relationship) of who matches with whom in a market that

consists of heterogenous agents – each having preferences over whom to transact with, and that the agents compete among each other to attain a partner or partnership that is believed to be better than the rest (e.g., Koopmans & Beckmann, 1957; Shapley & Shubik, 1972; Becker, 1973; Fox, 2009; Fox, 2016). According to Fox (2009):

> *"a matching model takes a set of payoffs or outputs for all possible matches and produces a set of matches where no pair would prefer to deviate and become matched, instead of their assigned matches. The robustness of the equilibrium to deviation by any potential couple suggests 'pairwise stability' as the term for an equilibrium. The economic notion in matching models is the rivalry to match with the most attractive partners. In marriage, men compete with each other to marry the most attractive women while women compete with each other to marry the most attractive men. There is scarcity on both sides of the market."* (p.1)

Two-sided Matching Models

In its simplest understanding, a two-sided matching model divides the competing agents into two sides, such as men and women (Roth & Sotomayor, 1990). It can be further classified into models with and without transfers. For instance, Gale and Shapley (1962) suggested a framework of multiple pairwise stable partners, where men have preferences over women and women have preferences over men, but, there is no transfer or exchange of money. In contrast, Becker (1973) study marriage models where matched agents can exchange money. Further, two-sided matching models can be categorized based on the choice of matching model that the market in question. In one-to-one matching models, such as marriage where one man marries at most one woman, typically only one set of pairwise stable physical matches are established. Likewise, there can be many-to-one, one-to-many and many-to-many two-sided matching models based on other contract elements, such as the hours of work in a labor market, are determined as part of the pairwise stable outcome.

In business model context, the theoretical model of two-sided matching model can be either a one-to-one, many-to-one, or many-to-many matching models based on the type of strategic partnership required for its business model. For example, the business model of the focal firm may require the firm to engage in a marketing partnership, a buyer-supplier partnership, as well as a R&D partnership at the same time point; however, each of these partnerships signifies matches meant for different functions in different contexts. In order to simplify this discussion, it is assumed that each business model is critical to the focal firm, and therefore, a generic form of strategic partnership occurs between a pair of focal and non-focal firm depending on its engagement requirement (such as, marketing, buyer-supplier, and others alike) of its business model. Accordingly, a focal firm is labeled as ***Fi*** and a non-focal firm is labeled as ***Ni***. The model integrates two-sided preferences as need for complementary resources for value creation. Therefore, the joint production function is defined as ***f* (*Fi, Ni*)**. In the following section, this study adapts the modeling approach in Fox (2010, 2016) and Mindruta et al. (2016) to explain the sequential process for identification of the joint production function (joint value creation function) of all pairs of focal and non-focal firms in business model context.

Two-sided Matching Model Specification

Let ***v*** denote the value of returns that a firm receives from a matched pair and ***t*** denote the transfer price it should pay to the other partner. By adding this transfer price (transfer utility) construct into the model, it creates conditions for firms by which they can pay higher transfer prices to attract better quality partners, or in contrast, lower its quality requirement in order to

attain greater quantity. As such, firms need to negotiate their transfer. Also, it is assumed that focal firm makes the transfer to non-focal firms in exchange of gaining the required complementary resources, however, transfers may be positive or negative. Therefore, the focal firm ***Fi***'s net payoff from a paired-match shall be ***vF(Fi, Ni) + tFiNi*** and the non-focal firm ***Ni***'s net payoff is ***vN(Fi, Ni) – tFiNi***.

The joint production surplus is the summation of both the payoffs, i.e.,

$$f(Fi,Ni)= vF(Fi,Ni)+ vN(Fi,Ni) \quad (1)$$

According to matching logic, the non-focal firm ***Ni*** selects the focal firm ***Fi*** depending on ***Ni's*** payoff from partnering with ***Fi***, depending on unlikelihood of getting a higher payoff for partnering with any other possible focal firms who may also willing to work with ***Ni***. Likewise, the focal firm ***Fi's*** willingness to enter into a partnership with ***Ni*** also depends on its own effectiveness from the relationship as well as its perception of ***Ni*** as the most "effective" choice set among the set of other non-focal firms who may also be willing to partner with ***Fi***, based on their own alternate choices. Thus, a "pairwise stability" is expected to be achieved based on this state of equilibrium portrayal of this matching model logic. As such, pairwise stability means that either of the firms in a stable match would not be willing to abandon its partner for another partner, since the payoffs of both focal and non-focal firms in paired match is expected to be higher than the payoffs they could get from any alternate partnership in the market.

Empirically, a pairwise stable configuration of partnership is defined based on two features. First, the payoffs of each partner in a match are significantly larger than their extreme choice of not to participate in the market. Let ***vF(Fi)*** denote the payoff of a focal firm if it doesn't need any partner resources and performs on its own (internally). Thus, voluntary participation for all matches (***Fi,Ni***) must satisfy:

$$vF(Fi,Ni) + tFiNi > vN(Fi) \quad (2)$$

and,

$$vN(Fi,Ni)> tFiNi \quad (3)$$

Therefore, based on the inequalities (2) and (3), it can be inferred that transfers sets a lower bound for firms to participate in the market for partnership, below which is it considered unacceptable.

Second, for any two probable matches ***(Fi, Ni)*** and ***(Fj, Nj)***, the following condition implies:

$$vF(Fi,Ni)+ tFiNi > vF(Fi, Nj) + ₮FiNj \quad (4)$$

where $₮FiNj = vF(Fi,Nj) – (vF(Fj,Nj) − tFjNj)$

To infer condition (4), ***₮FiNj*** can be considered as the maximum transfer that the non-focal firm ***Nj*** may be willing to pay to focal firm ***Fi*** when it anticipates switching from its current equilibrium partner ***Fj*** for partnering with ***Fi***. The maximum transfer limit for the non-focal firm ***Nj*** is the returns ***vN(Fi,Nj)*** that ***Nj*** can obtain in the alternate partnership with ***Fi*** and the net payoff ***vN(Fj,N j) – tFjNj*** that ***Nj*** enjoys in the relationship with its actual partner ***Fj***. The central theme of the inequality (4) is that even if the non-firm ***Nj*** offers to pay the focal firm ***Fi***, the maximum transfer, focal firm ***Fi*** still prefers to pair up with ***Ni*** and receive ***tFiNi,*** instead of partnering with ***Nj*** and receiving ***₮FiNj***.

The depicted relationships (1) to (4) suggest that that neither one-sided preferences for a particular partner nor dyad-level acceptance are sufficient conditions for a partnership to occur. Instead, each dyad-level decision is controlled by the opportunity cost of unavoidable

partnerships. This implies that pairwise stable equilibrium matches are unique and sustainable in one-to-one matching games with transfers (Roth and Sotomayor, 1990). Moreover, the equilibrium matches are considered as optimal since all the matches in the market is believed to collectively create the maximum summated value in the market (Becker, 1973).

Estimation Methods for Two-Sided Matching Models

Matching models are similar to that of the popular static and discrete Nash equilibrium pay-off games, such as entry models (Berry, 1992; Bresnahan & Reiss, 1991), prisoner's dilemma, coordination games, network traffic and alike. These models face estimation challenges. In game theory, the Nash equilibrium is a solution concept of a non-cooperative game involving two or more players in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only his or her own strategy (Osborne & Rubinstein, 1994). However, matching models use pairwise stability of the agents in the market (possibly in hundreds or even thousands), unlike the Nash equilibrium with few agents (e.g., three or four firms deciding to enter a market in some entry applications). This, the number of agents in matching empirical models can create computational challenges in estimating the right solutions. There are at least six known estimation methods for two-sided matching models (Fox, 2009) as explained below:

Nested Solution Methods. It uses simulated maximum likelihood or the simulated method of moments to estimate the outcomes. It solves the model by running it for a fixed number of times (each iteration) based on its outer optimization routine. They may appear conceptually straightforward but computationally burdensome. Using wage data of public school teachers, Boyd et al. (2003) applied this method of moments to study their matching c to the public schools in New York state. The study assumes that their wages are determined based on their skills.

Full Likelihood Methods. It uses an augmented full-likelihood approach where the unobserved payoffs of each match are treated as nuisance parameters and integrated out using a blocking structure in a Markov Chain Monte Carlo (MCMC) procedure. For instance, Sorensen (2007) used this approach in studying the matching of venture capitalists to entrepreneurs. This method can be computationally rigorous in large matching markets. However, likelihood become simpler if the values of realized matches are recorded in the data as payoffs of matches Akkus (2008).

Maximum Score Estimator Methods (Inequality Methods). This (inequality) method maximizes the number of inequalities implied by pairwise stability of the matches. One of the greatest advantage of this approach is that it breaks the computational curse of dimensionality (i.e., traditional assumptions of dimensionality is the data is not necessary) because the estimator doesn't require all inequalities to be included (Manski, 1975). This method applies for one-to-one as well as many-to-many match models. For instance, Fox (2009; 2016) applied a maximum score estimator to estimate a many-to-many matching game between automotive suppliers and automotive assemblers where transfers are endogenous, but not recorded in the data.

Logit Methods. This method studies matching games with transfers by assuming that the payoffs to matches have error terms that satisfy the parametric logit property (Dagsvik, 2000; Choo & Siow, 2006). Such as assumption allows studies to derive closed-form equations that allow estimation in extremely large markets (such as the US national marriage market).

Identification Methods. In marriage based matching model, the agents on the same side of the market (i.e., men) are also rivals to match with agents on the other side of the market (i.e., women). When a man couldn't not match with the most attractive woman, it does not imply that the man did not wish to marry that woman. This means that the equilibrium budget set of each agent goes unobserved until a stable match is found. Therefore, it becomes tough to identify immediately from data on who matches with whom. To resolve this identification problem, Fox (2009) suggested this identification method that first, identifies relative importance of complementary resources in payoffs (such as being better educated vs. being wealthier), and second, identifies those matches that gives higher payoffs (Fox, 2009).

Selection Correcting Outcome Equations. In this method, the auxiliary outcome equation of the matching model is parametrically selected (Sorensen, 2007). In his study of matching of venture capitalists to entrepreneurs, Sorensen (2007) doesn't determine the outcome (i.e., success of an investment) as part of the matching game, instead the outcome is only observed in the data for the realized matches. This approach is somewhat similar to using a single agent decision model to selection correct an outcome equation (Heckman, 1979 – Sample selection bias correction).

APPLYING MAXIMUM SCORE ESTIMATOR FOR TWO-SIDED MATCHING MODELS

Although matching theory existed since several decades in parcels (e.g., Gale & Shapley, 1962), the full-blow and empirically integrated studies have just started (e.g., Fox, 2009, Mindruta et. al, 2016). This is due to the complexities in estimating market outcomes (i.e., the matching parameters and their logics are interrelated and requires that simultaneously modelled). As a result, the data violates the traditional assumptions of dimensionality (Fox, 2016). In addition to it, the model becomes vulnerable to pre-determined equilibrium selection rule for a likely matched configuration to occur among all possible market-wide configurations (Fox, 2010). In order to avoid these two shortcomings, this study finds the maximum score estimator methodology (inequality method) to be appropriate to be applied to estimate matching models for strategic partners in business model context. Also, this maximum score identification approach (Manski, 1975; Fox, 2016) address the computational complexities of modeling matching games more efficiently than other matching model estimation techniques (Fox, 2010; Fox and Bajari, 2013). As a matter of fact, there are two key aspects in the joint production function that this estimation method focuses on. First, it identifies and sorts the combination of partner attributes (complementary or substitutable) that leads to partnership formation. Second, it assesses their relative importance in explaining the sorting patterns that controls the partner selection.

Similar to Fox (2010, 2016), this study suggests that the joint value creation function (Equation 1) is identified without any parameter specification, and suggests a technique to estimate the unknown parameters stated for the pairwise stability condition in equations 1-4 above. For the two matches ***(Fi, Ni)*** and ***(Fj, Nj)*** in a pairwise stable match, the following inequality applies:

$$f(Fi, Ni) + f(Fj, Nj) > f(Fi, Nj) + f(Fj, Ni) \quad (5)$$

According to Fox (2010, 2016) and Mindruta (2016), this local production maximization inequality (Equation 5) suggests that the sum of values created by any two pairs ***(Fi,Ni)*** and

(Fj,Nj) in should be equal to or greater than the sum of values created by the swapped (i.e., alternated) pairs ***(Fi,Nj)*** and ***(Fj,Ni)***.

Next step is to calculate the coefficient estimates by maximizing the number of inequalities that satisfies the inequality (Equation 5) for all possible pairs of observed and swapped matches in the market. The mathematical formula of the estimator is stated as follows:

Let the joint production function be expressed as:

$$f(Fi, Ni) = f(Fi, Ni)|\beta) + \varepsilon Fi,Ni$$

$$f(Fi, Ni) = \Pi \beta\,(XFi * XNi, XFi, XNi) + \xi Fi + \xi Ni + \varepsilon FiNi \quad (6)$$

where,

XFi and ***XNi*** represent the vector of attributes of focal and non-focal firms respectively,

ξFi and ***ξNi*** are focal and non-focal firms' fixed effects, and

εFiNi are match-specific errors (without any traditional assumption about its distribution)

(XFi*XNi) is a term to denote the interactions between the observed attributes of partners in a match.

Also, let h be a market index, ranging from ***1 to M***, where ***M*** is the total number of observed markets (i.e., ***M=1*** represents a single large market). Let ***Ph*** be the total number of partnerships in a market ***h***. Within each market, the focal firms are denoted by F^i_h, $i=1,\ldots P_h$ and non-focal firms are denoted as N^i_h, $i=1,\ldots P_h$. Therefore, the global maxima of the production function, Q(***β***), is numerically calculated by summating all predicted inequalities for all pairs of observed and swapped matches in all M markets:

$$Q(\beta) = \sum_{h=1}^{M} \sum_{1<j,j<Ph} 1[f(F_i^h, N_i^h|\beta)+f(F_j^h, N_j^h|\beta) > f(F_i^h, N_j^h|\beta)+f(F_j^h, N_i^h|\beta)] \quad (7)$$

where **1[.]** indicates that the inequality in the parenthesis as true, and **0** as false.

According to (Fox, 2016), when the local production maximization inequality (Equation 5) is used as the basis for estimation, the equilibrium outcome of the matching model need not be estimated. Thus, the maximum score estimator method explained above may be considered as an equivalently consistent to estimate the production function for more than one independent markets or a single large market (Fox, 2016).

DISCUSSION

Comparative Analysis of Binary Choice and Matching Models in Partner Selection

Literature on relationship studies suggests that standard discrete choice models is appropriate in situations where the focal firm (or partners) makes choices with homogenous (or closely substitutable) partners, or even when the focal firm selects from a set of heterogenous partners, but such selection choices are mutual and unhindered by any rivalry (Fox, 2016).

Advantages and Disadvantages of Maximum Score Estimator-based Matching Models

Matching model methodology has a set of advantages and disadvantages. First, it is not atypical in strategic management research to have incomplete dataset about the sample firms, their characteristics, their partner preferences, their current partnerships and alike. In this respect, matching models' maximum score estimator methodology serves as an ideal technique because the method allows the identification of matches based on the observed characteristics

of each matched pair, and the estimated equilibrium condition of all paired matches in the market, without holding the traditional assumptions, such as linearity, normal distribution, heteroscedasticity and so on. Second, this matching score estimator provides fairly accurate inferences from the estimates, even with missing data of firms or matches to the extent that there are no systematic sample selection biases in the data creation process (Fox, 2010). Third, this method allows an unrigid specification of the joint production function that can contain multinomial expressions of higher order, continuous variable (e.g., firm size, firm age) or discrete variable (e.g., total number of patents owned), and even, match-specific (e.g., how far are the partners located or how similar are their resources/capabilities) that can be used to develop variables for both observed and swapped matches. Also, the function is not required to be defined from a global or submodular level in identifying matched pairs with certain attributes.

There are a few negatives too. First, since the inequality (equation 5) function forms the basis to cancel out the "main effects" of firm characteristics XFi and XNi, this restricts its estimation to only dyadic level and the groupings of partner attributes. Thus, its applicability is limited in estimating the magnitude of the match production function (e.g., efficiency, total contributions, value from participation and non-participation in the market for partners). However, "main effects" can be identified for situation where the data on self-matched firms exist or information on transfers are available (Akkus et al., 2015). Second, based on Equation 6, the estimates are seemingly fixed for the production function. Thus, it hinders to observe other relevant variables that may be contributing in the joint value creation (e.g., technological turbulence, market maturity) Third, since the estimator does not make any assumption, there is a possibility that the estimates may be inconsistent if the attributes of focal firms also interact in the production function with the same or fixed effects of non-focal firms of the market or with the match-specific errors homogenously (e.g., job skills).

As such, to achieve consistency in estimation, this matching estimator method requires (a) careful and conservative delineation of boundaries of the market(s) and appropriate assignment of firms (agents) are assigned to the appropriate side of the market (Fox, 2010), (b) Since the joint production function does not have a natural scale or level, scale normalization is done by setting one of the β estimates to be ±1. As a result, other relationships are interpreted based on the impact of the baseline relationship on the joint value, (c) This estimator method relies on numerical optimization for obtaining the estimates. Therefore, it is computationally intensive and requires thorough robustness checks to conclude on optimized results.

Applicability of the Maximum Score Estimator in Partner Selection

In using the maximum score methodology for studying matching models, the conditions of applicability (i.e., limiting conditions) contribute significantly for its successful application. This technique assumes that the observed matched partners are stable. This assumption is empirically valid for matching when other market factors do not influence their stability. In business model context, such factors include associated time and costs in searching appropriate partners, information asymmetry in potential partners' competences, bounded rationality of managers, agency costs. First, it is difficult to estimate the amount of time that the focal firm may be willing to wait for finding the most that would perfectly match in its business model. Atakan (2006) suggests that when firms search partners independently (i.e., involves search costs), the matching becomes preferential and biased. Shimer & Smith (2000) suggest that when observed matches have strong complementary resources, this method may downplay other attributes that may affect the stability of pairs. Therefore, such situations require dynamic

matching model considerations to explicitly include the influence of search costs in pair formation or collapse. Second, the quality of data restricts the influence of information asymmetries in the matching model. In business model context, agency costs (e.g., manager's self-interest) may mislead the firms to select partners that may be having lower complementarities than others in the market. Third, in partner formation, firms tend to choose partner based on third party referrals to reduce uncertainty, rather than going by the true logic of complementarity. This leads to preferential partner selections and such matches may be severely affected by adverse selection and possible moral hazards of partners. In short, the above statement market factors tend to make firms opt for preferential matching, and that would make the maximum score estimated results irrelevant.

NEXT STEPS AND CONCLUSION

The next step in this research is to explore past research further to learn whether using one versus another partner selection method changes substantive conclusions. To gain better understanding, a simulation of realistic conditions could provide meaningful evidence on its applicability in partnership situations of business models. Applying the maximum estimator model to real world business models would either helps us to better predict strategic partnering and/or identify synergies across multiple partners.

This study contributes significantly to the field of strategic management by examining a new method for partner selection that has been aptly applied in other fields. More specifically, this study helps in demonstrating the possibilities of an alternate method of partner selection for conditions that suggest an unbiased two-sided matching approach for selecting partners from the market for strategic reasons, i.e., achieving successful business model. The logic of "market of partners" exemplifies the uniqueness of this study that the strategic partnership literature severely lacks. In addition to it, this study also explains the reason for a less known firm, but higher complementariness can pair with an established firm. On the contrary, a similar (less known) firm with insignificant complementariness tends to compromise with a less established firm.

Perhaps, matching models can not only be applied for unbiased partner selection in business model context, but in other strategy related issues such as selection of targets in M&As, identifying buyer and supplier matches, investment related decisions, and top-executive selection processes. In sum, this study attempts to create academic interest among business model researchers to explore the purposefulness of matching models in partner selection activities of business models, and its empirical applicability in not just one-to-one cases, but also one-to-many and many-to-many cases of two-sided firm matching process of creating joint value.

REFERENCES

Ahuja, G. (2000). The duality of collaboration: Inducements and opportunities in the formation of interfirm linkages. *Strategic management journal*, *21*(3), 317-343.

Ahuja, G., Polidoro, F., & Mitchell, W. (2009). Structural homophily or social asymmetry? The formation of alliances by poorly embedded firms. *Strategic Management Journal*, *30*(9), 941-958.

Akkus, O. (2008). Endogenous Matching and the Determinants of Underpricing in IPO Markets.

Akkus, O., Cookson, J. A., & Hortacsu, A. (2015). The determinants of bank mergers: A revealed preference analysis. *Management Science*.

Atakan, A. E. (2006). Assortative matching with explicit search costs. *Econometrica*, *74*(3), 667-680.

Baccara, M. (2012). A field study on matching with network externalities. *The American Economic Review*, *102*(5), 1773-1804.

Banal-Estañol, A., Jofre-Bonet, M., & Lawson, C. (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. *Research Policy*, *44*(6), 1160-1175.

Becker, G. S. (1973). A theory of marriage: Part I. *The Journal of Political Economy*, 813-846.

Berry, S. T. (1992). Estimation of a Model of Entry in the Airline Industry. *Econometrica: Journal of the Econometric Society*, 889-917.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2003). *Analyzing the determinants of the matching public school teachers to jobs: Estimating compensating differentials in imperfect labor markets* (No. w9878). National Bureau of Economic Research.

Bresnahan, T. F., & Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 977-1009.

Chatain, O. (2011). Value creation, competition, and performance in buyer-supplier relationships. *Strategic Management Journal*, *32*(1), 76-102.

Chen, J., & Song, K. (2013). Two-sided matching in the loan market. *International Journal of Industrial Organization*, *31*(2), 145-152.

Chesbrough, H. (2007). Business model innovation: it's not just about technology anymore. *Strategy & leadership*, *35*(6), 12-17.

Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and corporate change*, *11*(3), 529-555.

Dagsvik, J. K. (2000). Aggregation in matching markets. *International Economic Review*, *41*(1), 27-58.

Fernando, C. S., Gatchev, V. A., & Spindt, P. A. (2005). Wanna dance? How firms and underwriters choose each other. *The Journal of Finance*, *60*(5), 2437-2469.

Fox, J. T. (2009). Structural empirical work using matching models. *New Palgrave Dictionary of Economics. Online edition*.

Fox, J. T. (2010). Identification in matching games. *Quantitative Economics*, *1*(2), 203-254.

Fox, J. T. (2016). Estimating matching games with transfers. *National Bureau of Economic Research.*

Fox, J. T., & Bajari, P. (2013). Measuring the efficiency of an FCC spectrum auction. *American Economic Journal: Microeconomics*, *5*(1), 100-146.

Gabaix, X., & Landier, A. (2006). *Why has CEO pay increased so much?* (No. w12365). National Bureau of Economic Research.

Heckman, J. J. (1977). Sample selection bias as a specification error (with an application to the estimation of labor supply functions).

Koopmans, T. C., & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society*, 53-76.

MacDonald, G., & Ryall, M. D. (2004). How do value creation and competition determine whether a firm appropriates value?. *Management Science*, *50*(10), 1319-1333.

Mackey, A., Molloy, J. C., & Morris, S. S. (2013). Scarce human capital in managerial labor markets. *Journal of Management*, 0149206313517265.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics*, *3*(3), 205-228.

Mindruta, D. (2013). Value creation in university-firm research collaborations: A matching approach. *Strategic Management Journal*, *34*(6), 644-665.

Mindruta, D., Moeen, M., & Agarwal, R. (2016). A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances. *Strategic Management Journal*, *37*(1), 206-231.

Mitsuhashi, H., & Greve, H. R. (2009). A matching theory of alliance formation and organizational success: Complementarity and compatibility. *Academy of Management Journal*, *52*(5), 975-995.

Mortensen, D. T. (1988). *Equilibrium wage distributions: a synthesis*. Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Milgrom, P., & Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of accounting and economics*, *19*(2), 179-208.

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT press.

Ostrovsky, M. (2008). Stability in supply chain networks. *The American Economic Review*, *98*(3), 897-923.

Ozcan, Y. (2015). *Innovation and Acquisition: Two'Sided Matching in M&A Markets*. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Pan, Y. (2015). The determinants and impact of executive-firm matches. *Management Science*.

Coleman, J. S. (1991). Matching processes in the labor market. *Acta Sociologica*, *34*(1), 3-12.

Pohle, G., & Chapman, M. (2006). IBM's global CEO report 2006: business model innovation matters. *Strategy & Leadership*, *34*(5), 34-40.

Roth AE, Sotomayor MA. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press: Cambridge, UK.

Shapley, L. S., & Shubik, M. (1972). *Game Theory in Economics: Chapter 3: the" rules of the Games"*. Rand.

Shimer, R., & Smith, L. (2000). Assortative matching and search. *Econometrica*, *68*(2), 343-369.

Sorensen, M. (2007). How smart is smart money? A two-sided matching model of Venture Capital. *The Journal of Finance*, *62*(6), 2725-2762.

Stuart, T. E. (1998). Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. *Administrative science quarterly*, 668-698.

Teece, D. J. (2010). Business models, business strategy and innovation. *Long range planning*, *43*(2), 172-194.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Terviö, M. (2008). The difference that CEOs make: An assignment model approach. *The American Economic Review*, *98*(3), 642-668.

Yang, Y., & Goldfarb, A. (2015). Banning controversial sponsors: understanding equilibrium outcomes when sports sponsorships are viewed as two-sided matches. *Journal of Marketing Research*, *52*(5), 593-615.

Yang, Y., Shi, M., & Goldfarb, A. (2009). Estimating the value of brand alliances in professional team sports. *Marketing Science*, *28*(6), 1095-1111.

Zott, C., & Amit, R. (2010). Business model design: an activity system perspective. *Long range planning*, *43*(2), 216-226.

DECISION SCIENCES INSTITUTE

Managing Supply Chain Career Portfolio in Digital Market Environment:
An Examination of Individual Level Perspectives

Ghadeer Alsanie
Department of Information, Operations and Technology Management
College of Business and Innovation
The University of Toledo
ghdeer.alsanie@rockets.utoledo.edu

Paul Hong
Department of Information, Operations and Technology Management
College of Business and Innovation
The University of Toledo
2801 W. Bancroft St.
Toledo, Ohio, USA 43606
paul.hong@utoledo.edu

Euisung Jung
Department of Information, Operations and Technology Management
College of Business and Innovation
The University of Toledo
Euisung.jung@utoledo.edu

ABSTRACT

In view of an increasing prospect of supply chain resource constraints, managing supply talent is an important strategic priority of major firms. In particular, changing global market reality and digital market, and innovative technologies pose both risks and opportunities for supply chain professionals. We present a research model that defines three major concepts--digital market environment, digital work requirements, and digital career portfolios--and their key indicators. Propositions explain the relationships between these key concepts. Lessons and implications are discussed for planning future fruitful research.

KEYWORDS: Digital Market Environment, Digital Work Requirements, Individual Level Perspectives, Supply Chain Career Portfolio, Research Model, Key Indicators, Constructs and Definition.

INTRODUCTION

As the world increasingly experiences the impact of digital market environment, digital work requirements change (Dou & Chou, 2002; Gandini, 2016; Schulze, 2019). The development of digital innovative technologies has drastically changed the nature and structure of work environment, presenting both risks and opportunities for working individuals. The potential for innovative technologies to change the world of work are literally unbounded ("Workforce of The Future," 2018). However, the effect of digital information technologies on work is still debated. A recent McKinsey report on jobs predicted that while up to 40% of human activities could be replaced by digital innovations in the future, only a 5% effect is expected with regard to jobs (Lund et al., 2019) However, the COVID-19 crisis has strongly affected labor markets, and it could take years for workers to get back to the normal pre-crisis level. Nonetheless, the pandemic will not be the only trend shaping the future of work; the constant changes in digital innovative technologies will continue to raise. Individual workers stand in the middle of this changing process being pressured to acquire new skills and competencies to remain valuable. Researchers in robotics and automation argue that these processes eventually lead to net job losses (Acemoglu & Restrepo, 2017; Arntz et al., 2017; Decanio, 2016). At the same time, it is expected that technological innovations will give rise to new occupations and growing demand in technology-related sectors (Autor, 2015; Smith & Anderson, 2014).

The perceived effect of digital innovative technologies on individual workers can be expressed as new technology, processes, and applications related to their fields of employment. The way workers respond to these changes will largely determine their success and future career prospects. Still, the mechanism of workers' response to the changing digital market environment is not yet fully understood. How could workers not get lost in this process and become prepared for changes in the world of work? The answer to this question is hardly predetermined. And yet, much research of the related kind is focusing on organizations that seek to attain competitive advantage by utilizing globalization trends and technological progress.

For this purpose, we explore three specific research questions: (1) Why is the digital market environment so different from the pre-digital market environment? (2) How would organizations redefine their digital work requirements in contrast to pre-digital work requirements? (3) What individuals need to reshape their digital careers differently from pre-digital career patterns? Our specific focus is on the supply chain career portfolio. Lessons and implications are discussed for future research.

LITERATURE REVIEW

Concept Development

Digital Market Environment

Digital Market Environment refers to the domestic and global competitive landscape in which people use of new digital technologies such as mobile, artificial intelligence, cloud, blockchain, and the Internet of things (IoT) technologies. Organizations are increasingly offering new and innovative information technology (IT) to create the workplace of the future by effectively facilitating digital work (Warner & Wäger, 2019; Horton & Tambe, 2015). The digital market

environment is reflected in the intensity of the domestic competitive environment, as well as the global competitive pressure and the accessibility of digital technologies, besides the escalating customer expectations and the changing governmental regulations. Disruptive innovation and demanding customers are making the digital market environment more unstable (Henriette et al., 2015) this creates the need to explore this environment and the necessity for reinventing the digital market environment.

Digital Work Requirements

Digital Work Requirements refers to the combined sets of cognitive, technological, and social skills and competences that firms expect from their primary and supportive workforce. These work requirements should be evaluated regularly to thrive in a competitive environment specifically in a digital market environment. Digital Work Requirements are indicated by the adoption of digital technologies, advancement in job qualifications, improvement and development of cognitive and social skills, and the advancement of technological competences.

Digital Career Portfolios

Digital Career Portfolios refers to Individual initiatives, practices, and actions in response to the competitive market environment and changing work requirements. Digital career portfolios are indicated by the ability to sense emerging work patterns, planning flexible work scenarios, and the ability to gain diverse work experiences, the capacity in expanding network relationships, and the quantity and quality of diverse skill competences. Changes in organizations are causing destruction in traditional and digital careers. Adapting to the uncertainties in the future of a career is hard for workers as they face many difficulties. To overcome these difficulties, workers are advised to take control over their careers by becoming more diverse in their skills, adapting to change, and be dynamic in shaping their career portfolio (Brousseau et al. 1996).

A summary of the concept development is provided in Table 1:

Table 1: Digital Market Environment, Digital Work Requirements, Digital Career Portfolio

Construct	Definition	Indicators	Literature Base
Digital Market Environment	Domestic and global competitive landscape in which people use of new digital technologies such as mobile, artificial intelligence, cloud, blockchain, and the Internet of things (IoT) technologies	▪ Domestic competitive intensity ▪ Global competitive pressure ▪ Availability of digital technologies ▪ Increasing customer expectations ▪ Changing governmental regulations	Warner & Wäger, 2019; Horton & Tambe ,2015; Dittes et al., 2019; Bordi et al., 2018; Warhurst, 2008; Yovanof & Hazapis, 2009; Yee, Bailenson, Urbanek, Chang & Merget, 2007; Dou & Chou, 2002.
Digital Work Requirements	Combined sets of cognitive, technological, and social skills and	▪ Adopting digital technologies ▪ Upgrading job qualifications ▪ Clarifying cognitive skills	Hong et al., 2019; Fuchs & Sevignani, 2013.; Grantham, 2000; Gandini, 2016; Scholz, 2012; Horton, Kerr &

	competences that firms expect from their primary and supportive workforce	▪ Training technological competences ▪ Developing social skills	Stanton, 2017; Bucher & Fieseler, 2017.
Digital Career Portfolios	Individual initiatives, practices, and actions in response to the competitive market environment and changing work requirements	▪ Sensing emerging work patterns ▪ Planning flexible work scenarios ▪ Gaining diverse work experiences ▪ Expanding network relationships ▪ Diversifying skill competences	Brousseau et al. 1996; Birkinshaw, 2018; Weiner, 2016; Lent, Brown & Hackett, 2002; Willis & Wilkie, 2009; Hooley, 2017; Eshet, 2004; Balliester & Elsheikhi, 2018; Barley et al., 2017

RESEARCH MODEL AND PROPOSITIONS

The rapid development and adoption of digital technologies dramatically changed the business environment, leading to digital work requirements and transformation of a career portfolio. This paper draws mainly on the Dynamic Capability theoretical perspective that emphasizes the positive changes in response to environmental challenges that should lead to value outcomes. From the perspective of individual responses to globalization and technological transformations, this would involve engaging in activities and developing competencies to better position oneself in the changing world of work. Teece et al. (1997) defined dynamic capabilities as the ability to reshape competencies in response to changing environments.

In order to examine how individuals reshape value chain career in a digital environment; the paper begins with identifying the dynamic factors that lead to reinventing the digital environment that leads to redesigning work requirements that in turn, determine reshaping value chain career. This paper will offer valuable insights into the skills and competencies of workers that could become essential to possess and develop in the nearby future; in order to prevent and mitigate the risk that they face in the changing nature of work. The general research model is presented in Figure 1 below:

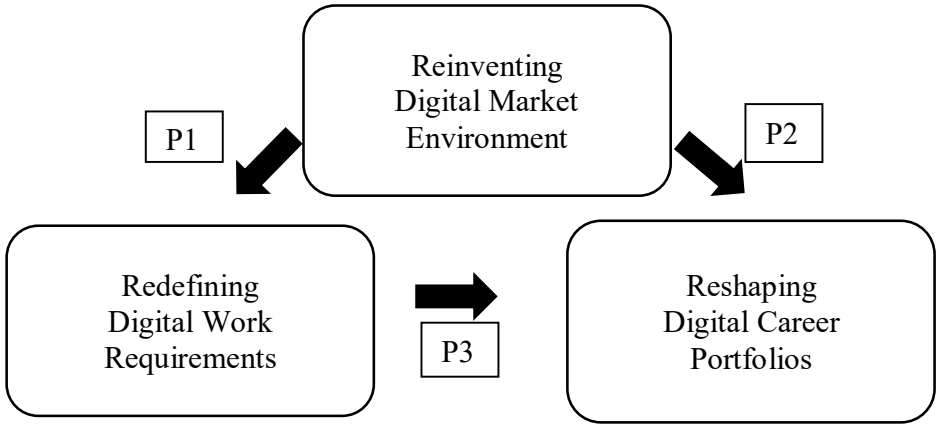


Figure 1: Reinventing Digital Market Environment, Redefining Digital Work Requirements, and Reshaping Value Chain Propositions

It is proposed that the digital revolution is taking place concurrently with and interrelated to broader socio-economic forces influencing the world of work. According to a recently released report by McKinsey, major transitions in the world of work will push 75 to 375 million of workers globally to change their occupations by 2030 (Lund et al., 2019). This brings an important implication for repositioning oneself by acquiring new skills and getting accustomed to new work practices.

Digital market environment and digital work requirements

In today's highly competitive digital environment, it is critical for firms to redefine their key business drivers and value chain process (Bordi et al., 2018; Connolly-Barker, 2018). Domestic competitive intensity and global competitive pressure necessitate firms to adopt digital technologies and upgrading job qualifications (Warner & Wäger, 2019; Horton & Tambe,2015; Hong et al., 2019; Fuchs & Sevignani, 2013). The increasing availability of digital technologies and rapidly changing customer expectations requires organizations to clarify the cognitive skills of their workforce and diversify training programs for equipping with technological competences (Dittes Richter, Richter, & Smolnik, 2019; Bordi et al., 2018; Gandini, 2016; Scholz, 2012; Horton, Kerr & Stanton, 2017). Changing governmental regulations expect firms to diversify organizational culture and thus developing social skills become more critical for greater collaborative work in harmony (Warhurst, 2008; Dittes ., 2019; Yee et al., 2007; Lent et al., 2002) To enable this transformation, firms need to redefine its work process followed by work requirement. Thus, drawing from this relationship, we propose,

P1. The more digital market environment reinvents its competitive realities, the more firms redefine work requirements.

Digital market environment and digital career portfolio

To survive in the digital market environment, it is essential to understand emerging digital tools and technologies which allow firms to reduce costs, increase productivity, improve new product development, achieve a faster market response, and enhance customer focus (Balliester & Elsheikhi, 2018; Hong et al., 2019; Pagani, 2013). The intensity of the domestic competitive environment, as well as the global competitive pressure, require firms and workers to plan flexible work scenarios and to keep sensing emerging work patterns (Birkinshaw, 2018; Henriette et al., 2015; Barley et al., 2017; Burke & Ng, 2006; Connolly-Barker, 2018). The accessibility of digital technologies besides the escalating customer expectations requires workers to gain diverse work experiences, and expand their network relationships (Rothwell et al., 2015; Davenport & Kirby, 2015; Hooley, 2017; Willis & Wilkie, 2009). Changes in governmental regulations require workers for quantity and quality of diverse skill competences (Warhurst, 2008; Yovanof & Hazapis, 2009). For future careers, individuals should understand the new digital value chain process with digital technology knowledge. Thus, building on this perspective, we propose,

P2. The more digital market environment reinvents its competitive realities, the more individuals reshape their digital career portfolios.

Digital work requirements and digital career portfolio

Diverse digital work requirements can be employed by firms for enhancing each stage of their value chain. The digital work requirement reflects a particular basket of new digital industrial technologies. Thus, individuals need to reshape their career portfolios by different digital work requirements. Adoption of digital technologies and the advancement in job qualifications support workers in sensing emerging work patterns and planning flexible work scenarios (Connolly-Barker, 2018; Lazazzara et al., 2020; Bondarouk & Ruël, 2009; Hooley, 2017; Grantham, 2000) Improvement and development of cognitive and social skills, and the advancement of technological competences contribute to gain diverse work experiences and capacity in expanding network relationships and diversifying skill competences (Willis & Wilkie, 2009; Eshet,2004; Bucher & Fieseler, 2017). Thus, deducing this causal relationship perspective, we propose,

P3. The more firms redefine their digital work requirements, the more individuals reshape their digital career portfolios.

IMPLICATIONS FOR FUTURE RESEARCH

Coming crisis in supply chain talent management

Supply chain talent management is becoming one of the key strategic challenges of firms in major industries to the extent that there might be a serious supply chain talent crisis (Cottrill, 2010; Dubey and Gunasekaran, 2015). The Supply chain profession is going through a rapid transition anticipating a large number of senior managers to retire within the next 5-10 years and education and training of vital supply chain talents are of vital importance (Leon & Uddin, 2016; McCrea, 2019).

Even large global firms are not yet ready to invest heavily to respond to massive changes in the digital market environment. Most firms take an incremental and cautious approach—"tortoise, not the hare" approach (Hartley and Sawaya, 2019). Digital transformation occurs slowly in "discovery-driven", not in delivery-directed (McGrath and McManus, 2020). Therefore, responses of educational institutions and individual responses are also much more reactive than proactive with a great deal of uncertainty of the future world of work. In this context, the coming supply chain talent crisis might not be an impossible scenario.

Fruitful research opportunities from multiple perspectives

A key macro-level changing factor is technological in general and digital market environment in particular. This paper examines issues related to the digital market environment, digital work requirements, and digital supply chain career portfolio. Managing a supply chain career portfolio demands a clear understanding of the digital market environment and changing digital work requirements. Our research model defines the essential dimensions of these three primary concepts and presents key indicators. Three propositions will be a foundation of future hypothesis development to test causal relationships which provide an important basis for future research that uses a variety of research methods. Follow-up research may involve case studies, pilot team projects, and large-scale surveys. Examining digital market trends, digital work contexts, and individual career scenario planning may further involve exploring big data as well as primary survey data. This will open fruitful research opportunities for supporting the efforts of effective supply chain talent management.

References

Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economic Letters, 159*, 157-160.

Autor, D.H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3-30.

Balliester, T., & Elsheikhi, A. (2018). *The future of work: a literature review*. International Labor Office, Working Paper 29.

Barley, S.R., Bechky, B.A., & Milliken, F.J. (2017). The changing nature of work: careers, identities, and work lives in the 21st century. *Academy of Management Discoveries, 3*(2), 111-115

Birkinshaw, J. (2018). How is technological change affecting the nature of the corporation? *Journal of the British Academy, 6*(s1), 185-214.

Bondarouk, T. V., & Ruël, H. J. (2009). Electronic Human Resource Management: challenges in the digital era. *The International Journal of Human Resource Management*, *20*(3), 505-514.

Bordi, L., Okkonen, J., Mäkiniemi, J. P., & Heikkilä-Tammi, K. (2018). Communication in the digital work environment: implications for wellbeing at work.

Brousseau, K. R., Driver, M. J., Eneroth, K., & Larson, R. (1996). Career pandemonium: Realigning organizations and individuals. *Academy of Management Perspectives*, *10*(4), 52-66.

Brown, J., Gosling, T., Sethi, B., Sheppard, B., Stubbings, C., Sviokla, J., & Zarubina, D. (2017). Workforce of the future: The competing forces shaping 2030. *London: PWC*.

Bucher, E., & Fieseler, C. (2017). The flow of digital labor. New media & society, 19(11), 1868-1886.

Burke, R.J., & Ng, E. (2006). The changing nature of work and organizations: implications for human resource management. *Human Resource Management Review, 16*, 86-94.

Connolly-Barker, M. (2018). Advanced automation technology, labor market insecurity, and collective joblessness: the determinants, constraints and employment effects of robots and artificial intelligence on the realm of work. *Psychosociological Issues in Human Resource Management, 6*(2), 92-97.

Cottrill, K. (2010). Are you prepared for the supply chain talent crisis. *MIT Center for Transportation and Logistics, Cambridge, MA*, 1-11.

Davenport, T., & Kirby, J. (2015). *Beyond automation*. Cambridge, MA: Harvard Business Review.

Dittes, S., Richter, S., Richter, A., & Smolnik, S. (2019). Toward the workplace of the future: How organizations can facilitate digital work. *Business Horizons*, *62*(5), 649-661.

Dou, W., & Chou, D. C. (2002). A structural analysis of business-to-business digital markets. Industrial marketing management, 31(2), 165-176.

Dubey, R. and Gunasekaran, A. (2015), "Supply chain talent: the missing link in supply chain strategy", *Industrial and Commercial Training*, Vol. 47 No. 5, pp. 257-264.

Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of educational multimedia and hypermedia*, *13*(1), 93-106.

Fuchs, C., & Sevignani, S. (2013). What is digital labour? What is digital work? What's their difference? And why do these questions matter for understanding social media?. *TripleC: Communication, capitalism & critique. Open access journal for a global sustainable information society*, *11*(2), 237-293.

Gandini, A. (2016). Digital work: Self-branding and social capital in the freelance knowledge economy. Marketing theory, 16(1), 123-141.

Gilliland-Swetland, A. J. (2000). *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment*. Council on Library and Information Resources, 1755 Massachusetts Ave. NW, Suite 500, Washington, DC 20036.

Grantham, C. E. (2000). The future of work: The promise of the new digital work society. New York: McGraw-Hill.

Hartley, J.L. & Sawaya, W.J. (2019). Tortoise, not the hare: Digital transformation of supply chain business processes. *Business Horizons,* 62 (6), 707-715.

Henriette, E., Feki, M., & Boughzala, I. (2015). The shape of digital transformation: a systematic literature review. *MCIS 2015 proceedings*, 431-443.

Hong, P. C., Kallarakal, T. K., Moina, M., & Hopkins, M. (2019). Managing change, growth and transformation. Journal of Management Development.

Hooley, T. (2017). Developing your career: harnessing the power of the internet for "digital career management". *Development and Learning in Organizations: An International Journal*.

Horton, J., Kerr, W. R., & Stanton, C. (2017). Digital labor markets and global talent flows. In High-skilled migration to the United States and its economic consequences (pp. 71-108). University of Chicago Press.

Horton, J. J., & Tambe, P. (2015). Labor economists get their microscope: big data and labor market analysis. *Big data*, *3*(3), 130-137.

Lazazzara, A., Tims, M., & De Gennaro, D. (2020). The process of reinventing a job: A meta–synthesis of qualitative job crafting research. *Journal of Vocational Behavior*, *116*, 103267.

Lent, R. W. (2018). Future of work in the digital world: Preparing for instability and opportunity. *The Career Development Quarterly*, *66*(3), 205-219.

Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career choice and development*, *4*, 255-311.

Leon, S., & Uddin, N. (2016). Finding supply chain talent: an outreach strategy. *Supply Chain Management: An International Journal,* Vol. 21 No. 1, pp. 20-44

Lund, S., Manyika, J., Segel, L. H., Dua, A., Hancock, B., Rutherford, S., & Macon, B. (2019). *The future of work in America: People and places, today and tomorrow*. McKinsey Global Institute.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ... & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation. *McKinsey Global Institute*, *150*.

McCrea, B. (2019). What's your worth? *Supply Chain Management Review*. 23(7), 18-23.

McGrath, R., & McManus, R (2020). Discovery-Driven Digital Transformation. *Harvard Business Review*. 98 (3), 124-133.

Pagani, M. (2013). Digital business strategy and value creation: Framing the dynamic cycle of control points. *Mis Quarterly*, 617-632.

Rothwell, W. J., Jackson, R. D., Ressler, C. L., Jones, M. C., & Brower, M. (2015). *Career Planning and Succession Management: Developing Your Organization's Talent—for Today and Tomorrow: Developing Your Organization's Talent—for Today and Tomorrow*. ABC-CLIO.

Scholz, T. (Ed.). (2012). Digital labor: The Internet as playground and factory. Routledge.

Teece, D., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7): 509–533.

Warhurst, C. (2008). The knowledge economy, skills and government labour market intervention. *Policy Studies*, *29*(1), 71-86.

Warner, K. S., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning*, *52*(3), 326-349.

Weiner, W. (2016). The influence of social media on job seekers in the digital age. *Career Planning and Adult Development Journal, 32*(3), 26-32.

Willis, L., & Wilkie, L. (2009). Digital career portfolios: Expanding institutional opportunities. *Journal of Employment Counseling*, *46*(2), 73-81.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior*, *10*(1), 115-121.

Yovanof, G. S., & Hazapis, G. N. (2009). An architectural framework and enabling wireless technologies for digital cities & intelligent urban environments. *Wireless personal communications*, *49*(3), 445-463.

DECISION SCIENCES INSTITUTE

How might the use of Blockchain Technology affect supply chain risks and their management?

Meike Schroeder
Hamburg University of Technology
Institute of Business Logistics and General Management
Am Schwarzenberg-Campus 4
21073 Hamburg, Germany
+49(0)40428784384
Email: meike.schroeder@tuhh.de

Sven Reimers
Hamburg University of Technology
Institute of Business Logistics and General Management
Am Schwarzenberg-Campus 4
21073 Hamburg, Germany
+49(0)40428784526
Email: sven.reimers@tuhh.de

ABSTRACT

Transparency in the supply chain is essential to maintain efficient supply chain risk management (SCRM). Blockchain technology can be applied to improve visibility in supply chains. But with increased transparency, SCRM is faced with new challenges. In this article, the application of blockchain technologies' effects on supply chain risks are analyzed. For this purpose, a fictitious case is created in which blockchain technology is applied to a manufacturer of medical technology products' supply chain. The findings of the paper include (1) the effects on supply chain risks (2) recommendations for the implementation of future SCRM in blockchain.

KEYWORDS: Supply Chain Risks, Transparency, Blockchain, Agency Theory, Use Case

INTRODUCTION

The push for ever more efficient supply chains has a decisive influence on the individual supply chains' competitiveness (Chang et al., 2019; Manuj and Mentzer, 2008; Tang, 2006). Blockchain has, therefore, attracted increasing attention both in research and in practice in recent years.
The use of blockchain technology can assure ownership tracking and traceability across the supply chain. It would also improve the availability and dissemination of information, thus, increasing the transparency for all supply chain partners involved (Azzi et al., 2019; Swan, 2018; Vinay Reddy, 2019). The use of blockchain technology, therefore, shows considerable potential for the future improvement of cooperation within supply chains. Since the blockchain research field is still quite young, there are only a limited number of scientific publications. They describe the technical implementation and the benefits and barriers (e.g. scalability, privacy, and governance) associated with the implementation of blockchain technology (Lee and Pilkington, 2017; Petersen et al., 2018; Nakamoto, 2008; Tönnissen and Teuteberg, 2019; Yang, 2019). Further papers deal with practical examples, such as the use of blockchain technology by British Airways, IBM and Maersk, Wal-Mart, UPS, Provenance, or Carrefour (Awwad et al., 2018; Chang et al., 2019; Jabbar and Bjørn, 2018), in which goods tracing, transparency measures, goods documentation, security improvements, and counterfeit prevention are analyzed.

Some articles also discuss individual risks such as technical information risks, risks of fraudulent food recalls fraud, and product losses (Bumblauskas et al., 2020; Fridgen et al., 2019; Fu and Zhu, 2019; Gao et al., 2019; Kumar et al., 2020; Montecchi et al., 2019).
However, none of the scientific publications has yet provided a holistic view of Supply Chain Risk Management (SCRM) in regards to a given technology. Holistic means that – in addition to internal risks – supply, demand, and environmental risks are also considered in the analysis. Furthermore, the consideration includes the individual process steps of SCRM, such as risk identification, analysis, assessment, handling, and control (Rao and Goldsby, 2009; Schröder, 2019).
While different aspects of SCRM, as well as the application of blockchain technology, have, particularly in the last two years, been the subject of scientific discussions, merging the two research areas remains relatively unexplored. Thus, a first attempt is made with this paper, addressing the following research questions:

1) How might the application of blockchain technology affect supply chain risks?
2) What recommendations result from this for future SCRM?

The paper is organized as follows: first we give a brief introduction to the topics of SCRM and blockchain. Second, a description of the methodological approach follows; relating the extensive literature review, which was applied, using a snowballing method. Well-described blockchain use cases were analyzed to answer a given research questions. Since there are only a few cases about practical blockchain applications, there exists currently no paper that has a level of detail, which could be used to analyze the topic of SCRM holistically. Relevant information that is needed for an analysis of a holistic SCRM approach is missing. Therefore, a fictitious use case is described in this paper, applying blockchain technology to a manufacturer's supply chain of medical technology products. Based on this case, blockchain technologies' effects on supply chain risks can be analyzed, using the categorization of the risks developed by Christopher and Peck (2004).
After all these categories are analyzed, recommendations for an SCRM are given, before the close of the article with summary and limitations.

SUPPLY CHAIN RISK MANAGEMENT

Although the term "supply chain" has been discussed in both academia and the market for several decades, there is still no consistent understanding of what it means. Lambert et al. (1998, p. 504) defines "supply chain" as *"the alignment of firms that bring products or services to market […]",* whereas Mentzer et al. (2001, p. 4) understand supply chains to be *"a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer"*. This paper is based on the definition from Christopher (2005), which is widely used in scientific literature. Accordingly, a supply chain is *"the network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hands of the ultimate consumer"* (Christopher, 2005, p. 17).
Trends like globalization or increasing digitalization force companies to collaborate more closely with supply chain partners distributed worldwide, and, consequently, to optimize their exchange of material, information, and capital to reduce costs and enhance customer value.
International supply chain integration not only increases dependency between the partners but also increases the vulnerability to risks in the entire network (Christopher and Peck, 2004; Giunipero and Aly Eltantawy, 2004). Major events, such as natural disasters or the spread of the coronavirus (COVID-19), have resulted in complete supply chain interruptions in the past. However, due to the "domino effect", these supply chain interruptions no longer affect just one

single company, but several, so that financial damages run through the entire supply chain (Jüttner et al., 2003; Manuj and Mentzer, 2008). For this reason, the management of supply chain risks should increase focus on decision-makers.
SCRM can be understood as "*a part of Supply Chain Management which contains all strategies and measures, all knowledge, all institutions, all processes, and all technologies, which can be used on the technical, personal, and organizational level to reduce supply chain risks."* (Kersten et al. 2011, p. 157). The main objective of SCRM is to increase the transparency and robustness of processes to withstand any kind of supply chain disruptions (Tang, 2006).
The typical SCRM process encompasses four steps: risk identification, analysis and assessment, handling (also called mitigation), and control. This paper focuses on the risk identification step. Therefore, this step will be described in more detail. In the first step (identification), potential risks are identified and classified. The range of the risk landscape to be identified plays an important role here. For example, sometimes either only the risks within a company or risks that occur within the entire supply chain are identified (Norrman and Lindroth, 2004). Therefore, it must be decided whether and how many supply chain partners should be included in this step of the risk identification process. Here, direct suppliers or customers (tier 1) can be integrated into the process of risk identification, or even expanded to include the first suppliers or end customers (tier 2 to n) (Kersten et al., 2011).
As part of risk identification, supply chain risks can be grouped according to different classifications (Ritchie and Brindley, 2004; Schröder, 2019). This article is based on the approach of Christopher and Peck (2004). They classify supply chain risks into three groups: company internal risks, supply chain internal risks, and environmental risks (Christopher and Peck, 2004).
Internal risks emphasize process and control risks within the company. While process risks relate to internal disruptions and to the value-adding and managerial activities of the company. Control risks arise from the misapplication of rules and procedures about how supply chain processes, like order quantities or batch sizes, are controlled. Supply chain risks consider both supply and demand risks generated by supply chain partners.
While supply chain risks relate to the flow of goods and information within the network, up the supply stream of the focal company, demand risks relate to the flow of goods and information between said focal company and the market. Environmental risks cover all risks caused by socio-political, macroeconomic, or natural disasters (Christopher, 2005; Christopher and Peck, 2004).

In business literature, there are a variety of explanatory approaches with different origins that are used to explain the existence and design of SCRM. For the theoretical framework of this paper, the principal agent theory is used. It provides an insight into the consequences that can result from the asymmetric distribution of information between the so-called “Principal” and the other operating “agents” (Ross, 1973.). A principal-agent relationship exists, for example, in a supply chain between the manufacturing company and its suppliers and customers. The principal client (e.g. the manufacturing company in focus) makes decisions that influence not only his own outcomes, but also the potential benefits of the agents (supplier and buyer in the supply chain). The principal has imperfect information about the occurrence of certain environmental conditions as well as about the agents’ behaviour. This in turn provides room for the agents to take more opportunistic action, since the agent usually has an information advantage in his field of activity (Ross, 1973; Stiglitz, 2002).
Principal-agent relationships can be classified according to the types of information asymmetry. Three different situations, leading to this asymmetry, can be distinguished (Stiglitz, 2002; Kleinaltenkamp and Jacob, 2002): The first problem which can occur is "hidden characteristics". This occurs prior to the conclusion of a contract (ex-ante), where the principal does not fully understand characteristics of or the services offered by the agent. If the agent does not disclose his capabilities, there is a risk that the Principal will select an unsuitable contractual partner (adverse selection).

The second concern is that of "hidden action" or "hidden information", which only occurs after the contract has been concluded. In this case the agent reduces his efforts and withholds information to his advantage. The principal can therefore only assess the result, but not the behaviour of the agent. Consequently, he cannot distinguish whether the agent is responsible for a poor result, or whether unfavorable environmental influences are responsible (Ross, 1973; Stiglitz, 2002).
In the final situation, "hidden intention", the principal has already made specific investments and advance payments ("sunk costs"). As a result of these investments, the principal becomes dependent on the agent because he is dependent on the agent's output (Kleinaltenkamp and Jacob, 2002).
In summary, the principal-agent approach helps to explain the relationships within the supply chain as well as the existence of SCRM in the company and supply chains. According to the principal-agent approach, there is the possibility of direct behavioural control of the agent by the principal, which can be achieved by improving the exchange of information thereby increasing transparency. This will be demonstrated below using the example of blockchain technology implementation.

BLOCKCHAIN IN SUPPLY CHAIN MANAGEMENT

The first, reliable public approach was presented by Satoshi Nakatmoto (although this identity is not verified) in October 2008. In his explanation, he presented a solution to the "double spending" problem for a digital currency and proposed a simple implementation called "Bitcoin". (Nakamoto, 2008). The blockchain technology behind Bitcoin is a distributed ledger of transactions that cannot be manipulated. This technology is tamper-proof via a series of cryptographic methods, whereby, no intermediaries participate in the activities (Lee and Pilkington, 2017). Blockchain technology has three basic characteristics: it is decentralized, verified, and immutable (Petersen et al., 2018). Decentralization is achieved through the peer-to-peer network operated by the members. Verification is achieved by requiring members to cryptographically sign their transactions before sharing them with the network. The remaining characteristic, immutability, is created with the help of the so-called consensus algorithm (Nakamoto, 2008). There are various implementations of blockchain technologies which can be categorized into public and private settings. The public blockchain implementations, like Bitcoin or Stellar, are open-accessible for everyone. On the contrary, private blockchains, like the Hyperledger-Fabric implementation, can only be joined by meeting the entrance regulations of the network operators (Swan, 2018).
There are many theories and practical suggestions regarding the proper application of blockchain in logistics and supply chain management. "Common use cases" are identified to track goods along the supply chain, for example, to improve tuna traceability in order to combat illegal fishing, or the use of blockchain to enable the traceability of wood, from the tree to the end customer in order to certify the origin of the material (Figorilli et al., 2018; Lin et al., 2019; Tönnissen and Teuteberg, 2019; Yang, 2019). To realize use cases like track and trace, "smart contracts" are necessary. A "smart contract" is code that facilitates, modifies or enforces the execution or action of a digital contract. To add a smart contract into a network, a trusted party should create said contract and add it to a network, so the network can execute it without relying on a third party (Swan, 2018). For these reasons introducing blockchain technology can ultimately create a new database of information, which can also be applied in other areas, such as the SCRM (Swan, 2018; Tang and Veelenturf, 2019).

METHODOLOGY

We conducted a literature review, using a snowballing approach (Wohlin, 2014) to answer our research questions, listed in the introduction. A set of internationally renowned research papers about blockchain was assembled and then evaluated using keywords (("Blockchain" OR

"distributed ledger") AND ("supply chain" OR "logistics")). For context, Scopus, Web of Science, and IEEE Xplore were the primary sources. The paper collection and analysis took place between January and February 2020. The initial set of papers included several different publishers, such as the Journal of Purchasing and Supply Management and the International Journal of Production Economics, using IEEE Access in the timeframe of 2016-2020. In total, 386 peer-reviewed relevant papers, with a focus on supply chains and/or logistics, were identified. Blockchain publications with no clear connection to risk management in the logistics and supply chain management field were excluded.
Additionally, the literature research was extended by searching for well-described blockchain use cases to analyze them in order to answer the first research question. Several use cases in the literature were identified: blockchain implementation at British Airways, IBM and Maersk, Wal-Mart, UPS, Provenance, and Carrefour (Awwad et al., 2018; Chang et al., 2019; Jabbar and Bjørn, 2018). However, none of these were suitable for this research project. In most of the cases, the handling was described only very superficially, hence, relevant information, which is needed for an analysis of a holistic SCRM approach, is missing. Therefore, a fictitious use case, applying blockchain technology to a manufacturer's supply chain of medical technology products is described.

THE USE OF BLOCKCHAIN TECHNOLOGY FROM AN SCRM PERSPECTIVE

Results of the literature analysis

Current literature reviews on blockchain in supply chain and logistics concentrate on specific industries or general discussions on where to use it (Awwad et al., 2018; Bermeo-Almeida et al., 2018; Casino et al., 2019; Czachorowski et al., 2019). Table 1 contains a summary of relevant work in the area of risk management in supply chain and logistics. Only articles that deal with the topic of risk management are listed here in an attempt to further develop the theory. Excluded are all articles that address the use case in general, but which do not address the theory. The articles listed are, believed to be, central articles describing the topic, although completeness cannot be guaranteed.
Taken together only a few articles focus on the general approach to supply chain risk management. The majority of the articles focus on specific areas of supply chain management, such as supply chain finance or logistical functions, like the tracking of goods. Even in the overview shown below, the articles are very focused on specific use cases like supply chain finance. The generalist approach to this new technology in SCRM is usually left out. Therefore, the focus of this article is the influence of the new technology on SCRM and how risks need to be reassessed if this technology is to be standardized.

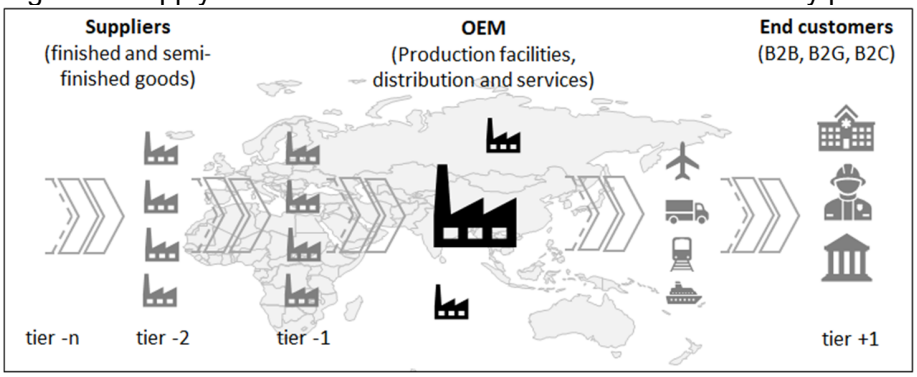
Article	Key Findings
Barczak et al. 2019	The authors "analyze the strength of correlation between digital technology [among them blockchain] and different risk types [macro environment, operational, functional and microenvironment risks]".
Fridgen et al., 2019	The authors "discuss the potential contribution of Blockchain technology to systematic risk management in global supply chains and networks."
Fu and Zhu, 2019	The authors "make big production enterprise as the research object to analyze by constructing a model its endogenous risk management mechanism, and analyze the management mechanism produced economic value."

Gao et al., 2019	The authors "explore Supply Chain Financing pledge risk controlling taking advantage of Blockchain technology, which is known for information transparency and tamper-proof, and proposes a strategy of Real-Time Stare in Market to mitigate risk pressure brought by a pledge of movables."
Liu et al., 2019	"Based on the influence of Blockchain technology on information sharing among supply chain participants, [the authors] use mean-CVaR (conditional value at risk) to characterize retailers' risk aversion behavior, while a Stackelberg game is taken to study the optimal decision-making of manufacturers and retailers during decentralized and centralized decision-making processes."
Montecchi et al., 2019	The authors develop "a provenance knowledge framework and show how its application can enhance assurances and reduce perceived risks via the application of Blockchain."
Zhao, 2019	"The paper proposes the mega-project supply chains infectious risks (MPSCIRs) synergy theory, defines the risk propagation speed, propagation density, and risk interconnectivity, distinguishes conventional risks from infectious risks, and identifies the risks facing major projects when passing through the "secret channel" of supply chains based on system dynamics theory and stability theory, and combining the new characteristics of major project supply chain risks after new technology reforms."
Zheng et al., 2019	The authors study" the risk decision-making problem faced by participants in a spacecraft supply chain, considering the adoption of the Blockchain technology to facilitate information sharing."
Zhu and Kouhizadeh, 2019	The authors describe how "Blockchain technology provides traceability, transparency, security, accuracy, and smart execution, which can all contribute to the product deletion and rationalization decision. Application recommendations and managerial insights into product deletion decision-making processes with Blockchain technology are provided."

The Use Case: Blockchain Technology for a Medical and Safety Products Manufacturer

The fictitious use case describes the supply chain of a leading international company in the fields of medical and safety technology (Figure 1). With more than 10.000 employees worldwide and distribution and services companies in over 40 countries, the OEM manufactures a large variety of medical and safety products in India, the UK, the Czech Republic, South Africa, China, and the USA. The manufacturer is known for products, like ventilators, for intensive and emergency care items (gas measuring technology), and for personal protection equipment that is produced for use in clinical, industrial, firefighting, and rescue services.

Figure 1: Supply chain of the manufacturer of medical and safety products



The industry of medical technology is subject to high design and production requirements, which results from numerous legal standards for manufacturers and importers regarding the quality and traceability of medical devices (see, e.g. Medical Production law, EU Guideline 67/548/EWG,

FDA's Regulation of Medical Devices). The legal requirements require each part used for the manufacturing of the medical and safety products to be documented. To make this possible, the OEM needs access to information about serial and batch numbers, as well as about certifications from suppliers, production facilities, and distribution partners of the entire supply chain. Here it is often difficult to gain visibility within the supply chain as the OEM usually has limited direct access to information on their direct suppliers (tier -1 respectively) or on their direct customers (tier +1) (Stevenson and Spring 2007; Lambert et al. 1998). Changes in documents and products must be traceable and tamper-proof, if product charges have to be recalled.

As described before, the use of blockchain technology helps to increase transparency in the supply chain. When increasing transparency using blockchain technology, new opportunities like the tracking of goods and emissions or even mapping the supply chain are made available. For this reason, in the use case, the OEM decided to implement a private-access blockchain. The company and transport symbols presented in Figure 1 illustrate the nodes of the blockchain. This includes both the OEM and supplier subgroups. The end customers within the supply chain are not included, as they bring too much uncertainty into supply chain planning. The blockchain presented here is a private blockchain in which the tracking data of the individual products are stored. With the help of the blockchain, all data on the sub-products are tracked, thereby preventing information asymmetries and gaps.

Effects of Blockchain Technology on Different Risks Categories

In the following section we describe how blockchain technology can improve SCRM by affecting different kinds of risks. To structure our procedure, we rely on the risk categorization system developed by Christopher and Peck (as described before; 2004), which divides risks according to their origin, into internal – containing process and control risks – supply chain, demand, and environmental risks.

Process and Control Risks

As described before, the asymmetrical distribution of information between the primary company, the suppliers, and the customers can be a major problem in the process flow (Stiglitz, 2002).

The principal – in our case the OEM – is often not able to track the activities of the agents – in this case, the suppliers. But within blockchain technology, the participants apply smart contracts, as explained in the blockchain section. Any supply chain partner using such a contract posts all applicable transactions to the blockchain – for example, when the products are transferred to a logistics service provider and loaded onto a truck. By using smart contracts, the blockchain can operate in a fully autonomous and decentralized manner and replace the need for human intervention (Bocek et al., 2017; Zhao, 2019). The information in the blockchain is reliable due to the security of the distributed ledgers. None of the partners can modify, append or delete any transaction without the others noticing. As a result, all handling steps are verifiable and traceable, and data quality risks will be reduced (Chang et al., 2019).

As defined by Christopher and Peck (2004), control risks might arise from the application or misapplication of pre-defined rules. By using blockchain technology, the handling process ensures that all supply chain partners meet the defined standards and that problems of confidence or concerns of manipulation are reduced and limited to the data entry phase (Chang et al., 2019). Therefore, using blockchain technology decreases control risks.

Furthermore, applying blockchain technology reduces the risks associated with information transfer and those associated with accrued during the overall handling process (Azzi et al. 2019;

Swan, 2018). Multiple parties, like shippers and carriers, logistics services providers, distribution and service companies, insurance providers, and customs agents, are involved in the transport process of medical and safety products, and they all need specific information at particular times. By using blockchain technology, each supply chain and service partner can view the progress of goods through the entire supply chain and track where products are in transit. Furthermore, these stakeholders can view bills of lading and see the status of customs documents, so that the transparency from end-to-end can be enhanced. For example, when the shipped semi-finished products from China arrive at the seaport in Germany, both the terminal officials and the freight forwarder must be provided with the same data. If the logistics service provider revokes a release order, the blockchain technology enables the distribution of this information in real-time to all actors involved. The data is replicated on the terminal note and also transmitted to all other notes (Hackius et al., 2019). Also, equal and simultaneous data access allows for faster customs clearances as the correct data is in the right place at the right time (Plowman, 1964). This level of transparency helps to reduce errors in the process execution and increases the speed of transit as well as the shipping process (Chang et al., 2019).

From an SCRM perspective, the composition of the blockchain network, in connection with the distribution of power and defined standards, should also be considered to reduce process risks. As of the publication of this document, uniform blockchain standards that govern the processing of transactions are still missing. In consequence, use of case-specific standards must be set (Lacity, 2018). But how and by whom should the blockchain network be controlled and the necessary standards defined?

In the use case, the manufacturer of medical and safety products is a powerful actor. They handle the private network by defining the standards and administrating smart contracts to reduce process and control risks. The OEM determines who gets access rights to the blockchain technology. For that, they must convince all supply chain partners, including all distribution and services companies from India, UK, Czech Republic, South Africa, China, and the USA, to take part in the blockchain. Implementing blockchain technology requires financial, personnel, and technical resources, as well as the ubiquitous use of the smart contracts by all partners (Montecchi et al. 2019). Convincing partners is particularly challenging when they are involved in multiple supply chains which do not use blockchain technology (Wust and Gervais, 2018). On the one hand, the OEM can motivate the supply chain partners to participate – e.g. by offering web interfaces or maintenance services. On the other hand, they could exploit their market power by terminating the collaboration if the supplier does not agree to accept these defined standards and use the blockchain technology (Corkery, Michael and Nathaniel Popper, 2018; Nikhilesh De, 2018). Here, the typical discussion around bargaining powers between buyers and sellers is explored (Scalco et al., 2017; Wolff, 2016). However, in the best case, after the transition to the new infrastructure, all supply chain partners will be delivering regular benefits. At a minimum, a common understanding of the system landscape is necessary to guarantee an efficient process flow (Christopher, 2005).

When defining the blockchain standards, the OEM must consider country-specific regulatory requirements (e.g. the FDA for medical devices in the USA) to minimize process risks. At the same time, supply chain partners might be afraid of information risks and, therefore, may still hesitate to share confidential data (Fridgen et al., 2019). Nonetheless, each partner must disclose some of their confidential information when participating in a blockchain so that the quality of the data collected ultimately leads to added value (Swan, 2018).

Supply and Demand Risks

Unforeseeable disruptions, which cannot be identified in time and consequently lead to an interruption of the supply chain, are the biggest challenges in supply chain management (Jüttner et al. 2003; Manuj and Mentzer, 2008). Often the causes cannot be identified and thus the consequences for upstream supply chain partners cannot be realistically estimated. By using blockchain technology, transparency can be increased (Azzi et al., 2019; Swan, 2018; Vinay Reddy, 2019). Supply chain partners can reliably track goods along the entire supply chain (Awwad et al., 2018; Chang et al., 2019; Jabbar and Bjørn, 2018). Critical data points along the supply chain can be accurately collected so that information related to the medical and safety products can be continuously updated. Blockchain technology allows data to be linked directly to all the OEM logistics services providers and customers. Based on real-time data, it is possible to map the supply chain, which protects it more robustly against disruptions and reduces the incidence of errors (Azzi et al., 2019; Ivanov et al., 2019; Montecchi et al., 2019). Supply chain partners can proactively take measures to counteract the risks and promptly react when the events occur (Azzi et al., 2019; Swan, 2018; Vinay Reddy, 2019).

Blockchain technology makes it possible to analyze the transaction time-stamps (Swan, 2018). In this way, a disturbance at the beginning of the supply chain – for example, the spread of the coronavirus (COVID-19) in Asia – can be detected in real-time. Thereby, each supply chain partner could foreseeably estimate the consequences of the interruption and adapt their respective forecasting and projections (Schlegel and Trent, 2014; Tang, 2006). The party responsible for the disruption as well as the immediately subsequent supply partner could thereafter be identified and traced (Svensson, 2000).

Furthermore, if supply chain partners use intelligent transport containers within the entire supply chain, like smart boxes or smart pallets, they can track the process' progress by the minute, and data record the entire process stream (modum, 2017). For example, temperature sensors could consistently record the surrounding temperature, and, if during shipment from China to Germany, temperatures exceed or fall below a certain level, the supply chain partners would be alerted to a potential impact on the quality or functionality of the shipped products. In this case, blockchain participants could check the temperature continuously during transport to reduce product quality risks (Tian, 2017).

If temperature tracking, e.g. detects adverse effects on product quality, the OEM could act immediately, informing their customer and initiating a new order from a second supplier to reduce supply risk. In this case, the applied blockchain technology reduces supply risks (Tian, 2017).

Due to the spread of the coronavirus, there is a significant worldwide increase in demand for ventilators and personal protective equipment – in particular, FFP masks and half masks. In the use case, the OEM produces the masks in Sweden and South-Africa. When the OEM recorded the increasing demand, all production capacities were ramped up and safety stocks were increased. By using blockchain technology, the OEM can track all materials and resources. Additionally, bottlenecks, due to the global supply chain, and reduced transport capacities can be identified quickly.

By merely comparing incoming and outgoing goods, as well as serial numbers and descriptions, every supply chain partner can compare the information and thus check whether the product comes from the real manufacturer or is a counterfeit (Bumblauskas et al., 2020; Fridgen et al., 2019; Fu and Zhu, 2019; Gao et al., 2019; Kumar et al., 2020; Montecchi et al., 2019). The use of blockchain technology prevents the circulation of falsifications and recalls can be executed faster (Azzi et al., 2019; Chang et al., 2019).

The value this provides can be exemplified in the FFP masks. In the corona crisis, masks have become a sought-after and valuable product, leading to questionable providers entering the world market: packaging can be misleading, advertised products may not really exist, and new masks

are replaced with expired ones on the way to the customer (Becker et al., 2020). Blockchain technology helps the OEM to protect themself against buying counterfeit goods, and to avoid legal difficulties that may result from mistakenly selling counterfeits. In doing such, the OEM reduces risks to their reputation and avoids demand-side risks, by creating trust among direct customers, who can individually check where the goods come from (Kshetri, 2018; Montecchi et al., 2019). These advantages can lead to better customer services, combined with reduced costs, and faster delivery (Wust and Gervais, 2018).

A lack of transparency and the withholding of important information and (real-time) data between partners in the supply chain not only poses risks on the flow of goods, but can also be crucial for supply chain finance, as missing information can result in payment delays (Jüttner et al., 2003; Manuj and Mentzer, 2008). By using blockchain technology, paper-based transactions can be automated. This automation can approve and verify invoices more quickly. With blockchain technology, everyone involved in the supply chain has easy access to real-time, finance-related data – such as the current invoice status – so that those payment transactions can be processed and monitored transparently (Hofmann et al., 2018; Omran et al., 2017). This extended period for financing improves the liquidity of the supply chain partners. The quick payment of invoices reduces the risk of liquidity bottlenecks and, thereby, the risk of supplier failures.

In addition to these positive aspects of increased transparency, supply chain transparency can also result in new risks from the equal visibility of the transaction history for all partners in the blockchain (Li et al., 2017). If there is only a small group of suppliers who can manufacture the semi-finished products, the competitor can quickly reconstruct order quantities and buyers. The supplier may, therefore, lose a competitive advantage. The OEM's delegation of power from the supply chain transparency should be critically considered. If the OEM has access to all blockchain data and transactions, they gain knowledge about each of the upstream suppliers as well as those respective sales volumes. It makes the OEM more powerful than before. There may be a risk that the OEM will consider skipping the supplier to negotiate price directly with the upstream suppliers.

Environmental risks

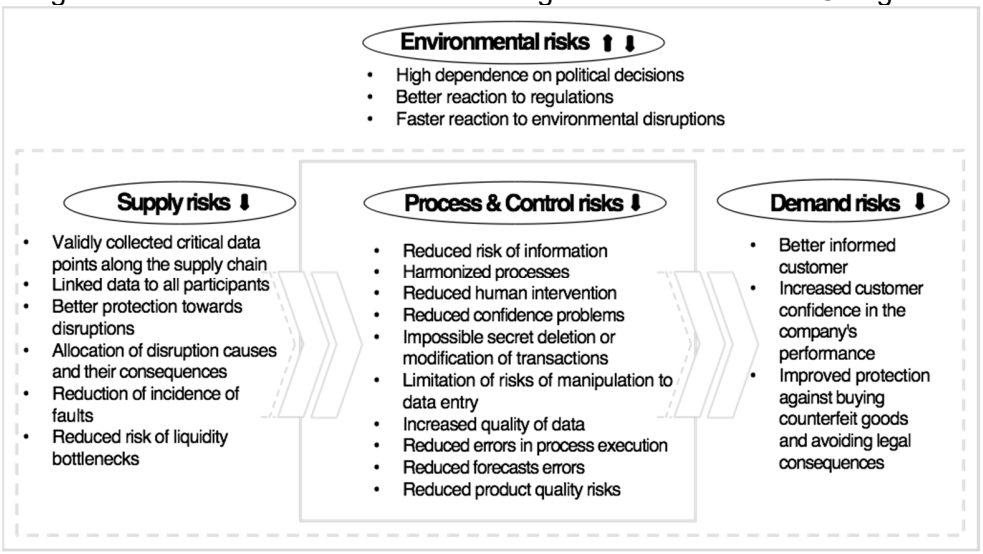
As described above, the occurrence of environmental risks, like natural disasters or worker strikes, cannot be influenced by individuals. However, the use of blockchain technology can help to identify these earlier so that measures can be taken before supply chain interruptions occur. Conversely, the transparency of the supply chain can help to accelerate the recovery process (Simchi-Levi et al., 2014).

Besides environmental risks, political risks also remain. In the use case, the supply chain partners are in different countries. Data must be stored in the blockchain to document the transactions and to achieve transparency. Therefore, the OEM must consider which data is subject to confidentiality by law (Tönnissen and Teuteberg, 2019). If political requirements prohibit the use or removal of personalized data, the transactions will be de-restricted and all the advantages of transparency, resulting from the use of blockchain, will be negated.

But due to the increased transparency in the supply chain, the OEM might react better to regulations, as with the obligation to report identified risks in products that can directly or indirectly lead to death or severe health impacts (Medical Devices Act; Medical device safety plan regulation). Should this reporting obligation be expanded in the future with further information, it could be saved or retrieved from the blockchain.

Figure 2 summarizes the results of how blockchain might influence the different risk classifications, based on Christopher and Peck (2004). Worth noting is that a non-overlapping allocation of the characteristics of different risk categories is not possible. An increase in transparency not only has a positive effect on internal processes and controls, but also on the entire supply chain (end-to-end-view) (Azzi et al., 2019; Swan, 2018; Vinay Reddy, 2019).

Figure 2: Effects of Blockchain Technologies on Different Risk Categories



RECOMMENDATION FOR SCRM

Applying the risk categorization developed by Christopher & Peck (2004), we analyzed how various risks might be influenced by the use of blockchain technology. While internal (process and control), demand, and supply chain risks might be reduced, due to defined blockchain standards and the transparency of the flow of goods and information, environmental risks remain volatile and cannot be influenced.

On the one hand, the management of these risks will decrease in scope, but, on the other hand, risks related to the distribution of power must be given more importance. Those risks, which could previously be classified as low in probability, should now be given greater weight. The changing power position of a given company in a private blockchain, such as the OEM in the fictitious use case, might have a significant influence on the bargaining power of buyers and sellers, and thus on their overall competitiveness. If the OEM is not the main customer for certain components, they have less bargaining power and thus, are dependent on the material or products of their suppliers. Here, the OEM could use the transactions entered in a private blockchain to deduce the inter-dependencies between the various suppliers and main customers. As a result, the OEM has information about their suppliers and main customers and can change the priorities of their supplier network.

We also detected a shift of the SCRM focus: the OEM should apply additional effort towards the strategic alignment of SCRM – such as the election of blockchain users or the development of blockchain standards. Consequently, this means that thorough planning is required. By contrast, the OEM can reduce the management of the operating supplier actions, because supply chain operations will fully conform to defined blockchain standards.

From the analyzes of the use case, it was discovered that the processes can be automated, which means that the individual phases of the risk management process (risk identification, analysis and assessment, handling and control) could also be further automated. The availability of uniform data records facilitates the use of risk analysis methods and enables the easy integration of risk control variables (e.g. delayed provision of goods or delivery reliability at various times).

However, for better risk assessments, early warning indicators need to be developed and implemented to analyze data from blockchain transactions. The OEM could then use the results to improve the cross-departmental exchange of risk-related information. In the use case, the

SCRM is decentralized at the OEM. Employees in operations, purchasing, R&D, sales, product management, quality-control, and finance deal with supply chain risks. Purchasing carries out supplier evaluations and supplier audits. If a supplier is consistently highlighted as a weak-link, a pre-supplier examination or the replacement of the supplier could take place. At the same time, R&D employees have to select future suppliers during the product development process. Hence, access to this supplier information during product development can reduce supply and process risks at a later stage.

Additionally, early warning indicators could become more efficient in combination with machine learning (Yong et al, 2020). With advanced analytical methods, the OEM could make continuous forecasts to determine the optimal resource allocation on short notice (Blackburn et al., 2015; Huang, 2020). Based in part on the real-time data of the delivery and production statuses in the supply chain, an optimal allocation of available transport and personnel capacities can be predicted.

Nevertheless, it must be clear to all blockchain participants how to interpret blockchain data, based on early warning indicators (Tang and Veelenturf, 2019). All partners must be aware of what subsequent measures should be taken in response and develop supply chain-wide business continuity plans so that delivery and production can be started quickly. As a result, cooperative risk management is necessary.

Also, the throughput speed of the individual phases of the risk management process will increase in real-time with data availability. This necessitates a more reactive, data-driven SCRM orientation in order to evaluate incoming information immediately.

Finally, it should be emphasized that the OEM should always carry out a cost-benefit analysis when implementing blockchain technology. As an essential requirement for the implementation of blockchain technology, all partners involved, regardless of their national origin, must use the uniform blockchain standards. This requires all stakeholders to first come together to discuss the procedures prior to implementation. But, does the added value from risk reduction exceed the effort to implement blockchain technology? Or, could measures with less effort but similar results, be adopted for the entire supply chain?

Until now, a holistic SCRM approach for the use of blockchain technology is still missing, but the results have shown that cooperative risk management for the entire supply chain will continue to grow in importance. For this reason, further research into this topic is encouraged.

CONCLUSIONS AND LIMITATIONS

An efficient SCRM relies on transparency about the flow of goods and information within the network. The fictitious use case, describing the supply chain of a leading international, medical and safety technology manufacturing company, explores the changes in supply chain risks' scope and weight, due to increased visibility and supplier monitoring with blockchain technologies. Human risks might decrease, while risks associated with the division of power might become more critical.

A company's traditional SCRM must be adapted to the new conditions in the supply chain when using blockchain technology.

Current risk measures should take an overall view of the supply chain, and not only refer to individual links or supply chain partners. A cooperative risk management approach is needed to mitigate supply chain interruptions through measures taken in parallel by multiple partners.

The results of the research are limited to the aforementioned fictitious use case, and a more detailed examination of the use case is required. The results should also be compared with other blockchain configurations, such as public blockchain, and additional use cases should be examined in terms of SCRM. The empirical review of the findings is still pending.

This paper provides a useful starting point to deepen research and scientific discussions on this subject.

REFERENCES

Awwad, M., Kalluru, S. R., Airpulli, V. K., Zambre, M. S., Marathe, A. and Jain, P. (2018) 'Blockchain technology for efficient management of supply chain', *Proceedings of the International Conference on Industrial Engineering and Operations Management*, vol. 2018, SEP [Online]. Available at https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067036639&partnerID=40&md5=4fbfe79a141e148ac265e9dfbf73c595.

Azzi, R., Chamoun, R. K. and Sokhn, M. (2019) 'The power of a blockchain-based supply chain', *Computers and Industrial Engineering*, vol. 135, pp. 582–592 [Online]. DOI: 10.1016/j.cie.2019.06.042.

Barczak, A.; Dembińska, I.; Marzantowicz, L. (2019). "Analysis of the Risk Impact of Implementing Digital Innovations for Logistics Management". *Processes*, 7, 815 [Online]. DOI:10.3390/pr7110815.

Becker, S.; Schmergal, C. and Traufetter, G. (2020) *Elf Millionen Masken, alle Schrott* [Online]. Available at https://www.spiegel.de/politik/deutschland/andreas-scheuer-und-der-corona-nepp-aus-china-elf-millionen-masken-alle-schrott-a-00000000-0002-0001-0000-000170604444.

Bermeo-Almeida, O., Cardenas-Rodriguez, M., Samaniego-Cobo, T., Ferruzola-Gómez, E., Cabezas-Cabezas, R., Bazán-Vera, W. and Del Cioppo-Morstadt J., Vera-Lucio N., Bucaram-Leverone M., Valencia-Garcia R., Alcaraz-Marmol G. (2018) 'Blockchain in agriculture: A systematic literature review', *Communications in Computer and Information Science*, vol. 883, pp. 44–56 [Online]. DOI: 10.1007/978-3-030-00940-3_4.

Blackburn, R., K. Lurz, B. Priese, R. Göb, and I.-L. Darkow. 2015. "A Predictive Analytics Approach for Demand Forecasting in the Process Industry." *International Transactions in Operational Research* 22 (3): 407–28. doi:10.1111/itor.12122.

Bocek, T., Rodrigues, B. B., Strasser, T. and Stiller, B. (2017) 'Blockchains everywhere - a use-case of blockchains in the pharma supply-chain', *Proceedings of the IM 2017 - 2017 IFIP/IEEE International Symposium on Integrated Network Management: May 8-12, 2017, Lisbon, Portugal.* Piscataway, NJ, IEEE, pp. 772–777.

Bumblauskas, D., Mann, A., Dugan, B. and Rittmer, J. (2020) 'A blockchain use case in food distribution: Do you know where your food has been?', *International Journal of Information Management*, vol. 52, p. 102008 [Online]. DOI: 10.1016/j.ijinfomgt.2019.09.004.

Casino, F., Dasaklis, T. K. and Patsakis, C. (2019) 'A systematic literature review of blockchain-based applications: Current status, classification and open issues', *Telematics and Informatics*, vol. 36, pp. 55–81 [Online]. DOI: 10.1016/j.tele.2018.11.006.

Chang, Y., Iakovou, E. and Shi, W. (2019) 'Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities', *International Journal of Production Research* [Online]. DOI: 10.1080/00207543.2019.1651946.

Chopra, S. and Sodhi, M. (2004) 'Managing Risk to Avoid Supply-Chain Breakdown', *MIT Sloan Management Review*.

Christopher, M. (2005) *Logistics and supply chain management: Creating value-adding networks*, 3rd edn, Harlow, Financial Times/Prentice Hall.

Christopher, M. and Peck, H. (2004) 'Building the Resilient Supply Chain', *International Journal of Logistics Management*, vol. 15, pp. 1–13.

Corkery, Michael and Nathaniel Popper (2018) 'Supply Chain Management with Blockchain Technologies', *The New York Times.*

Czachorowski, K., Solesvik, M. and Kondratenko, Y. (2019) 'The application of blockchain technology in the maritime industry', *Studies in Systems, Decision and Control*, vol. 171, pp. 561–577 [Online]. DOI: 10.1007/978-3-030-00253-4_24.

Figorilli, S., Antonucci, F., Costa, C., Pallottino, F., Raso, L., Castiglione, M., Pinci, E., Del Vecchio, D., Colle, G., Proto, A. R., Sperandio, G. and Menesatti, P. (2018) 'A blockchain implementation prototype for the electronic open source traceability of wood along the whole supply chain', *Sensors (Switzerland)*, vol. 18, no. 9 [Online]. DOI: 10.3390/s18093133.

Fridgen, G., Körner, M.-F., Sedlmeir, J. and Weibelzahl, M. (2019) '(How) can blockchain contribute to the management of systemic risks in global supply networks?', *CEUR Workshop Proceedings*, vol. 2397 [Online]. Available at https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068934406&partnerID=40&md5=ffa433c8e9bbff037ee835474e5bd17f.

Fu, Y. and Zhu, J. (2019) 'Big Production Enterprise Supply Chain Endogenous Risk Management Based on Blockchain', *IEEE Access*, vol. 7, pp. 15310–15319 [Online]. DOI: 10.1109/ACCESS.2019.2895327.

Gao, B., Zhou, Q., Li, S. and Liu, X. (2019) 'A Real Time Stare in Market Strategy for Supply Chain Financing Pledge Risk Management', *IEEE International Conference on Industrial Engineering and Engineering Management*, 2019-December [Online]. DOI: 10.1109/IEEM.2018.8607805.

Giunipero, L. C. and Aly Eltantawy, R. (2004) 'Securing the upstream supply chain: a risk management approach', *International Journal of Physical Distribution & Logistics Management*, vol. 34, no. 9, pp. 698–713.

Hackius, N., Reimers, S. and Kersten, W. (2019) 'The Privacy Barrier for Blockchain in Logistics: First Lessons from the Port of Hamburg', in Bierwirth, C., Kirschstein, T. and Sackmann, D. (eds) *Logistics Management: Strategies and Instruments for Digitalizing and Decarbonizing Supply Chains - Proceedings of the German Academic Association for Business Research, Halle 2019,* New York, Berlin, Springer; Springer [Distributor], pp. 45–61.

Hofmann, E., Strewe, U. M. and Bosia, N. (2018) *Supply Chain Finance and Blockchain Technology: The Case of Reverse Securitisation*, Cham, Springer International Publishing.

Huang, J. 2020. "How to Drive a Holistic End-to-End Supply Chain Risk Management." *Journal of Supply Chain Management*, Logistics & Procurement 2 (4).

Ivanov, D., Dolgui, A., Das, A. and Sokolov, B. (2019) 'Digital Supply Chain Twins: Managing the Ripple Effect, Resilience, and Disruption Risks by Data-Driven Optimization, Simulation, and Visibility', *International Series in Operations Research and Management Science*, vol. 276, pp. 309–332 [Online]. DOI: 10.1007/978-3-030-14302-2_15.

Jabbar, K. and Bjørn, P. (2018) 'Infrastructural grind: Introducing blockchain technology in the shipping domain', *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* [Online]. DOI: 10.1145/3148330.3148345.

Jüttner, U., Peck, H. and Christopher, M. (2003) 'Supply chain risk management: outlining an agenda for future research', *International Journal of Logistics Research and Applications*, vol. 6, no. 4, pp. 197–210.

Kersten, W., Hohrath, P., Böger, M. and Singer, C. (2011) 'A Supply Chain Risk Management Process', *International Journal of Logistics Systems and Management*, Vol. 8, No. 2, pp. 152–166.

Kleinaltenkamp M. and Jacob F. (2002) 'German Approaches to Business-to-Business Marketing Theory: Origins and Structure', *Journal of Business Research*, 55(2), pp. 149-155 [Online]. DOI: 10.1016/S0148-2963(00)00150-8.

Kshetri, N. (2018) 'Blockchain's roles in meeting key supply chain management objectives', *International Journal of Information Management*, vol. 39, pp. 80–89 [Online]. DOI: 10.1016/j.ijinfomgt.2017.12.005.

Kumar, A., Liu, R. and Shan, Z. (2020) 'Is Blockchain a Silver Bullet for Supply Chain Management? Technical Challenges and Research Opportunities', *Decision Sciences*, vol. 51, no. 1, pp. 8–37 [Online]. DOI: 10.1111/deci.12396.

Lacity, M. C. (2018) 'Addressing key challenges to making enterprise blockchain applications a reality', *MIS Quarterly Executive*, vol. 17, no. 3, pp. 201–222 [Online]. Available at https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053527381&partnerID=40&md5=e572d3319f403a2912c530fabefdcebe.

Lambert, D. M.; Stock, J. R.; Ellram, L. M. (1998b): *Fundamentals of logistics management*, Irwin/McGraw-Hill: Boston et al.

Li, Z., Wu, H., King, B., Miled, Z. B., Wassick, J. and Tazelaar, J. (2017) 'On the Integration of Event-Based and Transaction-Based Architectures for Supply Chains', *IEEE 37th International Conference on Distributed Computing Systems workshops - ICDCSW 2017: Proceedings : 5 June 2017, Atlanta, Georgia.* Piscataway, NJ, IEEE, pp. 376–382.

Lin, Q., Wang, H., Pei, X. and Wang, J. (2019) 'Food Safety Traceability System Based on Blockchain and EPCIS', *IEEE Access*, vol. 7, pp. 20698–20707 [Online]. DOI: 10.1109/ACCESS.2019.2897792.

Liu, L., Li, F. and Qi, E. (2019) 'Research on Risk Avoidance and Coordination of Supply Chain Subject Based on Blockchain Technology', *Sustainability (Switzerland)*, vol. 11, no. 7.

Manuj, I. and Mentzer, J. T. (2008) 'Global supply chain risk management strategies', *International Journal of Physical Distribution & Logistics Management*, vol. 38, no. 3, pp. 192–223.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D. and Zacharia, Z. G. (2001) 'Defining supply chain management', *Journal of Business Logistics*, vol. 22, no. 2, pp. 1–25.

modum (2017) 'Data integrity for supply chain operations, powered by blockchain technology.

Montecchi, M., Plangger, K. and Etter, M. (2019) 'It's real, trust me! Establishing supply chain provenance using blockchain', *Business Horizons*, vol. 62, no. 3, pp. 283–293 [Online]. DOI: 10.1016/j.bushor.2019.01.008.

Nakamoto, S. (2008) 'Bitcoin: A Peer-to-Peer Electronic Cash System' [Online]. Available at https://bitcoin.org/bitcoin.pdf.

Nikhilesh De (2018) *Walmart Tells Produce Suppliers to Use Blockchain by Next September* [Online]. Available at https://www.coindesk.com/walmart-will-start-tracking-groceries-on-blockchain-in-2019.

Norrman; A. and Lindroth, R. (2004) 'Categorization of supply chain risk and risk management', in Clare Brindley (ed) *Supply chain risk,* United Kingdom, Ashgate, pp. 14–27.

Omran, Y., Henke, M., Heines, R. and Hofmann, E. (2017) 'Blockchain-driven supply chain finance: Towards a conceptual framework from a buyer perspective. 2017. - IPSERA 2017. - Budapest - Balatonfüred'.

Petersen, M., Hackius, N. and See, B. von (2018) 'Mapping the Sea of Opportunities: Blockchain in Supply Chain and Logistics', *it - Information Technology*, vol. 60, pp. 263–271.

Plowman, E. G. (1964) *Lectures on elements of business logistics*, S.l., Stanford University Graduate School of Business.

Rao, S. and Goldsby, T. J. (2009) 'Supply chain risks: a review and typology', *The International Journal of Logistics Management*, vol. 20, no. 1, pp. 97–123 [Online]. DOI: 10.1108/09574090910954864.

Ritchie, B. and Brindley, C. (2004) 'Risk characteristics of the supply chain - A contingency framework', *Supply Chain Risk*, pp. 28–42.

Ross, S.A. (1973) 'The Economic Theory of Agency: The Principal's Problem', *American Economic Review*, 63(2), pp. 134-139.

Scalco, P. R., Lopez, R. A. and He, X. (2017) 'Buyer and/or seller power? An application to the Brazilian milk market', *Applied Economics Letters*, vol. 24, no. 16, pp. 1173–1176.

Schlegel, G. L. and Trent, R. J. (2014) 'Supply chain risk management: An emerging discipline', *Supply Chain Risk Management: An Emerging Discipline*.

Schröder, M. (2019) *Strukturierte Verbesserung des Supply Chain Risikomanagements*.

Simchi-Levi, D., Schmidt, W. and Wei, Y. (2014) 'From Superstorms to factors fires. Managing unpredictable supply-chain disruptions', *Harvard Business Review*, Jan/Feb2014, Vol. 92, p. 96-101.

Stevenson, M.; Spring, M. (2007): Flexibility from a supply chain perspective: Definition and Review, In: *International Journal of Operations & Production Management*. Vol. 27, No. 7, pp. 685-713.

Stiglitz, J.E. (2002) 'Information and the Change in the Paradigm in Economics', *American Economic Review, 92(3), pp. 460-*501 [Online]. DOI. 10.1177/056943450404800103.

Svensson, G. (2000) 'A conceptual framework for the analysis of vulnerability in supply chains', *International Journal of Physical Distribution & Logistics Management*, vol. 30, no. 9, pp. 731–750 [Online]. DOI: 10.1108/09600030010351444.

Swan, M. (2018) 'Blockchain for Business: Next-Generation Enterprise Artificial Intelligence Systems', *Blockchain technology: Platforms, tools and use cases.* Cambridge, MA, Academic Press, pp. 121–162.

Tang, C. S. (2006) 'Perspectives in supply chain risk management', *International Journal of Production Economics*, vol. 103, no. 2, pp. 451–488 [Online]. Available at https://econpapers.repec.org/RePEc:eee:proeco:v:103:y:2006:i:2:p:451-488.

Tang, C. S. and Veelenturf, L. P. (2019) 'The strategic role of logistics in the industry 4.0 era', *Transportation Research Part E: Logistics and Transportation Review*, vol. 129, pp. 1–11 [Online]. DOI: 10.1016/j.tre.2019.06.004.

Tian, F. (2017) 'A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things', *2017 14th International Conference on services systems and services management (ICSSSM).* 345 E 47th ST, New York, NY 10017 USA, IEEE.

Tönnissen, S. and Teuteberg, F. (2019) 'Analysing the impact of blockchain-technology for operations and supply chain management: An explanatory model drawn from multiple case studies', *International Journal of Information Management* [Online]. DOI: 10.1016/j.ijinfomgt.2019.05.009.

Vinay Reddy (2019) 'Enhancing supply chain management using blockchain technology', *International Journal of Engineering and Advanced Technology*, no. 8, pp. 4657–4661.

Wohlin, C. (2014) 'Guidelines for snowballing in systematic literature studies and a replication in software engineering', *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering.* London, England, United Kingdom, 5/13/2014 - 5/14/2014. New York, NY, ACM, pp. 1–10.

Wolff, F.-C. (2016) 'Bargaining powers of buyers and sellers on the online diamond market: a double perspective non-parametric analysis', *Annals of Operations Research*, vol. 244, no. 2, pp. 697–718 [Online]. DOI: 10.1007/s10479-016-2160-1.

Wust, K. and Gervais, A. (2018) 'Do you Need a Blockchain?', *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pp. 45–54.

Yang, C.-S. (2019) 'Maritime shipping digitalization: Blockchain-based technology applications, future improvements, and intention to use', *Transportation Research Part E: Logistics and Transportation Review*, vol. 131, pp. 108–117 [Online]. DOI: 10.1016/j.tre.2019.09.020.

Yong, B., J. Shen, X. Liu, F. Li, H. Chen, and Q. Zhou. 2020. "An Intelligent Blockchain-Based System for Safe Vaccine Supply and Supervision." *International Journal of Information Management* 52. doi:10.1016/j.ijinfomgt.2019.10.009

Zhao, N. (2019) 'Managing interactive collaborative mega project supply chains under infectious risks', *International Journal of Production Economics*, vol. 218, pp. 275–286 [Online]. DOI: 10.1016/j.ijpe.2019.06.008.

Zheng, K., Zhang, Z., Chen, Y. and Wu, J. (2019) 'Blockchain adoption for information sharing: risk decision-making in spacecraft supply chain', *Enterprise Information Systems* [Online]. DOI: 10.1080/17517575.2019.1669831.

Zhu, Q. and Kouhizadeh, M. (2019) 'Blockchain Technology, Supply Chain Information, and Strategic Product Deletion Management', *IEEE Engineering Management Review*, vol. 47, no. 1, pp. 36–44 [Online]. DOI: 10.1109/EMR.2019.2898178.

DECISION SCIENCES INSTITUTE
Engine Light Repair Shop Evaluation for a Brazilian Airline

Gustavo Henrique Melhado Trevilatto
Embry Riddle Aeronautical University
melhadog@my.erau.edu
Marcelo Miranda
Embry Riddle Aeronautical University
miranm10@my.erau.edu
Leila Halawi
Embry Riddle Aeronautical University
halawil@erau.edu

ABSTRACT

The maintenance cost of a commercial aircraft comprises airframe, components, and engines. The most expensive of them are engines, which may represent up to 60% of an aircraft cost, and therefore will have a significant impact on any airline finance. As a result, engine fleet maintenance management is crucial for any airline sustainability. Apart from heavy maintenance where life-limited parts are replaced and its performance restored, aircraft engines are often required to come off wing for light repair due to operational issues like foreign object damage, high oil consumption, and vibration issues. In other cases, even though being operational, engines are required to come off the wing and undertake repair processes to comply with lease return conditions. Although these light repairs may be simple and relatively short in time, they require the engines to be transported to dedicated repair shops impacting engine availability due to shop slots unavailability and the logistics. This drives airlines to have additional spare engines on its fleet, increasing its operating costs. This research is aimed at evaluating the pros and cons any Brazilian airline may have if it decides to have its engine shop for light repairs. Due to the complexity of the engine repair process, this research will be focused on a specific repair for a particular engine, which, if developed internally, may allow airlines to expand their light repair capability and, in the long term, improve engine availability.

Keywords: brazil airlines, engines repairs, Maintenance, repair and Overhaul (MRO)

INTRODUCTION

The manufacturing performance of aircraft components is relatively more important than the final sales price factor, the investment is high, and the control is thorough. (Mergent, 2019). MRO, Maintenance, Repair, and Overhaul is a service provider (which may include an OEM) that can perform maintenance, repair, and overhaul functions on an Aircraft or Engine (Scheinberg, 2017). Prices are very high compared to production, this is overwhelmingly reflected in costs at MROs, and airlines are paying this for replacement. (Vieira, 2016). An MRO must be certified by an applicable Aviation Authority (Scheinberg, 2017). All these requirements, in conjunction, contribute to a limited number of engine MROs in the world and especially in Brazil. As a result, engine repair cycles may not be as short as desired by Brazilian airlines due to limited repair stations available in the market or long transportation time to import-export engines. The International Air Transportation Association (IATA) Maintenance Cost Task Force estimated the global MRO in 2017 at USD 76 Billion, excluding overhead (IATA, 2018). This value comprises 42% spending on engines, 21%

spending on components, 20% spending on base maintenance, and 17% spending online maintenance and represents around 11% of airlines operational cost (IATA, 2018).

Due to the high technology, high costs, expertise, and complexity involved in the design and certification of an aircraft, there are only a few manufacturers in the world. According to IATA's MCTF, in FY2017, the world fleet count for all active aircraft in commercial operations was 25,870, being 80%, which are manufactured by Airbus and Boeing. The other 20% is represented by Bombardier, Embraer, Fokker, ATR, and others (IATA 2018). In the construction of large aircraft production, the negotiations to build large Turbofans engines are from large companies such as GE - General Electric, United Technologies, PW with the Pratt & Whitney brand, and Rolls Royce that together surpass 93 % from the market. (The Economist June 1, 2019).

There is no doubt that 93% is a very high percentage. The result is that it severely contributes to a reduction in the available choices for any airline when considering engine fleet selection. The lack of multiple engine manufacturers in the market creates a condition close to a monopoly where any engine related cost is imposed by a few manufacturers, leaving for airlines limited or no options for bargaining. It was not by chance that, in 2016, IATA filed complaints against CFM with the European Commission's competition office alleging abuses of dominant positions by manufacturers of aviation equipment. The allegation resulted in a conduct policy aimed at restoring fair commercial practices in the aviation engine market. CFM released the conduct policy in July 2018 with 41 pages. CFM broadly said that its service licenses and warranties do not discriminate against the use of non-OEM material (Gubisch, M., & Hemmerdinger, J. (2018). Non-OEM material refers to spares developed under a parts manufacturer approval (PMA) and repairs devised under a designated engineering representative (DER) certificate (Gubisch, M., & Hemmerdinger, J. (2018). The historic deal has implications beyond CFM and is intended to be a prime example for the other engine. Beneficiaries include airlines, aircraft outsourcers, parts manufacturers, and MRO maintenance companies and third parties. (Aircraft Commerce, 2019).

Engines are very complex and precise machines that demand large investments in infrastructure such as bench tests, apparatus, tooling, and also know-how to be repaired appropriately. Also, the lack of dedicated training aligned with the level of specialization and expertise required for performing an engine repair makes it challenging to create a capable research force. All these requirements, summed up, do not contribute to the development of engine repair shops in many places in the world and especially in Brazil. When the repair is done locally in Brazil, airlines count only with one repair shop, which tends to offer little to no margin for slot allocation or pricing negotiation. As a result, engines may become unavailable for more extended periods than desired what increases operational costs due to engine substitution for a spare engine. When the machines are sent to be repaired abroad, turnaround times tend to be high and also contribute to engine unavailability for more extended periods.

Therefore, it is clear that engine fleet optimization is of paramount importance for any airline finance. This research intends to provide insight on how Brazilian airlines could improve its repair cycle through internalizing a specific repair for a particular engine. This research aims to identify insights for alternatives to reduce airline costs by improving engine fleet availability by developing the airline's internal light repair capability. It is assumed that the airline already has some MRO capability, but for airframe only and not dedicated to engine repair. The research will focus on a specific repair procedure on a particular type of engine. We will investigate if this is technically and economically feasible. This scenario may contribute to a further and broader development of the airline repair capability in house. The ultimate aim of the research is to reduce the airline operational costs either by using its

internal workforce or by improving its engine availability in a way that allows its spare engine fleet to be reduced.

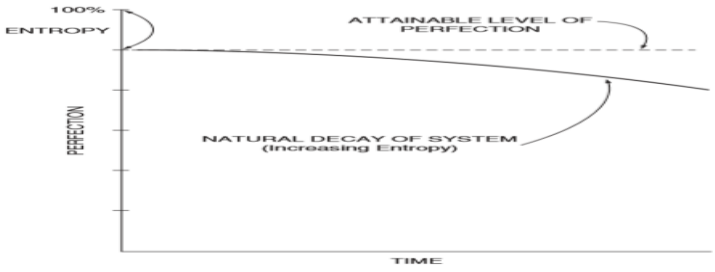
The analysis of the workforce consists of evaluating the training required to elevate the current know-how of the airlines' employees to the requirements needed to successfully perform the job as well as the gains it may offer for cost reduction. It is important to remember that the cost to repair an engine is comprised of material (new, used, or rebuilt parts) and labor. The prices of material are typically controlled by the OEMs and offer little to no opportunity for cost reduction. The costs of labor may provide opportunities for cost reduction for two reasons: 1 - local manpower costs do not fluctuate with exchange rates; 2 – local manpower costs consist of pure cost and do not bring any profit margin within its value (whereas outside MRO values for labor bring the MRO's profit margin). Considerations about repair costs taxation will also be taken into account for labor as well as material.

The analysis of the improvement of engine availability takes into account the slot availability and repair TAT that may be obtained if the repair is conducted within the airline borders.

Maintenance

Maintenance is summarized in an ongoing systematic process to ensure continuity of actual operation with the same level of safety and reliability as the model. (Kinnisson, 2012). In the real world, due to several constraints such as manufacturing, material, time, costs, etc. systems cannot be designed to 100% of perfection. However, every System is designed to a certain level of perfection that fits its intent. As time passes, due to wear and tear, there is a natural increase in the entropy of any system, which characterizes its deterioration. This deterioration distances the current system condition from its original design intent and can be seen in Figure 1.

Figure 1- Difference Between Theory and Practice



When the System deteriorates to certain levels still accepted by the operator, manufacturer, or regulator, it may continue to operate. However, an intervention is needed to prevent it from falling outside the limits of its design intents. This process is known as predictable, scheduled, or preventive maintenance. Preventative or scheduled maintenance is usually done using scheduled inspections or condition monitoring. This usually happens because the aircraft has parts that must be periodically inspected at a particular time, which ensures the operability of the designed components. The manufacturers typically request this inspection time in the MPDs. For example, on engines, LLPs replacement or EGT restoration generally are classified as scheduled maintenance because they can be predictable based on the operation.

There are also times when the System deteriorates to unacceptable levels by the operator, manufacturer, or regulator at a non-predictable pattern, usually sooner than expected. When this happens, an intervention is necessary to restore it to its original design intent. This process is known as unpredictable or unscheduled maintenance. The guideline is that

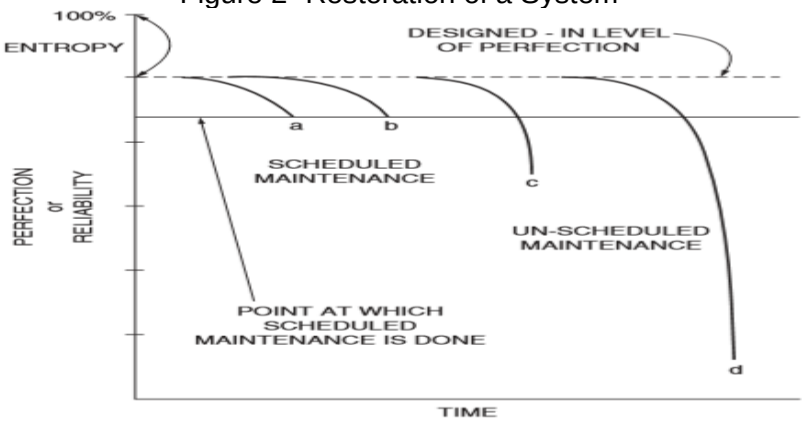
problems should be observed in two parts, such as primary and secondary damage, to increase the availability of equipment and its systems as a whole, to reduce repair time and increase productivity (Viles, 2007). It is also aimed at capturing relevant information for analysis. Table 1 highlights the process steps the engineers and mechanics usually take in unscheduled maintenance.

Table 1- Process Steps in Unscheduled Maintenance

1	Determination of the finding e.g., via inspection, exposure, or record in the technical logbook
2	Documentation of the finding
3	Troubleshooting/root cause analysis considering possible secondary errors and risks
4	Assessment and classification of finding including identification of applicable maintenance instructions / approved maintenance data
4a	If standard maintenance data are insufficient: Obtain individual maintenance instruction from the responsible design organization
5	Determination of personnel qualification, material requirements, time and financial effort
6	Rectification and documentation of the processing

The unpredictable or unscheduled maintenance can be seen in Figure 2, curve c, and d. Technical wear processes are not always predictable; this is a consequence of the fact that unpredictable processes (Hinsch, 2018). On engines, oil leakage, high oil consumption, abnormal vibration, foreign object damage are general types of unscheduled maintenance because they cannot always be predictable.

Figure 2- Restoration of a System



The term maintenance is also referred to as all actions which delay wear and maintain planned conditions. Just as they gradually wear and keep intentional states, the term overhaul is usually used to describe all steps taken to re-establish the intended state of a system or component.

Increasing and improving equipment availability is a focus of maintenance service, which means improving reliability and reducing repairs. (Viles, 2007). As a result, it requires high organizational flexibility in terms of factual conditions, space, as well as scheduling and capacities (Hinsch, 2018):

Factual flexibility needs because it is understood that it is a maintenance activity, often complex research with few interventions. As a result, maintenance personnel shall be having a lot of knowledge and be very well educated. This causes researchers to be gradually replaced by machines in maintenance activities that would be unlikely. (Hinsch, 2018).

Temporal and Capacitive flexibility It is fundamental for the component; failures in equipment and assemblies are unstable; as a result, necessary resources can be planned to a limited extent. Maintenance activities, also, production downtimes, are costly and should be taken seriously considering periods, and this can often happen (Hinsch, 2018).

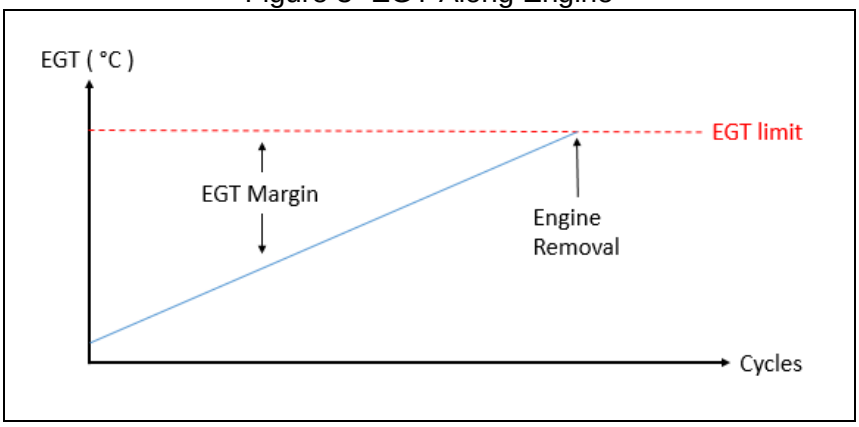
Spatial flexibility It is essential, as far as possible, that the maintenance is performed within its environment, due to the possibility of moving the equipment; this happens due to several factors, such as distance, weight, and others. (Hinsch, 2018).

The three main areas of concern on a commercial aircraft are components, engines, and airframe. They are thus subdivided to demonstrate the importance of their central regions and the cost difference between them over their lifetime, which in short has in common the impact on market value. (Ackert, 2010).

Engines are designed to operate several cycles (which are driven by their life-limited parts - a lifetime of LLPs, in most cases, are between 15,000 and 30,000 cycles (Ackert, 2010)) under certain operational limits. One of the central operating limits is the exhaust gas temperature (EGT). The EGT is researched in the unit of measurement in degrees Celsius, and its purpose is to inform the engine gas temperature, which relatively speaking, engine deterioration can occur with the high value or degree of EGT. (Ackert, 2010). High EGT can cause damage to parts and shorten component life, so EGT indication and performance must be monitored not to exceed limits. (Ackert, 2010).

When an engine approaches its EGT limit, it is time for it to be repaired. Figure 9 illustrates the EGT increase along with the engine life.

Figure 3- EGT Along Engine



Engines may come off the wing on a scheduled or unscheduled basis. Engines are typically removed from the branch on a planned fashion to be submitted to shop visits and have their LLP replaced or their EGT restored. Other times they can be removed from the wing to be

sent to 'hospital' shops, which carry out minor repairs and fan blade changes, but cannot perform module disassemblies and shop visits (Aircraft Commerce, 2001).

Airlines routinely monitor engine utilization by counting its cycles and measuring its exhaust gas temperature through an aircraft health monitoring system. With the collected values plus considering the current condition of an engine and its researched utilization, its maintenance may be scheduled. Scheduled removals are engine maintenance events typically driven by 1 - engine life-limited parts (LLP) expiration, 2 - poor performance (EGT loss), 3 - compliance with the engine return conditions to its lessor (either by being kept on the shelf to save its cycles or by being submitted to a repair to comply with its redelivery condition to its lessor) or 4 - even staggering purposes. Staggering is the process of removing an engine from the wing for later use.

There may be operational issues, however, which demand the engines to come off the wing on an unscheduled basis. Unscheduled removals are typically driven by operational matters such as foreign object damage, abnormal fuel or oil consumption, vibration out of limits, among other factors. In all cases, a minor intervention or a minor repair turns out to be necessary to return the engine under the same conditions, as usual, to operate so that it can be back to the wing.

The lifetime of an engine is fully persuaded by its thrust rating, either by the number of cycles or hours of operation. (Ackert, 2011).

After being removed, engines may be kept on stock, redelivered to its lessor, or submitted to the repair process. Either scheduled or unscheduled removals have an impact on airline finance and require a spare engine to sustain the aircraft in operation while the removed engine is under the repair cycle.

Everything must be taken into account in costs, such as fleet model, research methodology, maintenance scheduling, task planning, engine types and accessories, and so on. (Aircraft Commerce, 2018).

The maintenance process to avoid research shop engine costs can be broken as follows:

1 - Material Cost: primary cost driver, material replacements are around 60% to 70% of the cost of a machine shop visit; this can still increase costs if LLP parts require replacement.

2 - Direct labor: accounts for approximately 20% to 30% of the total cost;

3 - Repairs: accounts for 10% to 20%. (Ackert, 2011).

The type of aircraft and its engines vary greatly, so removal, repair, and installation procedures are complicated. Most of the time, however, it is easy to manipulate the way to connect and disconnect the hydraulic, electric, and pneumatic components, as well as the fuel lines, air intake, and exhaust. (United States Flight Standards Service, 2018).

The typical repair or overhaul cycle for an engine can be observed in table 2. A reduction in the length of the repair cycle illustrated in figure 3 improves the engine availability, therefore, requiring fewer spare engine for the same airline.

Table 2- Fundamental Engine Maintenance Procedure

Removal Transportation	Incoming Inspection Incoming Test Run	Disassembly	Cleaning Inspection NDT Repair Maintenance	Assembly	Test run QEC Buildup Completion	Transportation Installation Run-up Test flight

Competitiveness among airlines only tends to increase, and any reduction in costs and is significant. (Fritzsche, 2014). Research and maintenance procedures have always been influential in research to ensure the shortest possible ground time or unavailable aircraft. (Fritzsche, 2014).

Apart from scheduled (preventive) maintenance and unscheduled (unpredictable) maintenance, care can also be classified as Line Maintenance and Base Maintenance. Line maintenance is typically described as activities that are aimed at preserving the intended conditions of a component or System. Whereas base maintenance, under normal circumstances, refers to measures of an overhaul. Base maintenance typically refers to activities that are aimed at restoring the intended conditions of a component or System. The technical depth to which the tasks or activities demands is the main driver for classifying maintenance as base or line. In general, the eligibility conditions for essential maintenance are stricter and more demanding than the line maintenance. As a rule of thumb, for aircraft, events below the Check "C" are usually considered to be line maintenance. For engine events such as lube blade, oil filter replacement, boroscope inspection, chip detector inspection, trim balance are generally regarded as line maintenance. C-checks, D-checks are typical aircraft base maintenance activities, and engines LLPs replacement or EGT restoration are typically classified as an engine overhaul (table 3).

Table 3- Comparison of Different Line and Base Maintenance Checks for Widebodies

Check	Frequency	Working hours	Duration
Line Maintenance			
S-Check	weekly	10 - 50	3 - 5 hours
A-Check	every 4 - 8 weeks	50 - 250	Approx. 12 hours
Base Maintenance			
C-Check	ca. every 18 month	200 - 500	1 - 2 weeks
D-Check	every 6 - 10 years	30.000 - 50.000	4 - 8 weeks

It is essential to highlight that due to the complexity of the aviation industry both, line and base maintenance require a complete and dedicated organization to ensure the performed services comply with all the quality and safety requirements. Lots of effort are required to plan the maintenance activities properly, allocate the dedicated resources, monitor the progress of the scheduled activities, control costs, and document everything. Another fundamental aspect of maintenance is that the maintenance organization needs to be certified by local or international authorities. A brief overview of regulations for repair stations will be covered in the next section.

Regulation for Repair Stations

As soon as an aircraft enters operation after being released from its manufacturer, its maintenance is ruled by local authorities to always remain airworthy throughout the process. In general, every country or group of countries have their regulator for addressing aircraft maintenance and repair stations. In table 4 below, we can find a specific government or group of countries, its regulator, and the regulation which governs the repair stations.

Table 4- Location, Regulator, and Regulation for Repair Stations

Location	Regulator	Regulation
Brazil	ANAC	RBAC part 145
USA	FAA	FAR part 145
Europe	EASA	EASA part 145

Although each country or group of countries have their regulators, in practical terms, due to the intrinsic global characteristic of aviation, Brazilian Repair Stations Quality Departments often need to deal with international regulators such as FAA and EASA. The rules which govern aviation in each country can be found in each of the respective regulators for that country website.

The content addressed by RBAC 145 is vast, and any MRO which aims to service any Brazilian airline shall comply with it. Table 5 illustrates some of the aspects covered under RBAC 145.

Table 5- Content Addressed by RBAC 145

GENERAL
Applicability
Definitions
The requirement for certificate and operational specifications
CERTIFICATION
Requirements for certification
Issuance of certificate
Validity and certification renewal
Amendment of certificate transference
Categories and Classes
Limitations of certificate
INSTALLATION, RESOURCES, EQUIPMENT, TOOLING, MATERIALS, TECHNICAL DATA
General
Requirements for installations and resources
Changes of location (address), installation or resources
[Reserved]
Requirements for equipment, tooling, materials, and technical data
PERSONNEL
Requirements for personnel
Requirements for supervision personnel
Requirements for inspection personnel
Requirements for personnel authorized to approve an item to return to service
[Reserved]
Administration, supervision, and inspection personnel records
Training requirements
The danger of goods training
OPERATIONAL RULES
Certification prerogatives and limitations
Activities performed at other locations
Maintenance execution, prev. Maintenance, alteration for cert. owner based on

RBAC 121 / 135
[Reserved]
Maintenance organization manual
Content of maintenance organization manual
Quality control system

As we can see, the content covered by regulators is vast. This vast content covered by regulators is essential to assure that only qualified shops can be able to research on aircraft and engine maintenance. As a result, only a few organizations in the world are capable of servicing engines due to the stringent requirements to do the job and, therefore, large financial investments needed.

Engine Repair Landscape

According to Market Research Future, the construction of large-level aircraft in the construction of turbofan jets is limited to 4 large companies, thus sharing 93% of the manufacturing. (The Economist June 1, 2019). Similar situations, to a lesser extent due to the development of independent airline shops or independent repair stations, happens with engine MROs.

At this time, the older engine fleets are steadily declining due to the natural evolution of new-generation aircraft and engines introduced to the market. As a result, the older machines tend to consume more and reduce interest in use as well as reprocessing of parts to repair (Aircraft Commerce, 2018). However, although there is a massive increase in new-generation aircraft and engines, there is still a demand, for the time being, the repair and overhaul of older engines such as the Pratt & Whitney JT8D-200; PW JT9D, PW2000, PW4000-94, Rolls-Royce RB211-535, CFM56-3, and -5 series GE CF6-50 and -80C2 (Aircraft Commerce, 2018). As the new-generation engines have come online, OEMs have increased their presence in the maintenance repair and overhaul (MRO) market (Aircraft Commerce, 2009). Also, many of the significant maintenance providers for the older engine types leave the market in favor of pursuing the maintenance market for younger engines (Aircraft Commerce, 2018).

OEMs that undertook MRO services did so in a big way, with engine shops all over the world. It is fair to say that at least 30% of the global engine shops are a joint venture in some way with engine manufacturers (Aircraft Commerce, 2009). A few major airlines still have their shops for overhauling engines, such as Lufthansa, Air France Industries, Delta TechOps, and United Services. These airlines have developed in-house repair and overhaul capability for all or most of the engine types in their fleets and offer their maintenance, repair, and overhaul (MRO) services to other operators(Aircraft Commerce, 2014).

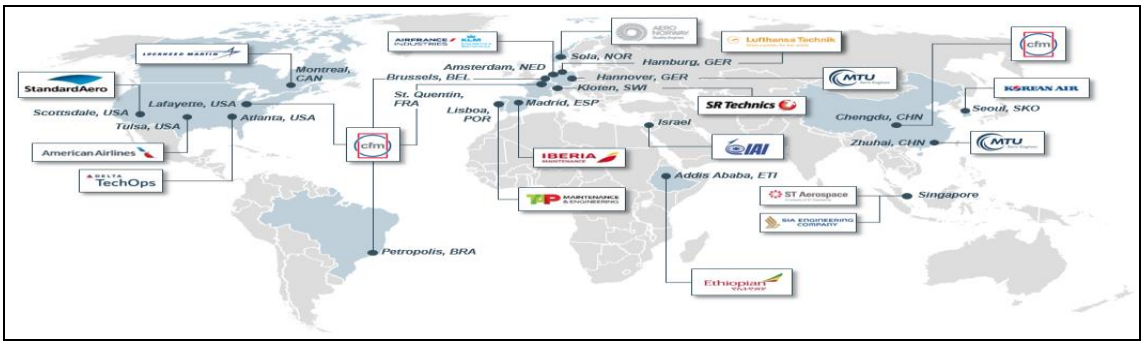
Not only do some have a leading shop at their base, but a few also have diversified into other continents. Lufthansa, for example, through JVs and wholly-owned subsidiaries, has many engine shops in Europe, Asia Pacific, and North America. About 50% of the global engine shops are connected to an airline. About 75% of international engine shops are connected to either an airline or an OEM. Some, such as TAESL and HAESL, is connected to both. This leaves just a quarter of engine shops in independent hands. The largest of these independent companies include MTU Maintenance, ST Aerospace, and SR Technics, all of which have invested in facilities on three continents. (Aircraft Commerce, 2009).

Research by Oliver Wyman shows OEMS 53% of engine maintenance services are always the same companies that did them, Chandler (2018, 03). Although there are a few

independent engine MROs in the world, they still depend on OEMs for the delivery of parts to be installed on the engines repaired at their facilities.

One important consideration is that although the number of engine MROs globally in quantity may seem to be high, very few of them are capable of repairing several different engine types. Figure 4 illustrates the main engine MROs in the world capable of overhauling the CFM56-7. This picture helps to understand that Brazilian airlines do not count on a privileged position for repairing their engines. Brazil counts with just one MRO in-country, located in Petropolis - RJ. As a result, depending on the type of engine which needs repair, it needs to be done abroad.

Figure 4- CFM56-7 Main MROs

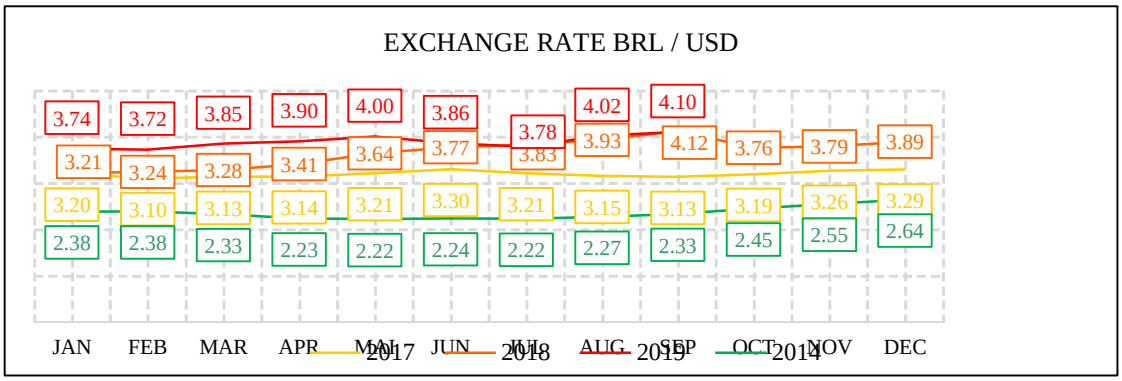


Another point that deserves attention is that even if an airline opts to repair its engine in Brazil, its repair cost will be dollarized. As a result, Brazilian airlines need to take into account currency exchange fluctuations.

Brazilian Current Exchange

Global economic forces - such as interest rates in the United States and European Union (Eidelchtein, 2009) - and internal politics uncertainties - such as politics succession perspectives, difficulties in approving economic reforms, taxes condition (Eidelchtein, 2009) - are the main drivers for the influx and outflux of dollars to and from Brazil. These forces cause the Brazilian currency to fluctuate. Based on the numbers extracted from Domestic Airfares 1st Quarter 2019. (ANAC, 2019) And Brazil Central Bank (Banco Central do Brasil, 2019) we can observe in figure 5 significant variations on current Brazilian exchange along the past five years. In practical terms, if an engine repair that was done in September 2014 were to be done in September 2019, it would cost 77% more.

Figure 5- BRL / USD exchange rate, ANAC (2019) and Brazil Central Bank



Engine Maintenance Cost

To provide aviation mechanism indicators worldwide, IATA (MCTF) collects data and all companies around the world annually and methodologically to define how much it costs for an airline to maintain its fleet for decision-making. In 2017 global spending on MROS was $ 76 billion, disregarding general costs. That means 11% of companies' operating costs. Estimating an annual increase of 4% per year, this could reach $ 118 billion by 2027.

Customs Clearance, Import, Export

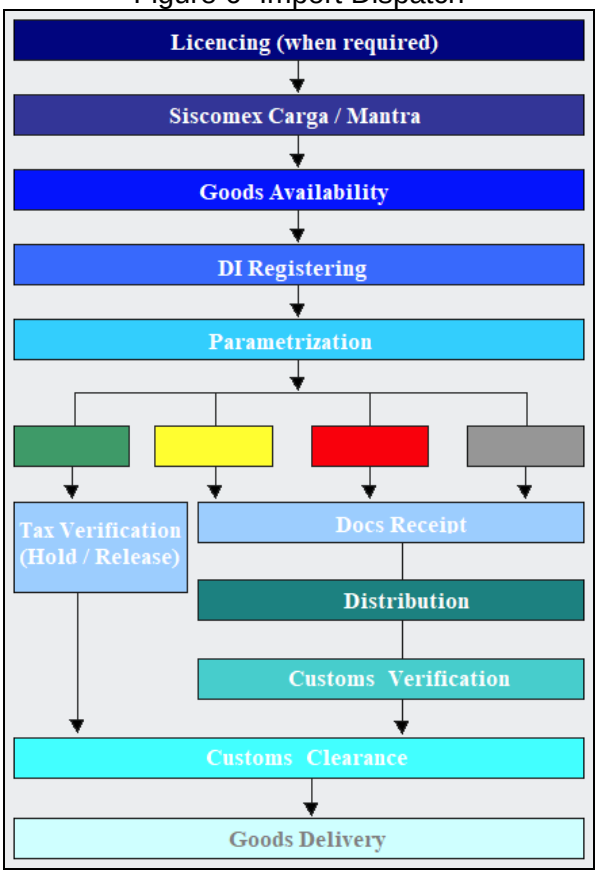
The management of customs activities, the monitoring, control, and taxation of foreign trade in Brazil is ruled by law decree 6.759 from February 5, 2009. This law decree is signed by the Brazilian president and issued by Ministério da Casa Civel (Civil House Ministry). The document provides all instructions and references for all kinds of foreign trade allowed to be done in Brazil.

An important governmental institution which heavily participates in the import and export of goods in Brazil is Receita Federal do Brazil (RFB). Receita Federal is responsible for the administration of federal taxes and customs control and acts to combat tax evasion (smuggling), embezzlement, piracy and drug, and animal trafficking.
After complying with the legal and commercial requirements of the country in force, any operation is defined by importation, whether it facilitates the entry of goods into a customs territory or not. (Assumpção, 2007). The import and export are done through Siscomex, an integrated and information system used by customs to receive all the data related to the goods to be imported and exported. Siscomex was created by law decree 660 from September 25, 1992, to register, follow up, and control through a single flow of information the import and export processes (Eidelchtein, 2009). The System is managed by the Foreign Trade Secretariat of the Ministry of Industrial Development and Trade (MDIC / SECEX), the Central Bank of Brazil, and the Federal Revenue Secretariat of the Ministry of Finance (MF / SRF), in their responsible areas of activity, respectively commercial, foreign exchange and tax. (Faro, 2007). The importation is the moment that configures the release of the goods; it is the established place for delivery of documents that must be presented with established deadlines. (Gama, 2013). Imports into Siscomex are processed in several stages to be performed by the importer, the depositary, the customs inspection, and the carrier.

It is the responsibility of the international maritime carrier, in the Siscomex Cargo module, to provide RFB with information about the vehicle and the domestic, foreign, and transit cargo carried therein, for each call of the vessel in a customs port. In the case of air modal, the electronic manifest must be informed in the Mantra system. The depositary is responsible for immediately reporting the RFB of the availability of the cargo collected in its custody. It is up to the importer to register the DI (Import Dispatch) or DSI (Simplified Import Dispatch) in the

System. Once analyzed, Siscomex will identify the configuration channel where it will remain until it obtains the customs authority, if applicable. In this place, there are four surveillance channels; they are Red, Green, Yellow, and Gray. (Gama, 2013). It is up to the customs control to check the customs and the clearance. Figure 6 illustrates not only the various dispatch steps but also other steps that are part of the import process.

Figure 6- Import Dispatch



Below is a brief description of the five possible parametrization channels:

Green: the System records the automatic clearance of the goods, exempting the documentary examination and the physical verification of the merchandise (RFB, 2014). The ID selected for the green channel in Siscomex may be subject to physical or documentary verification when evidence of import irregularities is identified by the RFB agent responsible for this activity (RFB, 2014);

Yellow: a documentary examination must be performed and, if no irregularity is found, the customs clearance has been made, and the physical verification of the goods is not required (RFB, 2014). In the event of the incomplete description of the goods in the DI, which requires physical verification for their perfect identification to confirm the correctness of the tax classification or declared origin, the RFB agent may condition the conclusion of the documentary examination to the physical verification of the goods (RFB, 2014);

Red: the goods are only cleared after the documentary examination and physical verification of the goods (RFB, 2014);

Gray: the documentary examination, the physical verification of the merchandise, and the application of a particular customs control procedure should be performed to verify signs of fraud, including the declared price of the merchandise (RFB, 2014).

Unfortunately, import and export processes in Brazil are occasionally impacted by frequent strikes, which bring substantial losses to the Brazilian industry and increase the risks to do business with the country. Most of the time, government agents halt their research for undetermined periods claiming better salaries and benefits. As an example, the last strike was initiated in November 2017 and endured until late July early August 2018; the estimated losses are around BRL 10 million per day (Terra, 2018). As a result of strikes, the import and export of goods, including engines, take longer than usual, increasing their turnaround time.

Homologation

The legal obligations and certification documents for a maintenance organization under EASA and FAA regulations are broadly RBAC 145 (ANAC) and FAR 145 (FAA) regulations. Among these documents, there are no significant differences as RBAC 145 is much a translation of FAR 145 into Portuguese. For the approval of a research shop, the approval steps are divided into 5 phases:

1. Certification Contents of the IS (Supplementary Instruction - IS 145.001 (d)) document, especially the section between sections 5.1 and 5.5 that describes all the steps from the preparation of the initial certification request to the issuance of the certificate/operative specifications.

2. Training: Contents of the IS (Supplementary Instruction - IS 145.010) document (a), especially sections 5.2.1 (a), 5.2.5 (c) and the passage between sections 5.2.5 (a) to (e), which describe the types of training required and how to design a Maintenance Organization Training Program. The FAA has a document called.

AC 145-10 - Repair Station Training Program that presents concepts similar to those described in ANAC IS.

3. Tools: Contents of the IS (Supplementary Instruction - IS 145.009 (b)) document, especially the section between sections 5.5.2.1 (c) through (e) describing how an organization should demonstrate the ability to perform maintenance services based on tools that you have or have access to. The FAA has a document called.

AC 145-9A - Guide for Developing and Evaluating RS / QCM that features virtually the same content as ANAC IS

4. Manufacturers own releases, etc.

For all purposes, a release from the article manufacturer is not required for a Maintenance Organization to perform maintenance on it, but in practice, by IS (Supplementary Instruction - IS 145.109-001 (c), a Maintenance Organization is required to prove which has access to the technical data issued by the manufacturer of the article to be maintained either by contract with the manufacturer itself or by assignment by the owner or operator of the item (ref .: sections 5.3, 5.4 and 5.5 of the document).

METHODS

As an airline runs, it accumulates historical data about its fleet. This historical data encompasses operational as well as maintenance records. Typical operational data for the engines comprise EGT, flight cycles, flight hours, fuel consumption, oil consumption, and vibration level. Specific maintenance data contains repair records, maintenance check

records, among others. From time to time, airlines can benefit from reviewing their historical data as valuable information may be extracted. If carefully reviewed, this historical data may allow good decision-making for processes improvements and cost reduction.

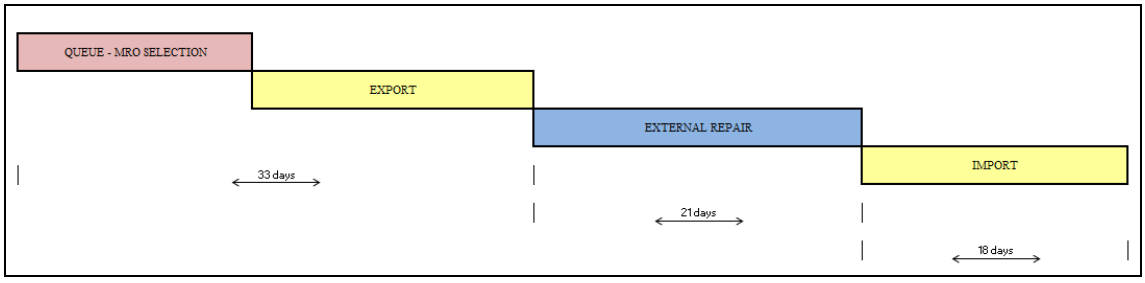
A group of engineers, by analyzing engine maintenance history, identified a repair, which was a potential opportunity to be internalized. There were 12 repairs under research done in the last four years resulting in an average of 4 repairs per year. Evaluation of the engine fleet condition indicated that at least 15 repairs should be done in the next five years as a minimum. The research, among other things, focused on alternatives to improve the engine turnaround time, therefore, enabling the engine to be available for more extended periods. This shorter turnaround time may, in the long run, reduce the company's spare engines and, consequently, the operational cost. To better understand the challenges that the repair presents if done in house, the engineers visited a specialized MRO and watched all the details involved in the repair process. During the visit to the MRO, they could better evaluate all the tooling required, the research force, as well as the risks involved in the operation. The resulting analysis was split into two: one to evaluate the turnaround time and others to assess the costs involved in the repair.

DATA SOURCE, COLLECTION, AND ANALYSIS

Turnaround Time, Costs, and Associated Expenses for External Repair

The repair under research, when done in a specialized MRO, based on historical records, is, on average, accomplished in 21 calendar days. However, often, due to MRO capacity, the engines need to wait a few days to be inducted and have its repair initiated. Although there are variations in the queuing time, on average, a machine needs to wait around 33 calendar days to be inducted. The time an engine needs to wait to be seated is undesirable because it extends the period under which the engine is unserviceable. The typical turnaround time for sending an engine to be repaired abroad (the United States or Europe) is, on average, 18 for export and 17 for the import processes under normal circumstances. It is important to remember, however, that import and export processes may suffer extensions from 10 to 15 days during periods when the Brazilian customs are on strikes. Figure 7 below illustrates the entire repair process for a typical engine.

Figure 7- Repair Cycle Turnaround Time



Historical records show that the repair under research costs, on average, USD 94.419,48 for labor and USD 203.375,08 for material, totalizing USD 297.794,56. Logistic, on average, accounts for USD 12.500,00 on the export phase and USD 17.500,00 on the import phase. It is essential to highlight that values for logistics vary significantly depending on the volume and weight of the transported product. As a result, due to the different engines on the market, transportation costs vary considerably due to their size and weight differences. Another aspect to be considered when evaluating transportation costs is the volume of goods to be transported under a specific contract. Usually, the higher the work under the transportation contract, the higher the chances to do a better deal. Additional expenses involved in the export and import processes, such as storage of the goods during custom clearance, fees, and customs broker as well as corresponding taxes to accomplish the

operation, are shown in Table 6 below. No insurance was hired to transport the engines because they are under the overall airline insurance. However, it represents a high cost and shall be considered for evaluating the repair viability if applicable:

Table 6- Repair Costs and Associated Expenses (BRL)

Export		Import	
Custom Broker	1.388,72	Customs Broker	1781,42
Insurance	0,00	Insurance	0,00
Freight	49.000,00	Freight	68.600,00
Fees	244,00	Fees	244,00
Storage	3.871,92	Storage	4.780,96
		Material	**797.230,32**
		Tax - COFINS	7.972,30
		Tax - ICMS	35.539,84
		Labor	**370.124,36**
		Tax - ICMS	16.499,82
		Tax - IRRF	65.316,22
		Tax - CIDE	43.545,13
Total Export	54.504,64	Total Import	1.411.634,38
Total Export & Import			1.466.139,02

Turnaround Time and Costs Internal Repair

The careful research of the engine and aircraft manual plus the visit to the specialized MRO enabled the engineers to identify all the steps involved in the process, the corresponding tooling necessary to do the job, the adequate personnel to do the job, the expected turnaround time to accomplish the repair in house and the risks involved in the operation. One crucial observation is that due to the complexity of the repair process, it is expected that the mechanics shall pass through a learning curve to achieve enough familiarization and expertise to be able to perform the repair as fast as a specialized MRO. The first repairs to be done in the house are estimated to be accomplished around 50 days (more than twice than a specialized MRO). As time passes and the mechanics accumulate experience, this number decreases and establishes about 28 days (seven days longer than a specialized MRO). The learning curve, as well as a comparison between the time it takes to perform the repair in an external MRO and internally, can be observed in Table 7 below:

Table 7- Comparison for Turnaround Time in Days Between External Repair and Internal Repair

Year	Repair #	External Repair				Internal Repair	Saving
		Export + Queue	Repair	Import	Total	Repair	
1	1	33	21	17	71	50	21
1	2	33	21	17	71	50	21
1	3	33	21	17	71	46	25
2	4	33	21	17	71	46	25
2	5	33	21	17	71	42	29
2	6	33	21	17	71	42	29
3	7	33	21	17	71	38	33
3	8	33	21	17	71	38	33
3	9	33	21	17	71	34	37
4	10	33	21	17	71	34	37
4	11	33	21	17	71	30	41

4	12	33	21	17	71	30	41
5	13	33	21	17	71	28	43
5	14	33	21	17	71	28	43
5	15	33	21	17	71	28	43
Total =	15	495	315	255	1065	564	501

The investment in material resources (fixtures and dedicated tooling) is USD 1.051.360,68. The facility (rent, electricity, etc.) costs are BRL 28.350,00 per month. The proposed headcount for a local repair shop is one supervisor, two lead mechanics, and two senior mechanics. An estimating cost for human resources is BRL 81.242,28 per month. The training to certify the personnel to do the job is USD 89.000,00. It is essential to highlight that this research considered the average exchange rate that occurred from January 1, 2019, to October 13, 2019 (1 USD = BRL 3,92) provided by Banco Central do Brasil (BACEN, 2019).

Based on the researched values for the length of an internal repair, it was possible to estimate its labor cost. Considering that it will be performed in 50 days, its price will be the daily labor rate multiplied by the number of days it takes. As a result, the labor cost for the first repair is BRL 135.403,80. Material costs do not change once the replacement parts are all the same and supplied by the engine OEM. Table 8 shows the values for the first repair done in the house. One crucial observation is that according to Brazilian regulation IN RFB N1700 March 14, 2017, the total investment in tooling may be depreciated within five years.

Table 8- Cost (BRL) for Internal Repair

Export		Import	
Custom Broker		Custom Broker	
Insurance		Insurance	
Freight		Freight	
Fees		Fees	
Storage		Storage	
		Material	**797.230,32**
		Tax - COFINS	7.972,30
		Tax - ICMS	35.539,84
		Labor	**135.403,80**
		Tax - ICMS	
		Tax - IRRF	
		Tax - CIDE	
Total Export	0,00	Total Import	976.146,27
Total Export & Import			976.146,27

RESULTS

We can observe from table 6 that the external repair cost (797.230,32 + 370.124,36 = 1.167.354,69) is increased (to 1.466.139,02) by 25,6% due to associated expenses related to the import and export process as well as the taxes incurred. We can also observe by

comparing Tables 6 and 8 that the internal repair cost for the first repair (BRL 976.146,27) is significantly lower (33%) than the external repair cost (BRL 1.466.139,02). To better visualize the cost differences, we combined the tables in Table 10 below.

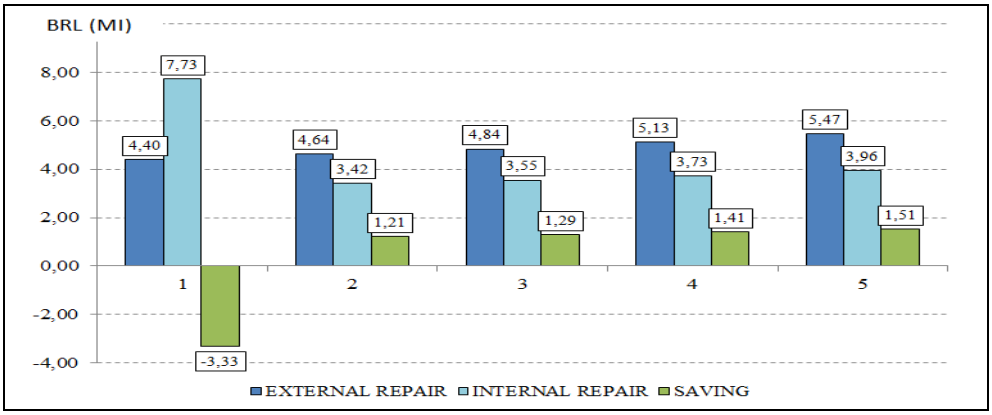
Table 10- Cost Comparison Between External and Internal Repair

External Repair				Internal Repair			
Export		Import		Export		Import	
Custom Broker	1.388,72	Customs Broker	1781,42	Custom Broker		Custom Broker	
Insurance	0,00	Insurance	0,00	Insurance		Insurance	
Freight	49.000,00	Freight	68.600,00	Freight		Freight	
Fees	244,00	Fees	244,00	Fees		Fees	
Storage	3.871,92	Storage	4.780,96	Storage		Storage	
		Material	**797.230,32**			**Material**	**797.230,32**
		Tax - COFINS	7.972,30			Tax - COFINS	7.972,30
		Tax - ICMS	35.539,84			Tax - ICMS	35.539,84
		Labor	**370.124,36**			**Labor**	**135.403,80**
		Tax - ICMS	16.499,82			Tax - ICMS	
		Tax - IRRF	65.316,22			Tax - IRRF	
		Tax - CIDE	43.545,13			Tax - CIDE	
Total Export	54.504,64	Total Import	1.411.634,38	Total Export	0,00	Total Import	976.146,27
Total Export & Import			1.466.139,02	Total Export & Import			976.146,27

This lower cost for the first repair tends to be even lower as time passes, and the team gains experience to execute it. The cost reduction between internal and external repair is the result of a lower cost of manpower, no expenses for logistics, and lower incurrence of taxes.

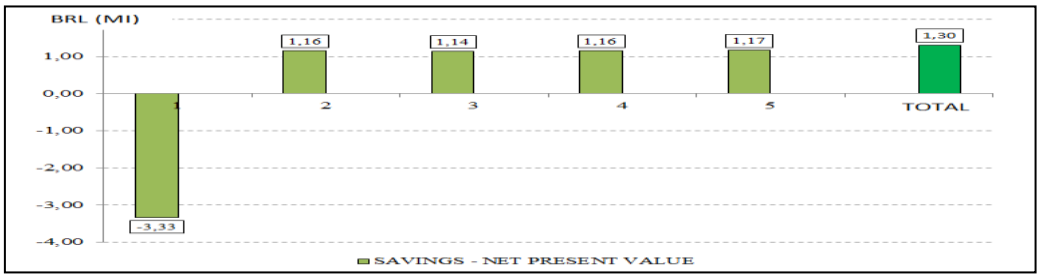
Although there are indications that the repair in the house is more economically attractive, more in-depth research is necessary to confirm if all the required investment is worth it. This research comprises evaluating the cash flow for both scenarios, internal and external repair, in the next five years, considering the expected volume of research to be accomplished per year. Another critical point to be taken into account is the economic adjustments that need to be researched in the future, considering inflation for the several sectors involved in the repair process, current exchange, interest rates, and all the incurred costs or expenses along the period. Values for the currency exchange rate and inflation (IPCA) for the next five years are based on research made by Banco Central do Brasil (BACEN, 2019). Material and labor escalation is based on industry practices. Custom broker and fees, due to a conservative approach, are considered unaltered throughout the period. The result of the analysis can be seen in figures 8 and 9.

Figure 8- Cash Flow for External and Internal Repair



The net present value of the accumulated savings of each year, as well as the total net current value for the entire period of 5 years, can be seen in figure 9 below:

Figure 9- Present Net Value for Accumulated Savings



By comparing both cash flows, we confirm that if the repair is internalized, the airline may have a saving BRL 1.3 million.

Table 13- Present Net Value (BRL) for Accumulated Saving the Next Five Years if the Repair is Internalized

External Repair - Cash Flow / Year	Internal Repair - Cash Flow / Year	Saving	Discount Rate	NPV
4.398.417,05	7.727.820,37	-3.329.403,31	5,91%	-3.329.403,31
4.637.235,23	3.424.887,16	1.212.348,07	4,82%	1.156.599,95
4.841.804,35	3.554.399,27	1.287.405,08	6,16%	1.142.334,78
5.134.408,12	3.727.642,11	1.406.766,02	6,61%	1.160.988,82
5.473.266,37	3.960.010,33	1.513.256,03	6,66%	1.169.246,55
	Total =			1.299.766,78

Apart from financial gains, the airline may benefit due to the shorter turnaround time for the repair process. This can be seen in figure 10 below.

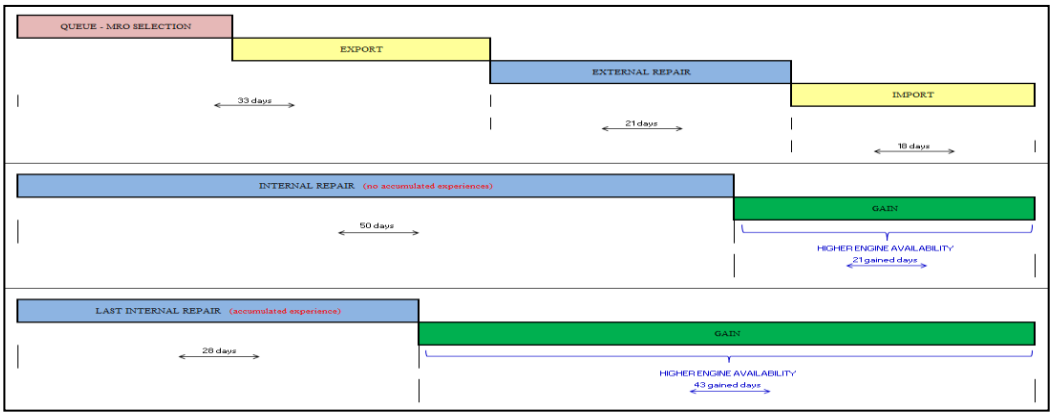


Figure 10- Turnaround Time Saving Due to Internal Repair

CONCLUSIONS AND RECOMMENDATIONS

The effort for evaluating if airlines can benefit from internalizing repair was, the internalization of a light repair is a very strategic choice for any Brazilian airline. If properly evaluated, it can provide financial as well as operational gains.

Airlines shall evaluate its engine repair fleet and historical data. By doing so, they can identify significant cost reduction and operational gains.
Key Lesson Learned: Taxes play a significant role in engine repair. If internalized, substantial savings may be obtained.

REFERENCES

Ackert, S. P. (2017). Engine Maintenance Concepts for Financiers - Elements of Turbofan Shop Maintenance Costs.

Aircraft engine & parts manufacturing - quarterly update 4/15/2019. (2019). (). Fort Mill, South Carolina: Mergent.

Airline Maintenance Cost Executive Commentary, An Exclusive Benchmark Analysis FY 2017 data by IATA's Maintenance Cost Task Force (2018). Retrieved July 14, 2019, from https://www.iata.org/whatwedo/researchgroups/Documents/MCTF/MCTF-FY2017-Research-Public.pdf

An Initial Appraisal of the Agreement between CFM & IATA (2019). Aircraft Commerce, 124, 50 - 66

ANAC (2019). Tarifas Aéreas Domésticas 1º Trimestre de 2019. Retrieved from https://www.anac.gov.br/assuntos/setor-regulado/empresas/envio-de-informacoes/tarifas-aereas-domesticas-1/relatorio-de-tarifas-aereas-domesticas-nacional

ANAC (2017). Regulamento Brasileiro da Aviação Civil, RBAC nº 145 Emenda nº 02. Retrieved from https://www.anac.gov.br/assuntos/legislacao/legislacao-1/boletim-de-pessoal/2018/6s1/anexo-ii-rbac-no-145-emenda-no-02

Banco Central do Brasil (2019). Exchange Rate BRL / USD. Retrieved from https://www.bcb.gov.br/acessoinformacao/legado?url=https:%2F%2Fwww4.bcb.gov.br%2Fpec%2Ftaxas%2Fport%2Fptaxnpesq.asp%3Fid%3Dtxcotacao

Chandler, J. G. (2018). MRO market for the next ten years is "basically bullish," but challenges
remain. Aircraft Maintenance Technology, 29(3), 28-37.

Engine Shop Visit Demand & Capacity (2001). Aircraft Commerce, 018, 25-33

Fritzsche, R., Gupta, J. N. D., & Lasch, R. (2014). Optimal prognostic distance to minimize total maintenance cost: The case of the airline industry. International Journal of Production Economics, 151, 76-88. doi:10.1016/j.ijpe.2014.02.001

Global Engine Activity Survey, (2009). Aircraft Commerce, 064, 53 – 66

Global Engine Maintenance Market 2014 (2014). Aircraft Commerce, 094, 35 – 46

Gubisch, M., & Hemmerdinger, J. (2018). CFM warranty change could adjust the sums on its parts. Flight International, 194(5656), 18-19.
Hinsch, M. (2018). Industrial aviation management: A primer in European design, production and maintenance organizations.

Kinnison, H. A. (2004). Aviation maintenance management. New York: McGraw-Hill.

Kinnison, H. A., 2012. "Aviation Maintenance Management." McGraw-Hill, New York, NY.

MRO Options Older & Mature Engine Types (2018). Aircraft Commerce, 118, 36 – 42

Scheinberg, R. (2017). The commercial aircraft finance handbook (First; Second; ed.). Milton:
Taylor and Francis. doi:10.4324/9780203713303.

The jet set. (2019). The Economist, 431 (9145), 6-7.

The United States. Flight Standards Service. (2018). Aviation maintenance technician handbook:
Volume 1, volume 2.powerplant. Oklahoma City, OK: U.S. Department of Transportation, Federal Aviation Administration, Airman Testing Standards Branch.

Vieira, D. R. (2016). Maintenance, Repair, and Overhaul (MRO) Fundamentals and Strategies: An Aeronautical Industry Overview.

Viles, E., Puente, D., Alvarez, M. J., & Alonso, F. (2007). Improving the corrective maintenance of an electronic system for trains. Journal of Quality in Maintenance Engineering, 13(1), 75-87. doi:10.1108/13552510710735131

DECISION SCIENCES INSTITUTE
Productivity management in waste recycling processes in South Africa

Barin N. Nag
Monarch University, Switzerland
Towson University, USA
bnag@towson.edu

Brian Tlakula
Monarch University, Switzerland
Brian.Tlakula@umonarch-email.ch

ABSTRACT

Waste generation and management are challenges in South Africa because of rapid population growth and increased industrialization. The mandate of government and waste management is to decrease waste disposed in landfills and promote recycling and re-use reducing greenhouse emissions and air pollution. A review of literature confirms a knowledge gap regarding the operations management concepts of productivity in waste recycling. The paper focusses on Nokeng to improve productivity in waste recycling processes. The paper concludes that the selected variables of inventory management, process design, quality management, time-oriented systems and HR management are benefits fundamental to improving productivity of waste recycling.

KEYWORDS: Productivity Management, Waste Recycling, Time-Oriented Systems, Inventory, Process Design, Quality Management, Human Resource Management

INTRODUCTION

Waste generation and management are currently a serious challenge in South Africa because of rapid population growth and increased industrialization within the sectors of mining, manufacturing and construction (Ginindza & Muzenda, 2016). This kind of predicament has a negative effect on the environment and wellbeing or health of South African citizens. Scholars reveal that waste managed improperly increases the emissions of greenhouse gases (see for example De la Barrera & Hooda, 2016).

Waste may be defined as operational components that do not add value to the products and are rejected (Nag, 2019). The largest waste quantities are generated from households and commercial businesses. Hence, the commitment of government and the waste management sector is to decrease waste disposed in landfills and focus more on recycling and re-use in order to prevent emissions and air pollution. This implies a shift away from landfilling towards waste prevention, reuse, recycling and recovery.

According to Godfrey & Oelofse (2017) only 10.8% of waste is recycled or re-used, while almost 90% ends up in landfills. Recycling is described by Ahsan *et al* (2014) as “the reprocessing of wastes into either the same product or different product and it is the key mechanism to recover products which is useful in reduction of waste quantity. In addition, Momoh & Oladebeye (2010) state that recycling processes provide the required raw material to industries and are found to be an efficient and effective method of solid waste management. It is the mandate of Nokeng Refuse Removal (Pty) Ltd (hereafter referred to as Nokeng) as enacted by the government’s legislation to ensure that more waste is effectively re-used, recycled and recovered. It is important to briefly reflect on the policy and regulatory environment of waste management in South Africa.

There have been several government policies or legislation between 1989 and 2017. The Environmental Conservation Act (Act 73 of 1989) was aimed at permitting, controlling, and managing waste disposal sites with the objective of reducing environmental impacts related with many poorly operated landfills whether controlled or uncontrolled. Following this the policies published include the National Waste Management Strategy (NWMS) and the White Paper on Integrated Pollution and Waste Management (IP&WM). These policies were meant to create regulations with regards to the reduction, reuse and recovery of waste. From 2008 until 2107, the National Environmental Management (NEM): Waste Act (Act 59 of 2008) and NEM: Waste Amendment Act (Act 26 of 2014) were promulgated. An impetus was created for a number of regulations for waste management and the secondary resources industry (Godfrey and Oelofse, 2017). However, even with the government's legislation in place, Nokeng is experiencing a challenge of effectively managing waste due to gatekeeping by local municipalities. There are few waste management companies operating in the formal economy. The policies mentioned do no effectively support the formal sector of waste management, and Nokeng is experiencing operational challenges with respect to time, process design and capacity building, inventory management, quality defects, and human resource management.

Operations Management literature shows that when there are time, inventory and quality problems there's a loss of output, waste of labour or human, materials and time resources (Nag, 2019). Quality defects play a critical role in production, and failure to meet quality standards can be disruptive having negative effects on revenue and profitability. Operations and/or productivity are fundamental aspects of waste management, particularly the productive processes of waste or raw materials (Teixeira et al., 2014). A review of South African literature confirms that there is a knowledge gap regarding the role played by operations management on the productivity of waste recycling processes.

The purpose of this paper is to develop a conceptual plan for Nokeng to improve productivity in waste recycling processes. The specific research objectives are: 1) to evaluate factors affecting productive performance at Nokeng; 2) to identify possible ways or systems of improving productivity at Nokeng in relation to operations management mainly focusing on process design and capacity building, inventory management, quality management, time-oriented management and human resource management; and 3) to develop recommendations on how to improve productivity of waste recycling processes at Nokeng. The results of this paper will not only benefit Nokeng but will also support the decision-making processes made by waste management agencies and policy makers.

THE IMPORTANCE OF OPERATIONS AND/OR PRODUCTIVITY MANAGEMENT IN WASTE RECYCLING PROCESSES

According to Remigios (2010), an operation is the core of every business organization, and of other organisations as well, including non-profit and government organisations. Remigios also maintains that managing this operational core makes the organisation more competitive in its industry environment by giving it the capacity to deliver a larger output of higher quality and lower price with the use of less input resources. This is the core of increasing productivity (Ghiani *et al.*, 2014). The significance of operations management lies in increasing productivity and making the business organisation more competitive, with increased revenues and market share, a greater profitability, and higher levels of quality and customer satisfaction.

In the context of waste management, where waste is a solid, physical product, Christensen (2010) defines it as a redundancy of material to be discarded as it has no marginal or salvage value. A concern here is that the definition of what concretely is defined as waste depends on the specific situation and owner's perception in addition to time,

location, state and personal preferences. According to Cadambi (2011), operations management is concerned with converting materials and labor into goods and services as efficiently as possible to maximize the profit of an organization.

According to Madu (2007) the most recognized environmentally friendly way of dealing with organizational waste is recycling, and recycling is defined it as the process of converting waste materials into resources. Recycling is based on separating recyclable material from other produced waste and then getting rid of the non-recyclable material and finding new purposes for the recyclable material. Not only does it decrease the amount of waste brought to landfills and incineration, it is also profitable for organizations. By recycling already existing, used material, organizations can save on costs of new material. In addition, it enables them to reduce the cycle time of introducing new products and processing time and they are more aware and in control of their supply chain.

It should be noted that the amount of materials used on a product and the way the product has been built from these materials affects its recyclability and the technology needed for it, separation of parts and the quality of the recycled materials (Sas *et al*., 2015). It must also be taken into consideration that some materials are very difficult to separate from each other, making the recycling process a lot more difficult, if not even impossible. Such materials include different kinds of plastics. In such cases, the benefits of recycling and/or separating the product materials from each other might result in low quality products, which in turn diminish the economic benefits in such measures that the product might not be worth recycling anymore.

It may be seen that the only option to get any benefits of the product is to use it as a resource of energy by burning it. Therefore, it is wise for an organization to design their products so that they are easily recyclable in the first place, as it will decrease the cost of the overall process (Sas *et al*., 2015). Further, what is opposing and highly dependent on the manufacturing field and the advancement of the used technology in recycling, is whether the process enables cost savings or whether it increases the price of the product. What should also be considered is that the customers' willingness to pay more for a recycled product than for a normal product is limited (Jonsson *et al*., 2011).

First, a waste management programme can be appropriate for Nokeng only if recyclers and remanufacturers are willing to invest significant capital in commodity industries with extremely thin margins which would enable them to build capacity for this activity. This implies that not only should Nokeng Refusal be engaged in this activity but also be able to access used materials of suitable quality in practical volume, but they would also need to establish longer term agreements to do the same. This would be critical for Nokeng to plan and invest in viable product flows. However, Jonsson *et al*. (2011) contends that it is significant for more research directed at effective management and operation of the waste management phase of the reverse supply chain that would be beneficial in addressing this issue.

Secondly, firms such as Nokeng Refusal that is engaged in waste management need to actively support and perhaps invests in newer processing options when setting up their reuse and recycling programmes. Consider the argument (Jonsson *et al*., 2011) that they need to be involved in creating a set of long term contracts with the waste generators so as to be assured of adequate supply for their activities. For firms such as Nokeng Refusal engaged in waste management activities, the focus of the state should be on how to improve their financial risk either through direct financial support or establishment of mandates for the waste generators which would require them to use the services of the firms engaged in reuse and recycling activities Sas *et al*. (2015).

According to Binnemans *et al*. (2013) effective waste management requires an understanding that the focus should be on addressing longer term issues rather than on merely tackling immediate problems. For example, there is the consideration that the scale of waste incineration facilities should be determined not only by focusing on current waste disposal but also keeping in mind how their use might decline in the longer term due to reuse and recycling efforts. This requires structuring of waste management activities such that the basics of a circular economy are considered (Chen & Chen 2013). Such a process would not only enhance economic, environmental, and social values but also extend product life cycles and facilitate the use of regenerative materials.

MANAGING INVENTORY IN WASTE RECYCLING AS A WAY OF PRODUCTIVITY AUGMENTATION

The quantity of inventory in waste management is particularly dependent on population, economic growth and efficiency of re-use and recycling systems (Chiemchaisri, Juanga & Visvanathan, 2007). This implies that the growth of the economy and population increases waste generation. In South Africa, the inventory of solid waste is high because only 10% of the waste generated is recycled or re-used. Godfrey and Oelefse (2014) argue that the recycling industry in South Africa is still far from maturity. This is confirmed to true because Nokeng is at a backlog in inventory of items to be recycled. Inventory levels are high, and an appropriate system is required in order to resolve this problem.

At the moment there is no inventory management system at Nokeng, thus, this is a serious setback for our organisation because storage holding costs are also high. The EOQ model can assist Nokeng to identify inventory systems in order to effectively manage its storage of recycle material. This is confirmed by Chen & Lin (2008) that waste inventory systems should be put in place as to effectively augment the productivity of recycle items. The EOQ inventory model is suitable for Nokeng because a) it predicts demand of stock in a probabilistic manner; b) it considers the lead time between order time and collection/ of items by suppliers; and c) it regards the importance of re-order points and safety stock in order to meet the demand at hand. This means EOQ inventory control software will be installed with the intention of eliminating the inventory backlog, increasing productivity of recycle materials and maximise profits or revenue (Bischak, Robb, Silver & Blackburn, 2013).

Standard inventory models depend upon time, quantity, and cost. In the special case of single period inventory, an item that must be discarded at the end of the period. It is stated that the optimal stocking level is developed from statistical analyses of costs and desired service levels. Decisions depend on models and formulas already developed, often with a dependence on total cost calculations with a view towards minimizing the total cost, and with the objective of determining the service level to optimally support a process. This paper follows these guidelines and discusses process design and capacity building as important aspects that will assist in improving operational productivity at Nokeng.

PROCESS DESIGN AND CAPACITY BUILDING IN WASTE MANAGEMENT

A key aspect of process design is the transformation of inputs to outputs as part of the value chain where the outputs are of greater value thereby generating the revenue that transformation is determined by revenues and profits and by customer demand and recent concerns over energy efficiency. Capacity is a decision associated with the process decision and the proper strategy is one which matches capacity with demand to achieve customer satisfaction (Braier et al., 2015). A structured process design is achieved by using flowcharts in a method of process reengineering by the use of technology and environmental awareness to achieve financial goals and customer satisfaction (Liang and Liu, 2018).

In the waste management activity of Nokeng, it is important to make sure that there is a structure in managing the waste removal. Firstly, the waste is collected from households and businesses, and then it is transported to the Nokeng depot. At the depot, there are staff members who specifically select recyclable refuse such as cans and plastic materials. The number of active staff members represents capacity. The recyclable waste is loaded in green bins and the unwanted waste is transported to the nearest municipal dumping site. The recycle refuse is sold to businesses that convert it into usable products. This can be explained as the segmentation of the service process. The timeline and capacity play a very important role in this process. Without the capacity of the fifty staff members, Nokeng would not be able to meet the deadline set by the municipality. Thus, the matching of capacity and demand is critical to the organisation.

The recycling process is important to environmental sustainability and is a way of maximizing profit and return on investment. Not only is product and process design essential elements of operations management, but also identifying a suitable location is also of vital importance. Nokeng has pick-up trucks with the latest technologies in removing waste. The technology assists the human capacity in delivering a fast and efficient service. Nokeng also makes use of flow charts in order to organise processes and routing to deliver services on time to ensure that productivity is enhanced, and customers are satisfied. Described below are some of the issues and techniques in the waste management system.

Quality Management as a Tool for Improving Waste Recycling

As stated by Sharma et al. (2017) and elsewhere, quality in the context of operations management is usually defined as consistency in satisfying customer expectations. In the last three decades, quality has been viewed increasingly as a component of strategy to develop markets and revenues. In addition, quality has been recognised as a means of reducing costs by eliminating defects of delivery and in increasing profitability as applied to both manufactured and service products. Research methods have focused on methods of improving quality and reducing defects to a level zero defects (Ghiani *et al*., 2014). Quality performance has become a prime objective in achieving the dual goals of profitability and customer satisfaction.

Time-Oriented Systems as an Important Factor in Waste Recycling Processes

Waiting lines appear in operations when a greater number of consumers arrive at a time period that is insufficient for them to be served at the same time. The managerial implications of waiting lines consist of the following: cost of providing working space, loss of business due to bulking loss of goodwill with unhappy customers, loss of customer satisfaction and loss of future business (Joshi and Ahmed, 2016). Managers of organisations often perceive waiting lines as an inconvenience because it has an adverse outcome on consumer satisfaction and increases operational costs, and cost factors are important because of budget allocations (Kholil, Alfa & Hariadi, 2018).

Nokeng does not have a proper waiting line system and this problem needs to be resolved. Recycling materials are in high demand; therefore, many customers end up waiting in long queues before receiving service. Waiting in line is a serious challenge because Nokeng is losing some loyal customers and as a result our profits are affected. Nag (2019) suggests that a solution to this problem is to design a service system which involves balancing the cost of providing the level of service with the cost of customer waiting in lines for a service. The balance of service and waiting in lines needs to be done between the times of high demand and low demand periods.

There are number of waiting in lines management methods including the queuing theory and simulation. The queuing theory is more complex and mathematical whereas simulation is simpler. Simulation is a computer-generated reconstruction of a waiting line where the customer arrival times and their service times are generated randomly from an established statistical probability distributed from an actual measurement of an actual operation (Paul, Ojekudo & Akpofure, 2015). The simulation method is appropriate in

assisting Nokeng to resolve the problem of long queues of customers in need of recycling materials. The advantage of using simulation is that it can help in changing the characteristics of demand and service with lesser effort (flexibility) in a short time (time compression). It can handle complexities and interference. On the other hand, Islam & Moniruzzaman (2019) contest that the simulation approach has disadvantages of uniqueness which makes it too expensive and it works on a trial and error basis (different results in different runs).

Successful time-oriented management system is characterised by the following:

I. Effective scheduling which compensates the whole supply chain by ensuring fast and reliable movement of goods and services. This system also assists in achieving higher productivity and less operational costs.
II. Delivery system involves faster delivery resulting from good scheduling and response strategy. The outcome of effective delivery systems leads to high customer satisfaction and greater marketability.
III. Scheduling Impact is an obligation to a good delivery system which results in dependability in satisfying customers and filling orders on time. This is a strategic and marketing tool.

Aziz *et al*. (2018) also states that scheduling concepts and methods in a deterministic scheduling environment are dependent on a production volume. A high-volume system in a scheduling environment is a flow system in a production process consisting of repetitive flow of products or services. At Nokeng, there is a mass of recycle materials which are assembled according to the material type in the flow system, for example, aluminium cans, plastics, cardboards and steel.

It is Nokeng's operation manager's responsibility to make sure that idle time, set up costs and job completion time are reduced. And so, management should establish administrative and operational strategies which encompass time-oriented processes. There are three methods of loading: 1) input-output control, 2) Gantt charts and 3) assignment method. In the case of Nokeng, the input-output control is appropriate because it will aid in minimising overloading of recycle materials which are consuming space at the recycle centre. There are also three techniques of managing work flow: a) adjusting performance, b) increasing capacity, and 3) controlling inputs and routing options (Nag, 2019).

The better technique for Nokeng management is controlling inputs and routing options. We can use local sub-contractors to assist in reducing the cluttered material and in transit of the recycle material to customers. This technique will improve efficiency, reduce clutter and overhead costs, improve work-in-process and augment quality. Guerreiro & Soutes (2013) indicate that when unproductive time is eliminated, organisations are able to decrease costs, improve quality, productivity and sustain good customer relations.

A waiting line system is also necessary for Nokeng to operate its recycling centre effectively and efficiently in order to avoid re-occuring of overloading. The waiting line structure or queue discipline can assist in monitoring the trucks queuing to deliver recycle materials to customers. Another possible option for Nokeng is the FIFO (First In, First Out) rule. This technique ensures that the first customer that ordered bulk material will first receive the order (Nag, 2019). FIFO is an effective and unbiased technique in a queuing system with some modifications. As priorities also need to be considered for loyal clients who order larger bulks of recycle materials. In order to prioritise successfully, the simplest way mentioned by Nag (2019) is to separate the queuing systems into two: 1) for general customers and 2) for loyal customers. There will be a specialised distribution or delivery to loyal clients. Thus, loyal customers will receive high priority.

Since the simulation method has a number of disadvantages as already mentioned; the FIFO technique will be the most cost effective for Nokeng. Additionally, Mahto (2014) advises that a planned schedule illustrating the FIFO technique should expose the critical activities (in view of time) that require special attention when a project is to be completed on time. Also, non-critical activities which have slack or floating time should be identified. This can be advantageous when such activities are used effectively and efficiently. It is important

to take into consideration that in order to implement time-oriented systems successfully, human resources should be managed effectually.

Utilising Human Resources in Waste Management to Increase Productivity

According to Armstrong (2016), Human Resource Management (HRM) is the comprehensive management of all aspects of human resources. Human resource is the only resource that involves human beings with all the complexities of needs that human beings have. Some of the most important of these needs are adequate compensation (pay), benefits for self and family, attractive incentives, and recognition for achievements. However, other scholars such as Storey (1996) defines HRM as an employment management technique for the strategic deployment of the workforce to achieve competitive advantage. In addition, Boxall and Purcell (2016) argue that HRM includes anything and everything associated with the management of employment relationships in the firm. The words anything and everything in the definition explains the wider range of issues comprising policies such employment contract and ways in which employees may be involved and participate in areas not directly covered by the employment contract thus ensuring suitable work life (Boxall & Purcell, 2016). Further, it goes beyond employment relations or industrial relations, which personnel management would not have been able to render in organizations.

Nokeng is also like other big organization that has an industry standard human resource management strategy. The organization consist of nine departmental heads who work to make the organization successful. The human resource management strategy consists of the following objectives: 1) managing the labour force; 2) designing suitable jobs, so that the people are productive in their respective areas, and are utilized with effectiveness and efficiency; 3) ensuring incentive and benefit schemes are implemented where necessary.

In any organization, employees are considered the essential strategic asset, and employees could be a competitive advantage if their organization provides more attention and invest in developing employees' skills at the workplace (Zaharie & Osoian, 2013) . At Nokeng, the management team discusses the performance of the employees and gives the needed feedback not only during the scheduled time of appraisal, but also all-round the year. This practice has helped employees to enhance their knowledge, skills, productivity and ability. In accordance to a strategic approach for reward and management, Nokeng has developed reward systems to encourage more productive and hardworking employees.

Retaining employees through strategic policies such as career path planning, pay and benefits; also engaging in operational strategies, such as re-training, re-deployment or relocation in order to make sure employees move in the right perspective by putting in place standards, good reward systems and employee-employer relationships (Mathis & Jackson, 2011). Nokeng requires employees who are highly capable and trained to do their job on time and productively. Nokeng needs to be fully committed towards empowering personnel through intensive training and/or skills development, which brings capacity building within the company and will enhance productivity.

CONCLUSION

In conclusion, this research paper has explicitly discussed the important concepts of operations management and its application to increasing productivity in the waste recycling processes. We recognisedthe operational gaps as an organisation and draw lessons from the identified literature and concepts. Literature also reveals that inventory management, process design, quality management, time related issues and the proper management of HR play a significant role in productivity and operations of waste management organisations. Scheduling and waiting-line systems are also critical aspects of time-oriented management because they contribute towards delivery time, overall costs, and quality.

In summary, the conclusion drawn is that all the five selected variables i.e. inventory management, process design, quality management, time-oriented systems, and HR

management, are fundamental aspects in improving productivity of waste recycling processes. The observations are drawn from the operations of one waste management company in South Africa. The operational characteristics of this company are typical for most waste management operations in South Africa, and the conclusions can be extended to most aspects of most of these operations. Further, the literature supporting this work is drawn from around the world, suggesting that the observations would be similar and resulting in similar conclusions in many environments and countries. An extension of this work would be to analyse economic conditions and incentives with a view towards increasing productivity in South Africa and extending the same analysis to other parts of the world.

REFERENCES

Ahsan, A., Alamgir, M., El-Sergany, M.M., Shams, S., Roason, M.K., & Nik-Daud, N.N. (2014). Assessment of municipal solid waste management system in a developing country. *Journal of Engineering, 1*, 1-11.

Armstrong, M. (2016). Armstrong's Handbook of Strategic Human Resource Management, London: Kogan Page

Aziz, F., Arof, H., Mokhtar, N., Shah, N.M., Khairuddin, A.S.M., Hanaf, E., et al. (2018). Waste levels detection and HMM based collection scheduling of multiple bins. *PLoS ONE, 13(8)*, 22-32.

Bing, X., Bloemhof, J.M., Ramos, T.R.P., Barbasa-Povoa, A.P., Wong, C.Y., & van der Vorst, I.G.A.J. (2016). Research challenges in municipal solid waste logistics management. *Waste Management, 48,* 584-592.

Binnemans, K. Jones PT, & Blanpain B. (2013) Recycling of rare earths: a critical review. Journal of Cleaner Production 51: 1–22.

Bischak, D.P., Robb, D.T., Silver, E.A. & Blackburn, J.D. (2013). Analysis and management of periodic review, order-up-to level inventory systems with crossover. *Journal of Production and Operations management*, *23(5),* 762-772.

Boxall, P. F., & Purcell, J. (2016). Strategy and Human Resource Management (4th Edition). Basingstoke: Palgrave Macmillan,

Braier, G., Durán, G., Marenco, J., & Wesner, F. (2015). An integer programming approach to a real-world recyclable waste collection problem in Argentina. *Waste Manag, 35*, 525-533.

Cadambi, B. V. (2011). Waste management, an operations and strategic performance driver. In 10th International Conference on Operations Management & Quantitative Management organized by AIMS International.

Chiemchaisri, C., Juanga, J.P., & Visvanathan, C. (2007). Municipal solid waste management in Thailand and disposal emission inventory. *Environ Monit Assess*: 1-8.

Chen, C. & Chen,Y. (2013) Energy recovery or material recovery for MSW treatments? Resources, Conservation and Recycling 74: 37–44.

Chen, T., & Lin, C. (2008). Greenhouse gases emissions from waste management practices using Life Cycle Inventory model. *Journal of Hazardous Materials*, *155*: 23–31

Christensen, T. (2010) Solid Waste Technology and Management. Available from:http://site.ebrary.com.libproxy.aalto.fi/lib/aalto/reader.action?docID=10510413 [Accessed on 06 January 2020].

De La Barrera, B., & Hooda, P.S. (2016). Greenhouse gas emissions of waste management processes and options: A case study. *Waste Management & Research, 34(7),* 658-665.

Ghiani, G., Laganá, D., Manni, E., Musmanno, R., & Vigo, D. (2014). Operations Research in Solid Waste Management: A survey of strategic and tactical issues. *Journal of Computers & Operations, 44,* 22-32.

Ginindza, B., & Muzenda, E. (2016). Waste management challenges to opportunities in the West Rand district municipality, Gauteng: Initiatives. Johannesburg, South Africa.

Godfrey, L., & Oelofse, S. (2014). Historical Review of waste management and recycling in South Africa. Resources*, 6(57),* 1-11.

Guerreiro, R., & Soutes, D.O. (2013). Time-based management practices: a study of Brazilian companies. R. Cont. Fin., 24(63), 181-194.

Islam, M.S., & Moniruzzaman, S.M. (2019). Simulation of sustainable solid waste management system in Khulna City. *Sustainable Environment Research, 29(4),* 1-9.

Jonsson, C., Felix, J., Sundelin, A. & Johansson, B. (2011) 'Sustainable Production by Integrating Business Models of Manufacturing and Recycling Industries.' In: Proceedings of the 18th CIRP International 201 Conference on Life Cycle Engineering; Braunschweig, Germany: 2-4 May. Berlin, Heidelberg: Springer. pp. 201-206.

Joshi, R., & Ahmed, S. (2016). Status and challenges of municipal solid waste management in India: A review. Cogent Environmental Science, 2(1), 1-22.

Kaenzig, J. & Wüstenhagen, R., (2010). The effect of life cycle cost information on consumer investment decisions regarding eco-innovation. Journal of Industrial Ecology, 14(1), 121–136.

Kholil, M., Alfa, B.N., & Hariadi, M. (2018). Scheduling of house development projects with CPM and PERT method for time efficiency (case study: house type 36). Bristol, UK: IOP Publishing.

Liang, J., & Liu, M. (2018). Network design for municipal solid waste collection: A case study of the Nanjing Jiang new area*. International Journal of Environmental Research and Public Health, 15(2812)*, 1-15.

Madu, C.N. (2007) Environmental planning and management. Available from: http://site.ebrary.com.libproxy.aalto.fi/lib/aalto/reader.action?docID=10188841

Mahto, D. (2014). Essentials of operations research - Chapter 10: Network techniques (PERT & CPM). Retrieved from https://dx.doi.org/10.2139/ssrn.2887249 on 16 November 2019.

Momoh, J.J., & Oladebeye, D.H. (2010). Assessment of awareness of attitude and willingness of people to participate in household solid waste recycling programme in Ado-Eketi, Nigeria. *Journal o*f *Applied Sciences in Environmental Sanitation, 2,* 1-20.

Morrissey, A.J., & Browne, J. (2014). Waste management models and their application to sustainable waste management. Journal of Environmental Informatics, 17(2), 65-74.

Nag, B.N. (2019). Managing operations in manufacturing, services and e-Business (2nd Edition). New York: Linus Publication Inc.

Nagapan, S., Rahman, I.A., & Asmi, A. (2012). Factors contributing to physical and non-physical waste generation in construction industry. *International Journal of Advances in Applied Sciences, 1(1),* 1-10.

Paul, A.N., Ojekudo, N., & Afpofure, O.N. (2015). Queuing theory and its application in waste management authority (A focus on Lawina Igando dumpsite, Lagos state). *International Journal of Science and Research, 4(9),* 1392-1400.

Remigios, M.V. (2010). An overview of the management practices at solid waste disposal sites in African cities and towns. Journal of Sustainable Development in Africa 12(7), 233-239.

Sas, I., Thoney, K. A., Joiness, J. A., King, R. E. & Woolard, R. (2015) Sustainable fashion supply chain management. Available from: http://link.springer.com.libproxy.aalto.fi/chapter/10.1007/978-3-319-12703-3_1 [Accessed on 06 January 2020].

Sharma, N., Gupta, N., Ahuja, M., Jain, M., & Shah, M. (2017). Study of operations research in waste management. International Journal of New Technology and Research, 3(10), 4-8.

Storey, J. (1996). From personnel management to HRM: the implications for teaching. Asia Pacific Journal of Human Resources, 33(3), 4-14

Teixeria, CA., Russo, M., Matos, C., & Bentes, I. (2014). Evaluation of operational, economic & environmental performance of mixed and selective collection of municipal solid waste: Porto case study. *Waste Management & Research*: 1-9.

Zaharie, M., & Osoian, C. (2013), Job recruitment and selection practices in small and medium organisations. Studia Universitatis Babes-Bolyai, 58(2), 86-93.

DECISION SCIENCES INSTITUTE

Reduction at Source of Product Packaging and Market-Based Instruments: The Moderating Role of the Stringency of Packaging Environmental Regulation

Karima Afif
HEC Montréal
Email: karima.afif@hec.ca

Claudia Rebolledo
HEC Montréal
Email: Claudia.rebolledo@hec.ca

Jacques Roy
HEC Montréal
Email: Jacques.roy@hec.ca

ABSTRACT

We examine the effect of market-based instruments (MBIs) on manufacturers' decisions to produce less packaging at the source, and the moderating effect of the packaging environmental regulation stringency on the relationship between MBIs and packaging weight. We analyze a longitudinal data set from 848 manufacturers. We find an inverted U-shaped relationship between economic instruments and packaging weight. The effect of MBIs and packaging regulatory strictness varies across industries. We provide empirical evidence for the perceptional and behavioral gap between policymakers and implementers of environmental policies. Policymakers should not expect the latter will always produce the same effect in all industries.

KEYWORDS: Reduction at source of product packaging, Market-Based Instruments, Stringency of Packaging Environmental Regulation, Institutional theory, Longitudinal analysis.

INTRODUCTION

Packaging is increasingly seen as a major issue for today's societies. It represents a significant waste stream that accounts for about 30% to 35% of the municipal waste in industrialized countries and about 15% to 20% in developing countries (Tencati et al. 2016: 35; Wiesmeth, 2018: 1004). Most packages either end up in incinerators and landfills or are thrown into nature (e.g. Yamaguchi et al., 2016). The recent closure of the Chinese waste management market, which was the main buyer of recyclable materials, will aggravate the situation. Many thousands of tons of material are likely to accumulate in sorting centers and cause a multitude of environmental, social and economic problems. This challenge leads academic communities, businesses, and governments to reflect on the environmental impact of packaging (e.g., Wiesmeth et al., 2018; Orzan et al., 2018).

Adopting efficient and effective packaging increases the environmental, the economic, and the operational performance of firms (Humbert et al., 2009). These outcomes require the implementation of three main practices: reduction at source of product packaging by using fewer resources and creating less waste; reuse or refilling; and recycling (Livingstone et al., 1994). According to the Waste Management Hierarchy (Directive 94/62/EC), the reduction at source

remains the best alternative to manage waste. It is also one of the most important practices of firms' environmental initiatives (Chappin et al., 2009). Dewees and Hare (1998) examined the regulation of packaging waste in Canada and their results indicate that 88% of reduction in waste in the province of Ontario was reached by source reduction. Therefore, this paper focuses on the reduction at source of product packaging. To encourage firms to adopt such practice, governments in developed countries have launched incentive-based environmental policies called market-based instruments (e.g. Sharma, 2000). However, Chappin et al. (2009) were cautious about the idea that these instruments always lead to the desired change in firms' environmental practices. This is attributed to the fact that manufacturers might be strongly inclined to avoid costs deriving from environmental policies. Besides, earlier studies have revealed that manufacturers adjust their environmental practices according to the stringency of waste regulations to which they are subject in their industries (Wang et al., 2018; Zhang et al., 2012). Hence, policymakers have emphasized the urgent need to resolve problems connected with source reduction of packaging waste and to promote it more effectively (Yamaguchi & Takeuchi, 2016), especially because of the perceptional gap that exists between policymakers, implementers, and targets of policies (Harland et al., 2019).

Although several previous studies have examined consumer perceptions, behavior, and willingness-to-pay for products with reduced packaging (e.g., Rokka & Uusitalo, 2008; Çakır & Balagtas, 2014; Yamaguchi & Takeuchi, 2016), few studies have examined manufacturers' behavior with regard to packaging source reduction and specifically the implications of market-based instruments that incite them to adopt such practice (e.g., Fernie & Hart, 2001; Cela & Kaneko, 2013). Some studies have assessed the effectiveness of regulatory measures on packaging waste (e.g, Dewees & Hare, 1998; Røine & Lee, 2006; Rouw & Worrell, 2011; Bartl, 2014; Tencati et al., 2016; Ren et al., 2018) such as environmental taxes (Cela & Kaneko, 2013). However, none of the reviewed studies have empirically examined the moderating role of the stringency of packaging environmental regulation on the relationship between market-based instruments and the reduction at source of packaging. It is therefore of interest to fill this gap in the current literature. This can yield to important implications for designing more consistent and effective instruments and hence decrease the environmental consequences of packaging waste (Tencati et al., 2016).

The purpose of this study is twofold. First, we empirically assess at a panel level the individual effects of market-based instruments on manufacturers' decisions to produce less packaging at the source. Secondly, we investigate the moderating role of the stringency of packaging environmental regulation on the relationship between market-based instruments and the reduction at source of packaging.

The Province of Quebec case offers an opportunity to investigate how manufacturers respond to the green policies regarding packaging waste. We examine the relationships of interest using an unbalanced longitudinal data set for the packaging weights from 848 Quebec firms during the period 2005-2017. We constructed a new measure for the stringency of packaging environmental regulation that is based on the variation of packaging emission intensity across industries and over time. Our findings provide empirical evidence for the perceptional gap between policymakers, implementers and targets of environmental policies. Policymakers should not expect that the latter will always produce the same effect in all industries.

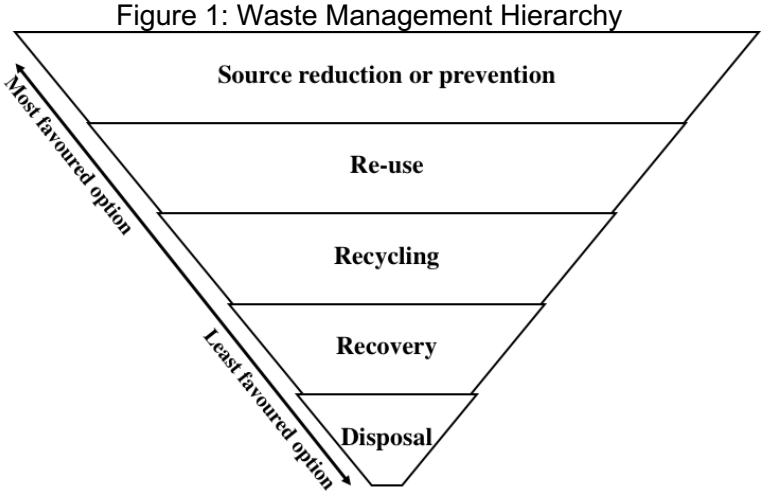
This paper is structured as follows. The theoretical background and research hypotheses are first outlined. Subsequent section describes the methods and research design. After that, the empirical results and discussion are presented.

LITERATURE REVIEW

Reduction at source of product packaging

The reduction at source of product packaging is implemented through two main approaches: package "downsizing" and package "eco-compatibility" (Hazen et al., 2012). According to Vernuccio et al. (2010), "package eco-compatibility" minimizes the environmental impact of packaging materials by reducing their quantity and variety, while ensuring that an optimal combination is used to guarantee the effectiveness of the packaging/product pairs across the supply chain. According to Çakır and Balagtas (2014), "package downsizing" or "package lightweighting" (Tencati et al., 2016) optimizes packaging size, volume, and weight. The importance of these practices stems from the "Waste Management Hierarchy" (Figure 1). The latter specifies the order of priority of each waste management option. The preferred alternative is prevention or source reduction by using less packaging materials in manufacturing and in product designs (Article 1 of the Directive 94/62/EC; Bartl, 2014). These practices are important because they are introduced before materials, packaging, and products become waste (Hazen et al., 2012; Bartl, 2014). They provide more efficient physical distribution of products (i.e. transportation, handling, storage, recovery, reuse or disposal) due to volume/weight efficiency, overpackage elimination, and systematic reduction of upstream waste sources. Given that source reduction of packaging materials is seen as the best solution, a number of scholars have pointed out the need to investigate the external factors that induce firms to adopt such green practice (e.g., Hazen et al., 2012; Bartl, 2014; Çakir & Balagtas, 2014).

Figure 1: Waste Management Hierarchy



Adapted from Article 1 of the Directive 94/62/EC: European Parliament and the Council of the European Union (1994) and Bartl (2014)

Market-Based Instruments (MBIs)

Many governments have launched specific market-based instruments that are based on price and market mechanisms in order to induce firms to adopt sustainable practices (Porter et al., 1995; Sharma, 2000). These instruments take a variety of forms. They can be subsidies, tax abatements, tradable emission permits, pollution fees, or other specific programs designed to encourage green behaviors (Jaffe et al, 2002; Liu et al., 2014). According to Porter et al. (1995), these incentives could be a major driving force that induces firms to undertake environmental practices. They not only enhance environmental performance of firms but also their economic

performance (win-win theory). Williams et al. (2011) show that market-based environmental regulation is an important way to encourage managers in the food processing industry to adopt environmental packaging practices. Similarly, Zhang et al. (2008) and Tsireme et al. (2012) find that incentive-based environmental instruments play a critical role on managers' decisions to adopt green supply chain practices. However, Chappin et al. (2009) were cautious about the idea that these instruments always lead to the desired change in firms' environmental practices. They suggest that these market mechanisms present firms with the following dilemma: either decision makers adopt green practices to avoid potential costs related to environmental sanctions, or they bear these costs when they are lower than those incurred by adopting green practices. This implies that firms might be strongly inclined to avoid costs deriving from environmental policies. In this line of thinking, Harland et al. (2019) highlight the perceptional gap that exists between policymakers, implementers, and targets of policies. The distance between what implementers and targets of policies want and what policymakers think they want, prevents many incentive-based environmental policies from producing the desired effects. Against this background, it is important to ascertain under what conditions the MBIs generate the intended change in firms' behavior regarding producing less packaging at the source.

The institutional theory viewpoint

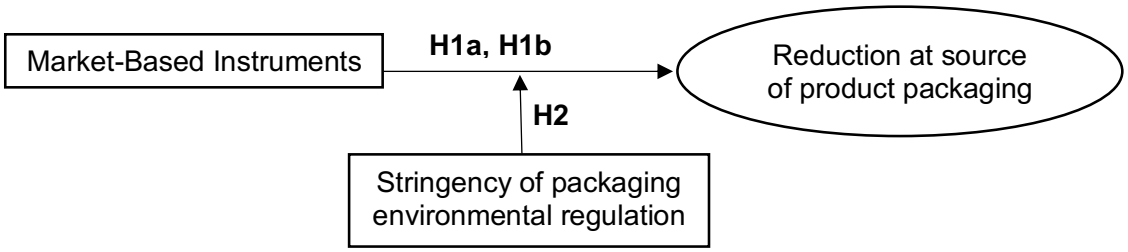
The institutional theory provides relevant lenses to investigate how external institutional forces shape decision-making of firms for sustainability (Dobrzykowski, 2019; Montabon et al., 2016). These forces drive firms to make sustainable strategic choices and to adjust their policies, structures, and processes accordingly (Clemens & Douglas, 2006; Zailani et al., 2012). In the field of sustainability, the importance of the institutional theory in explaining organizational responses to environmental issues has been recognized both conceptually and empirically. Past research has adopted this theory to explain how firms address green issues and, in particular, how they implement green strategies (e.g. Delmas, 2002; Tate et al., 2010; Zailani et al., 2012; Liu et al., 2014). These studies suggest that there are external institutional pressures that play a substantial role in determining the adoption of green initiatives. The influence of institutional drivers is determined by three mechanisms (Zailani et al., 2012: 724), namely : (1) coercive isomorphism resulting from "pressures exerted by other organizations on which the firm depends" ; (2) normative isomorphism related to "expectations in particular organizational contexts about what constitutes appropriate and legitimate behavior"; and (3) mimetic isomorphism that follows "the firm's rational desire to imitate the behavior of other organizations." These drivers affect how organizations make decisions related to environmental issues (Li, 2014). However, firms' strategic choices are heterogeneous depending on how they respond to these institutional pressures and according to their needs and specificities of the industry in which they operate (Seo & Creed, 2002; Saldanha et al., 2015; Dobrzykowski, 2019). Wang et al. (2018) and Tate et al. (2010) pointed out that government incentive-based measures represent a major formal institutional driving force that induces firms to set up proactive environmental strategies. That's why this study focuses mainly on the effect of market-based environmental tools which are important to drive organizational responses to environmental issues (Liu et al., 2014).

HYPOTHESES DEVELOPMENT

Figure 2 presents our conceptual framework. We relied on the institutional theory viewpoint to conceptualize the role of MBIs that act as an institutional antecedent that could potentially impact firms' decisions to implement the reduction at source of packaging. Besides, our conceptual model highlights the moderating role of the stringency of packaging environmental

regulation on the relationship between MBIs and the reduction at source of packaging. The following section clarifies the causality between these variables.

Figure 2: Conceptual framework



Several authors argue that manufacturers have to manage the negative impacts that their product packaging generates at the end of the life cycle (e.g., Fullerton & Wu, 1998; Wiesmeth et al., 2018). Particularly in developed countries, governments tried to remedy this situation by setting up economic incentives to encourage firms to undertake appropriate environmental initiatives. These incentives could be a powerful agent for change of environmental practices of firms (Porter et al., 1995; Sharma, 2000). However, the perceptional gap that exists between policymakers, implementers, and targets of policies could prevent these incentives from producing the desired effects (Harland et al., 2019).

The existing literature shows that Extended Producer Responsibility (EPR) is one of the most important regulatory measure that prolongs firm responsibility for packaging waste across the supply chain. This is based on the "polluter pays" principle which implies that producers contribute financially to the EPR schemes. This instrument presents manufacturers with strong incentives to make optimal packaging choices regarding volume, weight, and materials (Fullerton & Wu, 1998; Bailey, 1999; Dewees & Hare, 1998; Mayers & Butler, 2013). Dewees and Hare (1998) have economically assessed the effectiveness of this measure on packaging waste. They established that 88% of waste reduction in Ontario (Canada) was reached by the reduction at source. Since EPR policy is driven by weight rather than product volume, it induces firms to reduce their packaging weight at the source. These results suggest that firms are inclined to reconsider their packaging design when they bear the financial burden of eliminating its waste (Ferreira et al., 2017). Nevertheless, Roine and Chin-Yu (2006) found a weak causality between Norwegian EPR and technological change and innovation for plastic packaging. The EPR has an indirect and no significant effect on institutional change and innovation, but it has rather a direct effect on downstream operations through increased recycling. Mayers and Butler (2013) and Ferreira et al. (2017) maintained that the costly financial contribution to the EPR schemes forces manufacturers to reconsider their packaging, giving rise to a “voluntary producer responsibility”. This suggests that EPR policy induces manufacturers to optimize their product packaging in order to reduce costs and avoid penalties (Røine and Lee, 2006). Based on the foregoing, the EPR policy implies that manufacturers are forced to internalize costs that were previously externalized to society, but they also have a way to reduce costs and avoid potential environmental sanctions or penalties. According to the institutional theory, the formal pressure imposed on firms externally would give rise to two forms of strategies. Firms could either bear these costs when they are lower than those generated by reducing their product packaging, or they would implement this practice to avoid environmental sanctions. In other words, sustainable packaging becomes more attractive when the total

disposal costs that must be borne by firms are high. Based on the above, we formulate the following hypothesis:

H1a*: The costs of packaging waste that must be borne by manufacturers induce to a higher source reduction of their product packaging.*

Moreover, packaging taxes or fees and subsidies to environmental designs represent strong economic incentives that encourage firms to produce less packaging (Fullerton & Wu, 1998; Dewees & Hare, 1998; Mayers & Butler, 2013). Many scholars have discussed the effect of these price mechanism instruments on firms' decisions to undertake environmental practices. Lin and Ho (2011) carry out an empirical research on the adoption of green practices by Chinese logistics firms and find that they are more encouraged to optimize the ecological quality of their packaging when there are economic incentives. Tsireme et al. (2012) find that incentive-based environmental instruments play a critical role on managers' decisions to adopt green supply chain practices. These authors argue that the high environmental taxes positively affect the decision of managers to introduce green principles in their supply chain. By contrast, Fernie and Hart (2001) were cautious about this approach. They argue that it is difficult to replace short-term benefits with long-term environmental objectives particularly in industries where cost is the main decision driver such as the retail industry. This involves large investment to modify or replace existing machinery which may represent a barrier to implement a new packaging that meets the desired environmental requirements. If the costs borne by firm are high, this will impact both product prices to the consumer and the profit margins of firms. Furthermore, Gray and Guthrie (1990) and Verghese and Lewis (2007) suggest that the uncertainty of commercial benefits and difficulty to predict the financial reward could make manufacturers reluctant to make significant investments to improve the ecological quality of their packaging. Added to this is the uncertainty related to the full costs of packaging in the distribution chain so that the benefits of any change can be accurately evaluated.

A few researchers have demonstrated that decision makers are only willing to adopt environmental packaging practices when the costs are appropriate (e.g. Pullman & Wikoff, 2017; Arnaud, 2017). This stems from the fact that manufacturers are predominately driven by a cost reduction approach (Vandermerwe & Oliff, 1990; Thøgersen, 1999; Rokka & Uusitalo, 2008). Kassaye (2001) points out that cutting waste and garbage removal costs is one of the main reasons why large, medium, and small companies engage in green packaging practices. Verghese and Lewis (2007) argue that sustainable packaging innovation is only adopted by firms when it delivers economic benefits such as increased efficiency and cost reductions. Orzan et al. (2018) find that protecting the environment by implementing ecological packaging practices is seen as a cost-reducing and a profit-maximizing approach by decision makers particularly when they expect significant economic benefits. Based on the above, it is plausible to assume the following hypothesis:

H1b. *Economic incentives based on price mechanisms are positively associated with packaging reduction at source. Manufacturers are willing to produce less packaging when there are short-term economic benefits higher than the implementing costs required.*

Moreover, the existing literature reveals that stricter environmental regulation is more likely to have a significant effect on corporate green packaging strategies. Firms mainly adjust their packaging buying behavior to comply with laws and regulations (Gray and Guthrie, 1990; Vandermerwe & Oliff, 1990; Livingstone & Sparks, 1994; Røine & Lee, 2006). Stringent packaging waste regulations have made packaging operations and planning a critical green logistics issue. These regulations also influence product innovation processes of manufacturers

(Dewees & Hare, 1998; Ciliberti *et al.*, 2008). In line with this, Wang et al. (2018) carry out an empirical research on corporate environmental actions in China and find that companies who operate in the more stringently regulated industries are the ones that are more engaged in environmental management practices. This fact is particularly noticed among firms who operate in heavy-pollution industries. Consequently, they are obliged to comply with stricter regulations to avoid potential penalties and legal liabilities for noncompliance (Zhu & Sarkis, 2014). Conversely, companies who operate in environmentally friendly industries are not subject to stringent regulations. This implies that companies adjust their buying behavior mainly to comply with the level of regulatory stringency they face. Wang et al. (2018) and Brunel and Levinson (2016) capture regulatory stringency influences by using sectorial gas emission. They find that stricter sectorial regulation reduces emissions faster than weaker sectorial regulation. In the same line of thinking, Zhang et al. (2012) explain that local governments generally allocate more resources and implement more specific measures in environmentally disruptive industries in order to encourage targeted firms to engage in more environmental practices. Therefore, we argue that the relationship between MBIs and reduction at source of packaging is likely to be moderated by the stringency of packaging environmental regulation. Based on the foregoing, the following hypothesis is formulated:

H2. *The relationship between MBIs and the reduction at source of packaging is intensified by the stringency of packaging environmental regulation.*

METHODS

Choice of the study area

The province of Quebec is an interesting study area because of the emphasis it places on the transition to a green and responsible economy. The Quebec government has introduced economic incentives to discourage environmentally harmful activities and to stimulate environmental innovation. These economic incentives mainly involve cost internalization based on the ‘user-payer’ or the ‘polluter-payer’ principle and include fiscal measures such as taxation, credits, tax exemptions, and tax refunds (Sustainable Development Strategy, 2015–2020: 22).

Data and sample characteristics

To test our hypotheses, we used an unbalanced longitudinal data set for the plastic packaging weights produced by firms operating in six important Quebec industries (Table 1). Data were obtained from Éco Entreprises Québec (ÉEQ), a private and non-profit organization that finances municipal curbside recycling services in Quebec. Firms who placed packaging on the Quebec market have the obligation to report the total quantity of each packaging material. They accordingly pay contributions to the compensation program of ÉEQ. These funds are used to finance the selective collection of packaging material.

The period chosen is 2005-2017 which was determined by the availability of complete annual ÉEQ data. The year 2005 marks the creation and the accreditation of ÉEQ. Our initial sample contained 923 firms, of which 75 were excluded from the data set due to outliers. The final sample retained includes 848 firms. In addition to this data, we used other longitudinal data from the Quebec Institute of Statistics, Statistics Canada, the Official Gazette of Quebec, Recyc-Québec, and the Ministry of the Environment and the Fight against Climate Change in Quebec. Table 2 presents the data generated from each of these sources.

Table 1: Sample characteristics	
Industries	**Number of firms**
Food processing	445
Furniture and related products	119
Retail industry	102
Electronics	83
Chemical and pharmaceutical	83
Beverage and tobacco	16
Total	848

Unit of analysis

Given the strict measures for confidentiality of ÉEQ data, the names of firms in the data set have been hidden by ÉEQ and were replaced with reference numbers instead. Considering that no firm-level information was available, the unit of analysis for this research is the firm embedded in its industry. We used industry-level proxies for MBIs. Measuring these variables at a more aggregate level enables linkages among firms within the same industry to be considered. For the stringency of packaging environmental regulation variable, we constructed a new measure to characterize the level of packaging regulatory strictness faced by firms in each industry.

Description of variables and measures

The selection of variables and their measures were based on our literature review. Moreover, the collaboration of ÉEQ specialists helped ensure that variable choices and database manipulation are as appropriate as possible, which enhances the internal validity of the study. Table 2 summarizes all the variables included in the model, their measures, and the data sources.

Dependent variables. The reduction at source of packaging is the variable of interest in this study. The most important aspect of this reduction is the optimization of packaging weight. Hence, we used the weights of two categories of packaging as a proxy: the high-density polyethylene (HDPE) and the low-density polyethylene (LDPE) plastic packaging. If the weights of these packaging decreases, there will be a reduction at source. The data was obtained from ÉEQ and represents the annual packaging weights of products from 848 firms. This variable was transformed into natural logarithms in case a non-linear relationship exists between the dependent and the independent variables. The weight of HDPE and LDPE plastic packaging represents a relevant indicator for the reduction at source of packaging for many reasons. First, they are single-use packaging, which has a negative impact on the environment and represents a major issue for Quebec especially due to the closure of the Asian market. Second, Recyc-Québec (2018: 3) states that the market for disposing of this packaging is “fragile and depends on the Asian market.” It is actually difficult for sorting centers to find local markets for used packaging because they are concentrated almost exclusively in Asia. Third, the Quebec Industrial Research Center (QIRC) estimated that the cost of plastic bag impacts throughout Quebec was $4.1 million in 2010. Consequently, the government through its regulatory agencies developed some measures to encourage firms to optimize their product plastic packaging (Recyc-Québec, 2018).

Independent variables. External institutional drivers are defined as explanatory variables. Considering the characteristics of the MBIs and the availability of provincial data, three

indicators were retained to measure the effect of MBIs. First, the tax imposed on plastic packaging, which represents the contribution of firms to the compensation scheme. Second, the share of recycling costs borne by Quebec firms. Third, the sales price index of plastic packaging material on the market.

Moderator variable. The stringency of packaging environmental regulation is our moderator variable. Brunel and Levinson (2016) highlight several challenges surrounding the measurement of environmental regulatory stringency such as the heterogeneous nature of environmental issues, the various instruments that address them, and data availability. Hence, to operationalize the level of environmental regulatory stringency faced by firms in each industry, we constructed a new measure that could be used to characterize packaging regulatory strictness across industries and over time. We followed the recommendations of Brunel and Levinson (2016) to construct an appropriate measure.

Our packaging regulatory strictness measure is based on the variation of packaging emission intensity across industries and over time. The equation (1) represents our emissions-based measure of stringency. We used the total annual weight of all packaging categories (i.e. plastic, glass, metal, paper, aluminum, steel packaging) placed on the market by firm *i* from 2005 to 2017. We divided this total annual packaging weight (E_{it}) by the annual GPD of the industry where the firm *i* operates over the 2005-2017 period (GDP_{it}).

$$\text{Index of packaging regulatory strictness} = \frac{\text{Eit}}{\text{GDPit}} \quad (1)$$

The idea behind using packaging emission intensity as a measure of regulatory stringency originates with the approach outlined in Brunel and Levinson (2016) suggesting that most regulated firms will emit less packaging on the market. Where we see higher packaging emission intensity, we would conclude that the cost of polluting is lower because regulations are less stringent in that industry. Where we see lower packaging emission intensity, regulations must be more stringent. Stricter environmental regulation reduces emissions faster than weaker sectorial regulation.

Control variable. We controlled for industry size and competitiveness using the log-transformed GDP at basic prices of each industry. This indicator has been widely used in prior literature as a reliable proxy for this variable.

Table 2: Summary of variables, measures and data sources

VARIABLES	MEASURES	UNITS	DATA SOURCES
Reduction at source of packaging (RSP)	Weight of plastic packaging put on the market	Ton	Éco Entreprises Québec
Market-based instruments (MBI)	Taxation imposed on plastic packaging (TAX)	$/Metric ton	Official Gazette of Quebec
	Share of recycling costs borne by Quebec firms (RECYCOST)	%	Éco Entreprises Québec
	Plastic price index (PRICE)	$/Metric ton	Recyc-Québec
Stringency of packaging environmental	Index of packaging regulatory strictness calculated using the equation (1).	%	Éco Entreprises Québec; Quebec Institute of Statistics

regulation (SPER)			
Industry size and competitiveness (GDP)	Gross Domestic Product (GDP) at basic prices by industry	K $	Quebec Institute of Statistics

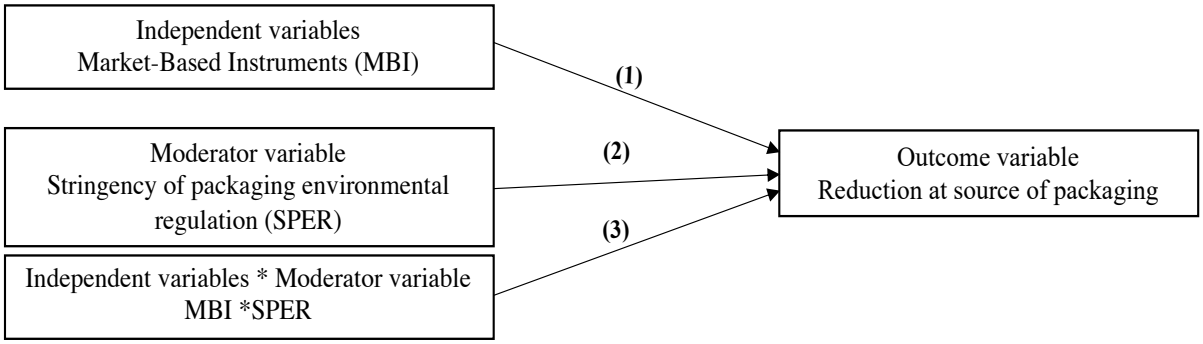
Empirical model for the econometric analysis

Figure 3 presents the empirical model. The first path represents the impact of MBI on RSP, the second path denotes the impact of the moderator (SPER), and third path is the interaction term between MBI and SPER, which modifies (strengthens or weakens) the relationship between MBI and RSP. The empirical model identifying the relationships between reduction of plastic packaging by firms in industry i at time t, the independent, the moderator and the control variables is specified as follows.

$$RSP_{it} = \alpha_0 + \alpha_1 RECYCOST_t + \alpha_2 TAX_t + \alpha_3 PRICE_t + \alpha_4 SPER_{it} + \alpha_5 TAX_t{*}SPER_{it} + \alpha_6 RECYCOST_t{*} SPER_{it} + \alpha_7 Price_t{*} SPER_{it} + GDP + c_i + \eta_t + \varepsilon_{it}$$

Where α_0 is the constant, ε is the error term, c_i represents time invariant industry specific effect and η_t represents industry invariant time specific effect.

Figure 3: Empirical model



Model selection

We analyzed our panel data using Stata 14.1. Longitudinal regression models were employed to estimate relationships among variables longitudinally. The first concern in a panel data analysis is to choose a valid model. The ordinary least squares (OLS) method might be biased and inconsistent because of the unobserved firm-specific and time-specific heterogeneities (Wooldridge, 2002). To choose the most appropriate model that yields consistent and efficient estimates, we estimated both fixed and random effects models. Then, we conducted the Hausman specification test to decide whether to use the fixed or the random effects model. The Hausman's test yields a p-value to be greater than 0.05, suggesting that the random effects model represents a valid model, which means that the random effects estimates are more consistent and efficient than fixed effects estimates (Hausman, 1978). Hence, to examine the individual effect of MBI and the moderating role of SPER on the MBI–RSP relationship, we used random effect model for the full sample and then for each group of industry. We added the option 'robust' to control for heteroskedasticity (Wooldridge, 2002).

RESULTS

Random effects estimation for the full sample

The results for the random effects estimation for the full sample are shown in Table 3. Model (1) contains the linear and squared terms of the market-based variables. Models 2 and 3 contain separate regressions involving the interactions of regulatory stringency and the market-based variables, one of which includes the effect of the control variable. Year effects are included in all models.

Model (1) shows interesting results related to the relationship between the economic incentives and the reduction in the weight of packaging. The coefficients for the linear and the squared terms of taxation and material price variables are statistically significant with a positive and a negative sign, respectively. This suggests an inverted U-shaped effect of material price and taxation on packaging weight, which means that material price and taxation at first increases packaging weight, then decreases it after a certain level of cost/taxes is attained. Using the coefficient values obtained for material price (0.330) and square of material price (-0.039), the turning point of the curvilinear relationship is situated at a material price optimal value of 68$ per ton (i.e., exponential (0.330) divided by 2*(-0.039)). Using the coefficient values obtained for taxation (9.951) and square of taxation (-0.826), the turning point of the curvilinear relationship is situated at a taxation optimal value of 413$ per ton (i.e., exponential (9.951) divided by 2*(-0.826)). These findings demonstrate a non-linear relationship between the economic incentives and the reduction in the weight of packaging.

Otherwise, it is not surprising that the GDP at basic prices is found to have a significant positive impact. It seems intuitive that a raise in the economic growth increases the amount of packaging placed on the market. To better understand the results obtained of the moderating effect of the stringency of packaging regulation, we estimated the model involving interactions between regulatory stringency and the market-based variables for each industry. The estimation results for each industry are presented in Table 4.

Random effects estimation by industry

Hypothesis 1a was supported in the food processing, the electronics, and the furniture industries, as shown by the consistent results of the recycling costs variable, which has a significant and negatively signed coefficient. This suggests that the increase in recycling costs of packaging waste borne by manufacturers encourages them to reduce the weight of packaging they place on the market. These results are consistent with the “Extended Producer Responsibility” which states that producers are only willing to deliver optimal choices of packaging when they are responsible for the costs of packaging waste at the post-consumption stage. These results are also consistent with the Directive 94/62/EC on packaging and packaging waste, which has major implications for packaging decisions of companies. This instrument incentivizes firms to produce packaging with optimal weights in order to decrease the disposal costs that they bear at the end of the product life cycle. However, Hypothesis 1a was not supported in the chemical and pharmaceutical, and the retail industries, as shown by the consistent results of the recycling costs variable. This instrument does not seem to provide enough incentive to reduce packaging weight for firms operating in these industries.

Hypothesis 1b was supported in the food processing, the electronics, and the beverage and tobacco industries, as shown by the consistent results of taxation variable, which has a significant and positively signed coefficient. However, Hypothesis 1b was not supported in the

chemical and pharmaceutical, and the retail industries, as shown by the consistent results of the recycling costs variable. The coefficients of the material price are statistically significant with a negative sign in the retail industry.

Regarding the stringency of packaging regulation, it is found to have a significant negative impact in the electronics and in the chemical and pharmaceutical industries, which indicates that an increased level of the stringency of packaging regulation in these industries would induce manufacturers to produce less packaging. The consistent results of the moderating effect of the stringency of packaging regulation showed support for hypothesis 2, but with dissimilar effects depending on the industries.

First, the estimated coefficient for the interaction between the stringency of packaging regulation and the taxation is negative and significant in the electronics industry, suggesting a steepening of the relationship between those two variables. These findings demonstrate that the increased of packaging regulatory pressure strengthens the effect of taxation on the reduction of packaging weight at the source. Otherwise, the coefficient of the interaction between the stringency of packaging regulation and taxation is positive and significant in the chemical and pharmaceutical industry. This means that an increase in the stringency of packaging regulation weakens the effect of taxation on packaging weight in this industry.

Second, the interaction coefficients between the stringency of packaging regulation and material price are positive and significant in the food processing, the chemical and pharmaceutical, the retail, and the beverage and tobacco industries. These results suggest that an increase in the packaging regulation stringency weakens the effect of material price on the packaging weight in these industries. Whereas the interaction coefficients between those two variables are negative and significant in the electronics and the furniture industries, which suggests that an increase in the regulation stringency strengthens the effect of material price on reduction at source of packaging in these industries.

Third, the interaction coefficients between the stringency of packaging regulation and recycling cost are negative and significant in the chemical and pharmaceutical and the beverage and tobacco industries. This suggests that an increase in the regulation stringency strengthens the effect of recycling cost on reduction at source of packaging in these two industries. The interaction coefficient between those two variables is positive and significant in the electronics industry, which means that an increase in the stringency of packaging regulation weakens the effect of material price on the packaging weight in this industry.

Table 3: Random effects estimation for the full sample

Variables	**Model (1)**	**Model (2)**	**Model (3)**
LOG_TAX	9.951***	0.064	0.028
	(1.521)	(0.092)	(0.098)
RECYCOST	-1.657	-0.621***	-0.585***
	(1.460)	(0.130)	(0.139)
LOG_PRICE	0.330**	-0.029	-0.040
	(0.153)	(0.023)	(0.025)
SPER	7.967***	-27.392	-28.167

	(1.201)	(21.860)	(22.251)
TAX*SPER		4.886	5.011
		(4.492)	(4.566)
PRICE*SPER		1.800**	1.847**
		(0.903)	(0.920)
RECYCOST*SPER		2.432	2.297
		(7.120)	(7.219)
TAX_SQUARE	-0.826***		
	(0.133)		
PRICE_SQUARE	-0.039*		
	(0.023)		
RECYCOST_SQUARE	0.145		
	(0.969)		
LOG_GDP	0.247***		0.298***
	(0.050)		(0.048)
Constant	-31.826***	1.340***	-2.310**
	(3.933)	(0.441)	(0.944)
Observations	4,826	4,826	4,826
R^2 within	0.1198	0.1356	0.1293
R^2 between	0.8072	0.8270	0.3791
R^2 overall	0.1209	0.1360	0.1298
Year effect	Yes	Yes	Yes
Wald chi^2	366.72	569.35	413.29
	0.000	0.000	0.000
Number of Year	12	12	12

Table 4: Random effects estimation by industry

Industry groups	**Group (1)** Food	**Group (2)** Electronics	**Group (3)** Chemical &	**Group (4)**	**Group (5)** Furniture	**Group (6)** Beverage &
Variables	processing		pharmaceutical	Retail		tobacco
LOG_TAX	0.296**	2.106***	-0.571*	-1.786***	0.194	0.923*
	(0.125)	(0.735)	(0.305)	(0.392)	(0.373)	(0.502)
RECYCOST	-0.873***	-4.072***	1.189***	1.897***	-1.148*	-1.362
	(0.182)	(1.450)	(0.294)	(0.624)	(0.655)	(0.926)
LOG_PRICE	0.034	0.177	-0.058	-0.375***	0.040	-0.130
	(0.029)	(0.113)	(0.046)	(0.099)	(0.065)	(0.084)
SPER	-0.203	-2.014*	-0.794***	-1.667	2.195	-0.010
	(0.671)	(0.180)	(0.548)	(0.905)	(0.788)	(0.509)
TAX*SPER	2.936	-3.859*	3.757***	2.482	-3.521	7.667

	(0.506)	(1.673)	(0.123)	(1.357)	(0.625)	(0.365)
PRICE*SPER	1.249*	-2.821*	7.316***	8.850*	-9.897*	4.783***
	(0.700)	(22.571)	(1.075)	(5.048)	(5.520)	(20.407)
RECYCOST*SPER	3.386	4.011*	-8.813***	6.176	93.908	-1.807***
	(5.699)	(329.783)	(10.277)	(67.190)	(77.807)	(118.132)
Constant	0.421	-8.985***	3.074**	11.064***	-0.191	-4.591**
	(0.616)	(3.061)	(1.419)	(1.529)	(1.602)	(2.283)
Observations	2,435	471	484	497	868	71
R^2 within	0.1469	0.2454	0.2839	0.2772	0.2779	0.6125
R^2 between	0.1316	0.6003	0.8080	0.9292	0.6877	0.9464
R^2 overall	0.1469	0.2507	0.2848	0.3038	0.2801	0.6293
Year effect	Yes	Yes	Yes	Yes	Yes	Yes
Wald chi^2	343.28	230.16	572.24	613.19	738.39	117.42
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Year	12	12	12	12	12	12

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

DISCUSSION AND CONCLUSIONS

Based on the institutional theory viewpoint, this paper examines at a panel level the individual effect of MBIs on manufacturers decisions that would lead to producing less packaging at the source. We also investigate the moderating effect of the stringency of packaging environmental regulation on the relationship between MBIs and the reduction at source of packaging. We first estimate a random effects model for the full sample and then by industry.

Theoretical contributions

Our findings demonstrate a non-linear relationship between the economic incentives and the reduction in the weight of packaging. There is an inverted U-shaped relationship between taxation and packaging weight, and between material price and packaging weight. These results contradict those of previous studies who claim a linear relationship between environmental instruments and green practices (e.g. Tsireme et al., 2012). Since decision makers focus predominately on cost reduction and profit maximization, they are only willing to reduce their packaging at the source if the economic benefits are significant. The estimated turning points of the curvilinear relationship suggest optimal values for taxation and plastic market price. Based on these findings, we can argue that the motivation behind producing less packaging at the source might not be a legitimate practice that is based only on ethical sensitivity, but this choice is rather promoted through pragmatic values (i.e., economic incentives). The latter seems to be a particularly strong institutional driver that shapes organizational decision-making regarding this green practice. Being predominately driven by a cost-reduction approach, firms tend to be committed to reduce their packaging waste (e.g., Bone & Corey, 2000; Pullman & Wikoff, 2017).

Moreover, our estimated results show that the effect of different MBIs are dissimilar depending on the industries. We found that industries are not homogenous in their product packaging

practice despite experiencing the same incentive-based instruments. Since firms are embedded in different institutional environments, they respond heterogeneously to the similar institutional pressures they undergo. This implies that applying the same instruments in all industries might be an inadequate approach to achieve institutional expectations regarding environmental issues. Therefore, we argue that the MBIs should rather be personalized by regulatory institutions according to the peculiarities of the institutional field in which the firm operates. This will enhance the environmental responsiveness of firms and increase the chances of achieving the expected change in firms' environmental behaviors in terms of packaging green practices.

On the one hand, our findings show that the costs of packaging waste that must be borne by manufacturers are associated with higher source reduction of product plastic packaging in the food processing, the electronics, and the furniture industries. This implies that higher disposal cost increases the benefit of reducing waste and therefore increases the attractiveness of the reduction at source of packaging in these industries. This is consistent with the EPR policy which suggests that producers are more inclined to deliver optimal product packaging when they are responsible for packaging waste costs at the post-consumption stage. This instrument incentivizes organizations to produce packaging with optimal weights to benefit from cost-saving opportunities at the end of the product life cycle. However, these findings have not been supported in the chemical and pharmaceutical, and the retail industries. The costs of packaging waste that must be borne by manufacturers do not seem to provide enough incentive for firms operating in these two industries to produce less packaging at the source. On the other hand, the results show that higher taxation is associated with the production of less packaging at the source in the food processing, the electronics, and the beverage and tobacco industries but not in the chemical and pharmaceutical, and the retail industries. The plastic price variable has failed to display any significative influence on the production of less packaging at the source except in the retail industry.

These results confirm that there is a heterogeneous environmental responsiveness of industries to seemingly similar incentive-based instruments. These results might be attributed to distinct characteristics of the ecological attributes of the plastic packaging that is used by firms operating in these industry groups. Reliance on market mechanisms also lead firms to conflicts between economic and ecological imperatives (Bailey, 1999). The institutional theory attributes these heterogeneous responses to similar institutional pressures to the differences in composition of the institutional field in which organizations are embedded. Our results suggest that organizations tend to produce less packaging at the source based on the logic of cost reduction, but they respond heterogeneously to the similar institutional pressures they receive depending on their institutional field. These findings contradict those of previous studies (e.g. Li, 2014) that found that coercive regulatory pressures always lead to a significant positive impact on firm's environmental behaviors.

Furthermore, our research contributes to the literature analyzing the moderating role of the stringency of environmental regulation. First, we contributed to the measurement of this variable by developing a new measure that is based on the variation of packaging emission intensity across industries and over time. Considering the challenges surrounding the measurement of the stringency of environmental regulation (Brunel & Levinson, 2016), our measure is specifically adapted to address the packaging environmental issue. Second, our findings suggest that an increased level of the stringency of packaging regulation in the electronics and in the chemical and pharmaceutical industries induce manufacturers to produce less packaging. Intuitively, as an industry becomes more polluting, the stringency of regulation rises, thus inducing firms to engage more in environmental practices. These findings confirm those reported in previous studies suggesting that firms operating in heavy-pollution industries are

more subject to stringent regulations. Hence, they tend to be more proactive to show their compliance with regulations (Wang et al., 2018; Brunel & Levinson, 2016; Zhu & Sarkis, 2004; Zhang et al., 2012). Third, the relationship between MBIs and the reduction at source of packaging is moderated by the stringency of packaging regulation in different ways (strengthening or weakening effect) depending on the industries. Let's remember our main results suggesting that the stringency of packaging regulation strengthens the effect of taxation on the reduction at source of packaging in the electronics industry, whereas it is weakening this effect in the chemical and pharmaceutical industry. Besides, the stringency of packaging regulation weakens the effect of plastic price on the reduction at source of packaging in the food processing, the chemical and pharmaceutical, the retail, and the beverage and tobacco industries, whereas it is strengthening this effect in the electronics and the furniture industries.

Finally, the stringency of packaging regulation strengthens the effect of recycling cost on the reduction at source of packaging in the chemical and pharmaceutical and the beverage and tobacco industries, whereas it is weakening this effect in the electronics industry. These findings support the argument that the institutional environment significantly influences the implementation of green practices. As suggested by Thornton and Ocasio (2008), there will always be trade-offs to be made as organizations face several and sometimes conflicting institutional logic within the same institutional field. Our research provides strong evidence that organizations adjust their response to the incentive-based instruments according to the level of regulatory stringency they face in their institutional field. They act in accordance with the standards and expectations of the institutional context where they are embedded. Our findings provide empirical evidence on the perceptional gap that exists between policymakers, implementers and targets of policies (Harland et al., 2019).

Policymakers implications

The practical implications of this study are relevant to policymakers in regulatory institutions. Since the competition on green products is escalating, it is of utmost importance to understand the antecedents of implementing the reduction at source of packaging which is the most preferred alternative in the waste management hierarchy. Policymakers should recognize source reduction as an important means of reducing waste. Without effective economic incentives, few manufacturers are likely to deliver optimal packaging weight. Our results are helpful for policymakers in regulatory institutions to evaluate the implementation effect of MBIs on this green practice. We demonstrate the existence of an inverted U-shaped relationship between taxation and the reduction at source of packaging and between plastic price on the market and the reduction at source of packaging. We also determine the turning points for these inverted U-shaped relationships. These results would send the right signals to policymakers to formulate better targeted instruments that encourage firms to reduce their product packaging at the source. Furthermore, our findings provide new empirical evidence for the perceptional and behavioral gap between policymakers, implementers, and targets of environmental instruments. This gap can prevent many incentive-based environmental instruments from producing the desired effects. Policymakers should not expect these instruments will always produce the same effect in all industries. When implementing these instruments, policymakers should consider the level of packaging regulatory stringency that businesses face in each industry in order to enhance their environmental responsiveness. This paper provides an important basis to strengthen the effectiveness of such instruments.

Limitations and implications for future research

This research has three main limitations. First, this study shows that the effect of MBIs and packaging regulatory strictness varies across industries. Future research might collect qualitative and quantitative data from each of these industries to provide more in-depth understanding for these variations. Second, our study examines the effect of MBIs on the reduction at source of packaging. It would be useful, both theoretically and empirically, to examine the effects of this practice on firm performance. Future research could shed light on its effects on the environmental and operational performance by using intra-firm indicators. This could give more insights about the performance outcomes and encourage firms to adopt such green practice. Third, firms operating in developed countries face different incentive-based environmental instruments. The idiosyncratic environmental regulations of the developed countries limit the generalizability of our findings to the less developed countries. Future research might collect data from different countries to explore the effects of different incentive-based instruments on reduction of packaging at the source. This could provide more in-depth understanding of antecedents of this practice in different contexts.

REFERENCES

Accorsi, R., Cascini, A., Cholette, S., Manzini, R., & Mora, C. (2014). Economic and environmental assessment of reusable plastic containers: A food catering supply chain case study. International Journal of Production Economics, 152, 88-101.

Arnaud, B. (2017). Extended Producer Responsibility and Green Marketing: An Application to Packaging. Environmental and Resource Economics, 67(2), 285-296.

Anonymous (1994). European Parliament and Council Directive 94/62/EC of 20 December 1994 on packaging and packaging waste. Brussels, Belgium: European Parliament and the Council of the European Union.

Bailey, I. (1999). Principles, policies and practice: evaluating the environmental sustainability of Britain's packaging regulations. *Sustainable Development,* vol. 8, no. 1, pp. 51-64.

Bartl, A. (2014). Moving from recycling to waste prevention: a review of barriers and enables. Waste Management & Research, 32(9), 3-18.

Bone, P.F. and Corey, R.J. (2000). Packaging Ethics: Perceptual Differences among Packaging Professionals, Brand Managers and Ethically-Interested Consumers. *Journal of Business Ethics,* vol. 24, no. 3, pp. 199-213.

Brunel, C. & Levinson, A. (2016). Measuring the Stringency of Environmental Regulations. Review of Environmental Economics and Policy, 10(1), 47-67.

Çakır, M. & Balagtas, J.V. (2014). Consumer Response to Package Downsizing: Evidence from the Chicago Ice Cream Market. Journal of Retailing, 90(1), 1-12.

Cela, E., Kaneko, S., 2013. Understanding the implications of environmental taxes: the case of the Danish weight-based packaging product charge. Environmental Policy Governance, 23 (4), 274–282.

Chappin, M.M.H., Vermeulen, W.J.V., Meeus, M.T.H. & Hekkert, M.P. (2009). Enhancing our understanding of the role of environmental policy in environmental innovation: adoption

explained by the accumulation of policy instruments and agent-based factors. Environmental Science and Policy, 12(7), 934-947.

Clemens, B. & Douglas, T.J. (2006). Does coercion drive firms to adopt 'voluntary' green initiatives? Relationships among coercion, superior firm resources, and voluntary green initiatives. Journal of Business Research, 59(4), 483-491.

Clemens, B. (2006). Economic incentives and small firms: Does it pay to be green? Journal of Business Research, 59(4), 492-500.

Delmas, M.A. (2002). The Diffusion of Environmental Management Standards in Europe and in the United States: An Institutional Perspective. Policy Sciences, 35(1), 91-119.

Dewees, D.N. & Hare, M.J. (1998), "Economic Analysis of Packaging Waste Reduction", *Canadian Public Policy / Analyse de Politiques,* vol. 24, no. 4, pp. 453-470.

Dobrzykowski, D. (2019). Understanding the downstream healthcare supply chain: Unpacking regulatory and industry characteristics. Journal of Supply Chain Management, 55(2), 26–46.

Fernie, J. & Hart, C. (2001), "UK packaging waste legislation: Implications for food retailers", *British Food Journal,* vol. 103, no. 3, pp. 187-197.

Ferreira, S., Cabral, M., da Cruz, N.F., Simões, P., Marques, R.C. & Marques, R.C. (2017), "The costs and benefits of packaging waste management systems in Europe: the perspective of local authorities", Journal of Environmental Planning and Management*,* vol. 60, no. 5, pp. 773-791.

Fullerton, D. & Wu, W. (1998). Policies for Green Design. Journal of Environmental Economics and Management, 36(2), 131-148.

Gray, V. & Guthrie, J. (1990). Ethical Issues of Environmentally Friendly Packaging. International Journal of Physical Distribution & Logistics Management, 20(8), 31-36.

Gouvernement du Québec (2018). Institut de la statistique du Québec et Ministère de l'Environnement et de la Lutte contre les Changements Climatiques. Recueil des indicateurs de suivi de la Stratégie gouvernementale de développement durable 2015-2020. Bibliothèque et Archives nationales du Québec, 4e trimestre 2018, ISSN 2291-9562 (PDF)

Guba, E. G., & Lincoln, Y. S. (1994). « Competing paradigms in qualitative research ». In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 105-117). Thousand Oaks, CA: Sage.

Harland, C., Telgen, J., Callender, G., Grimm, R., & Patrucco, A. (2019). Implementing government policy through public procurement. Journal of Supply Chain Management, 55(2), 6-25.

Hausman, J.A. (1978). Specification Tests in Econometrics, Econometrica, 46(6), 1251-1271.

Hazen, B.T., Hall, D.J. & Hanna, J.B. (2012). Reverse logistics disposition decision-making. International Journal of Physical Distribution & Logistics Management, 42(3), 244-274.

Jaffe, A.B., Newell, R.G. and Stavins, R.N. (2002). Environmental Policy and Technological Change. Environmental and Resource Economics, 22(1), 41-70.

Kassaye, W. W. (2001). Green dilemma. Marketing Intelligence & Planning, 19(6/7), 444-455.

Li, Y.N. (2014). Environmental innovation practices and performance: Moderating effect of resource commitment. Journal of Cleaner Production, 66, 450–458.

Lin, C. & Ho, Y. (2011). Determinants of Green Practice Adoption for Logistics Companies in China. Journal of Business Ethics, 98(1), 67-83.

Liu, X., Yamamoto, R. & Suk, S. (2014). A survey of company's awareness and approval of market-based instruments for energy saving in Japan. Journal of Cleaner Production, 78, 35-47.

Livingstone, S. & Sparks, L. (1994). The new German packaging laws: Effects on firms exporting to Germany. International Journal of Physical Distribution & Logistics Management, 24(7), 15-25.

Mayers, K. & Butler, S. (2013), "Producer Responsibility Organizations Development and Operations", *Journal of Industrial Ecology,* vol. 17, no. 2, pp. 277-289.

Molina-Besch, K., Wikström, F. & Williams, H. (2018). The environmental impact of packaging in food supply chains—does life cycle assessment of food provide the full picture? International Journal of Life Cycle Assessment, 24(1), 37-50.

Montabon, F., Pagell, M. & Wu, Z. (2016). Making Sustainability Sustainable. Journal of Supply Chain Management, 52(2), 11-27.

Orzan, G., Cruceru, A., Bălăceanu, C. & Chivu, R. (2018). Consumers' Behavior Concerning Sustainable Packaging: An Exploratory Study on Romanian Consumers. Sustainability, 10(6), 1787-1791.

Pålsson, H. & D. Hellström (2016). Packaging logistics in supply chain practice - current state, trade-offs and improvement potential. International Journal of Logistics, 19(5), 351-368.

Prendergast, G.P. (1995). The EC directive on packaging and packaging waste: Current status and logistical implications. Logistics Information Management, 8(3), 10-18.

Prendergast, G. & L. Pitt (1996). Packaging, marketing, logistics and the environment: are there trade-offs? International Journal of Physical Distribution & Logistics Management, 26(6), 60-72.

Porter, M.E. & Claas van der Linde (1995). Toward a New Conception of the Environment-Competitiveness Relationship, The Journal of Economic Perspectives, 9(4), 97-118.

Prakash, A. & Kollman, K. (2004). Policy modes, firms and the natural environment. Business Strategy and the Environment, 13(2), 107-128.

Pullman, M. & R. Wikoff (2017). Institutional sustainable purchasing priorities. International Journal of Operations & Production Management, 37(2), 162-181.

Recyc-Québec (2018). Collecte sélective - Sacs et pellicules de plastique. Fiche informative (Septembre 2018). https://www.recyc-quebec.gouv.qc.ca/sites/default/files/documents/fiche-sacs-pellicules.pdf

Ren, S., Li, X., Li, D., Yuan, B. & Chen, X. (2018). The effects of three types of environmental regulation on eco-efficiency: A cross-region analysis in China. Journal of Cleaner Production, 173, 245-255.

Røine, K. & Lee, C. (2006), "With a Little Help from EPR? Technological Change and Innovation in the Norwegian Plastic Packaging and Electronics Sectors", Journal of Industrial Ecology, 10 (1/2), 217-237.

Rokka, J. & Uusitalo, L. (2008). Preference for green packaging in consumer product choices - Do consumers care? International Journal of Consumer Studies, 32(5), 516-525.

Rouw, M., Worrell, E., 2011. Evaluating the impacts of packaging policy in The Netherlands. Resources Conservation and Recycling 55(4), 483-492

Rundh, B. (2009). Packaging design: creating competitive advantage with product packaging. British Food Journal, 111(9), 988-1002.

Saldanha, J. P., Mello, J. E., Knemeyer, A. M., & Vijayaraghavan, T. A. S. (2015). Implementing supply chain technologies in emerging markets: An institutional theory perspective. Journal of Supply Chain Management, 51(1), 5–26.

Seo, M. & Creed, W. (2002). Institutional contradictions, praxis, and institutional change: A dialectical perspective. Academy of Management Review, 27(2), 222-247.

Sharma, S. (2000). Managerial Interpretations and Organizational Context as Predictors of Corporate Choice of Environmental Strategy. The Academy of Management Journal, 43(4), 681-697.

Tate, W. L., Ellram, L. M., & Kirchoff, J. F. (2010). Corporate Social Responsibility reports: a thematic analysis related to Supply Chain Management. Journal of Supply Chain Management, 46(1), 19-44.

Tencati, A., Pogutz, S., Moda, B., Brambilla, M. & Cacia, C. (2016). Prevention policies addressing packaging and packaging waste: Some emerging trends. Waste Management, 56, 35-45.

Thφgersen, J. (1999). The Ethical Consumer. Moral Norms and Packaging Choice. Journal of Consumer Policy, 22(4), 439-460.

Tsireme, A.I., Nikolaou, E.I., Georgantzis, N. & Tsagarakis, K.P. (2012). The influence of environmental policy on the decisions of managers to adopt G-SCM practices. Clean Technologies and Environmental Policy, 14(5), 953-964.

Vandermerwe, S. & M. D. Oliff (1990). Customers Drive Corporations Green. Long Range Planning, 23(6), 10.

Verghese, K. & H. Lewis (2007). Environmental innovation in industrial packaging: a supply chain approach. International Journal of Production Research, 45(18/19), 4381-4401.

Vernuccio, M., Cozzolino, A. & Michelini, L. (2010). An exploratory study of marketing, logistics, and ethics in packaging innovation. European Journal of Innovation Management, 13(3), 333-354.

Wang, R., Wijen, F. & Heugens, P.P.M.A.R. (2018). Government's green grip: Multifaceted state influence on corporate environmental actions in China. Strategic Management Journal, 39(2), 403-428.

Wiesmeth, H., Shavgulidze, N. & Tevzadze, N. (2018). Environmental policies for drinks packaging in Georgia: A mini-review of EPR policies with a focus on incentive compatibility. Waste Management & Research, 36(11), 1004-1015.

Williams, H. & Wikström, F. (2011). Environmental impact of packaging and food losses in a life cycle perspective: a comparative analysis of five food items. Journal of Cleaner Production, 19(1), 43-48.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). Cambridge, Mass: MIT Press.

Yamaguchi, K. & Takeuchi, K. (2016). Consumer preferences for reduced packaging under economic instruments and recycling policy. Waste Management, 48, 540-547.

Zailani, S.H.M., Eltayeb, T.K., Hsu, C. & Choon Tan, K. (2012). The impact of external institutional drivers and internal strategy on environmental performance. International Journal of Operations & Production Management, 32(6), 721-745.

Zhang, B., Bi, J., Yuan, Z., Ge, J., Liu, B. & Bu, M. (2008). Why do firms engage in environmental management? An empirical study in China. Journal of Cleaner Production, 16(10), 1036-1045.

Zhang, Q., He, K. & Huo, H. (2012). Cleaning China's air. Nature, 484(7393), 161-162.

Zhu, Q. & Sarkis, J. (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. Journal of Operations Management 22(3), 265-289.

DECISION SCIENCES INSTITUTE

An Assessment of an International Marketing Multiple-Choice Question Bank Taxonomy

John R. Dickinson
University of Windsor
Email: MExperiences@bell.net

ABSTRACT

Published banks of multiple-choice examination questions and, the focus of this study, the taxonomies into which the questions are classified, are ubiquitous. Despite this, they have been subjected to little published independent assessment. This study assesses the taxonomy of a widely adopted bank of such questions in international marketing. Basic item analysis criteria of item difficulty (percent correct) and item discrimination (point-biserial correlation) are reported broken down by each of two dimensions into which questions are classified by the publisher. Toward establishing norms for such published question banks, comparable results are presented for two other texts.

KEYWORDS: Multiple-choice questions, Taxonomy, Item analysis

Full text and references available upon request.

DECISION SCIENCES INSTITUTE
Analytical Approaches to Tackling Fraud in a Remanufacturing Environment

Aditya Pandit
Northeastern University
Email: pandit.ad@northeastern.edu

Surendra M. Gupta
Northeastern University
Email: s.gupta@northeastern.edu

ABSTRACT

Fraud in remanufacturing is a seldom discussed topic primarily due to the lack of its observable volume. However, it can be foreseen due to the ever-rising demand for remanufactured products. Previous research has targeted related topics that include warranty policies, maintenance and pricing strategies. With the advent of big data and the availability of several methodologies, the identification of fraud is now possible. This study reviews the prevailing theories and methodologies employed in dealing with fraud, highlighting their effectiveness and shortcomings. The motivation is an attempt to outline how these methodologies could be adapted to deal with remanufacturing warranty fraud.

KEYWORDS: End-of-life, Reverse supply chain, Fraud, Remanufacturing, Warranty

INTRODUCTION

Advances in technology have fueled the rise in electronics and thus endowed the customers with access to inexpensive high quality consumer goods. The abundance of high quality and inexpensive products has also changed has also altered the consumers behavior. One noticeable change is that now products are quite often disposed of before they actually spoil or fail. This leads to products becoming obsolete in a much shorter time even though they still have remaining life. Consequently since there is a shortened product life cycle this leads to a steady degradation of natural resources to compensate for the ever increasing demands of customers and introduces even more waste into the environment.

Environmental legislation has encouraged firms into moving away from disposal and more toward End of Life (EOL) strategies. Environmentally conscious manufacturing (Ilgin & Gupta, 2010) can be carried out at any stage of the products life, right from a products design phase to its EOL, which are all instrumental in getting a product to meet environmental standards. By keeping waste to a minimum, the firm can reduce disposal costs and permit requirements, avoid environmental fines, boost profits, discover new business opportunities, rejuvenate employee morale, and protect and improve the state of the environment (Hanna, Newman & Johnson 2000).

One such waste stream arises from the presence of fraudulent activities in the supply chain which not only generates excessive costs but also requires time and resources to deal with. Fraud is defined as “the wrongful or criminal deception intended to result in financial or personal

gain". There may or may not be any actual physical loss or damage in order for one to say that fraud has been committed for e.g. It may cause no loss of money, property or legal right but still be an element of another civil or criminal wrong. From the legal perspective, wrongful deception cases refer to where a fraud victim may sue the fraud perpetrator to avoid the fraud or recover monetary compensation. Whereas a criminal deception case is when a fraud perpetrator may be prosecuted and imprisoned by governmental authorities.

Frauds in the new product industry have been generally well covered in recent literature (Mu & Carroll, 2016; Green & Choi, 1997; Zhang, Hu, & Zhou, 2010) however issues of fraud in the remanufacturing industry have yet to be explored in a meaningful way. Fraud can occur at any stage of the products life cycle such as during the manufacturing process (eg the use of substandard components in product assembly) and even after the products EOL (eg. Improper disposal). This study first briefly reviews many of the research papers that study the issue of fraud in different industries, next we briefly highlight the more prevalent methodologies that are more frequently employed in dealing with fraud, and lastly we discuss how that knowledge maybe be used in curtailing fraud in remanufacturing sector(namely with regard to product warranties).

LITERATURE REVIEW

Frauds can potentially exist in any sectors where there is a monetary transaction and through multiple medium such as through mail, wire phone and more recently the internet. Fraud can also be carried out by almost any of the parties involved. One case is when it is the seller committing the fraud. One way of tackling fraud is through government legislation. In the United States organizations such as the SEC and FTC are responsible for monitoring and punishing fraud. Fraud is also monitored and investigated at the much more local level. Practices such as internal audits and random inspections help in routing out any discrepancies, as it is of personal interest to companies to at least appear to be honest.

When the guilty party is the consumer, frauds are usually routed out through investigations carried out by the other party. Traditionally companies use a predetermined set of business rules to evaluate claims that are sent to them. But smart fraudsters can work around any publicly known set of rules to their own advantage. Approaches using business rules, anomaly detection, advanced analytics and social network analysis into hybrid fraud algorithms to score each claim can free up auditors to look at claims that are a 100% fraudulent and not some random selection of claims (Hawkins et al, 2011).

Conventional logic assumes that fraud occurs on between two players, namely the seller and the consumer; however it is more complex when we talk about fraud in the industry. Often product manufacturers are required to outsource services to outside companies or service providers. In such cases the level of trust the seller places is of importance (Barber et al, 2003).The most prominent type's of fraud are insurance frauds, financial frauds and business frauds and the type of frauds that they deal with are distinct and may offer clues as to how to solve remanufactured related frauds.

Fraud in Healthcare

Fraud can be seen in all types of insurance sectors, but most notably in healthcare. The study of fraud in this area has been off key interest to scholars for many years (Leap, 1948). Fraud in health insurance is the intentional deception or misrepresentation for gaining some shabby benefit in the form of health expenditures. Frauds that exist in the health care insurance sector are estimated to cost between 2 to 10 % of total value (Desjardins Insurance, 2019). Insurers make use of flagging systems (Desjardins Insurance, 2019) as well as random checks to evaluate claims sent to them. Fraud in health care is generally considered to be committed on an individual basis between an insurance holder and provider, however there have been reports of large criminal conglomerates exploiting the relatively transparent claim review process and racking up millions of dollars in fraud claim before being caught (Sparrow, 1997).

To help combat fraud and abuse, the federal government's False Claims Act (FCA) was instituted (Rudman, 2009) and since 1986 the law has helped the government recover more than $30 billion (Mitka 2012). Once someone is caught committing health insurance fraud it is most often up to the legal system and not the insurer to decide a suitable punishment. The Vevera (2015) describes many scenarios of persons convicted of frauds and how the cases were settled. Busch (2012) details the various parties involved in the health insurance system and their exchanges and relative responsibilities and provide details to guide an auditor to deal with all manner of health care frauds.

Data mining tools and techniques can be used to detect fraud in large sets of insurance claim data. Based on a few cases that are known or suspected to be fraudulent, the anomaly detection technique calculates the likelihood or probability of each record to be fraudulent by analyzing the past insurance claims. In one example (Kirlidog and Asuk**,** 2012), data mining was used to tackle the health insurance market in Turkey. A number of studies made use of real world data sets and data mining techniques (Yang and Hwang,2006 ; Bayerstadler, Dijk and Winter,2016). The main objective of these analysis were to narrow the target for detecting fraud and establishing fraud patterns.

Typically fraud detection looks at the data at just the transaction-level, which sometimes isn't enough to detect frauds effectively. Multidimensional data models and analysis techniques have shown to help predict the likelihood of fraudulent in this area (Thornton et al, 2013; Sparrow, 1997).There have been many new methodologies (such as biometrics) that have been developed or are being developed to counteract healthcare fraud.

Fraud in Finance

Financial fraud can be broadly defined as an intentional act of deception involving financial transactions for purpose of personal gain. Financial fraud is a crime, and is also a civil law violation. While all frauds do have a financial component ,financial fraud in the context of this study refer to crimes such as money laundering, credit card fraud, stock market fraud and other frauds of that nature. Zareapoor, and Shamsolmoali (2015) enumerated the challenges and presented a comprehensive review of the common techniques in detecting credit card frauds. They further showed the superiority of one particularly innovative fraud detection technique called the bagging ensemble classifier by comparing its results with more standard techniques.

Many studies have focused on data mining approaches to evaluate credit card fraud. Studies such as Bhattacharyya (2011), Lin et al (2015) and others have shown relative effectiveness of certain data mining techniques over others more conventional approaches. Corporate financial fraud has a severe negative impact on investors and the capital market in general. The current resources committed to financial fraud detection (FFD), however, are insufficient to identify all occurrences in a timely fashion. Methods for automating FFD have mainly relied on financial statistics, although some recent research has suggested that linguistic or vocal cues may also be useful indicators of deception. Tools based on financial numbers, linguistic behavior, and non-verbal vocal cues have each demonstrated the potential for detecting financial fraud. However, the performance of these tools continues to be poorer than desired, limiting their use on a stand-alone basis to help identify companies for further investigation. A review paper (West & Bhattacharya, 2016a) pointed out that many fraud studies neglected to address the association among fraud types. West and Bhattacharya (2016) presented an analysis of existing fraud detection literature based on key aspects such as detection algorithm used, fraud type investigated, and performance of the detection methods for specific financial fraud types. West and Bhattacharya (2016b) investigated several important fraud detection experimental issues including problem representation, detection algorithms, feature selection, and performance metrics. They noted that the techniques were not all equally sensitive to all performance areas.

Fraud in automotive industry and other frauds

Actual criminal Automotive “fraud” amounts to probably less than 1 percent, but the gray area known widely as “soft fraud” that may be upward of 30-40 percent in some areas (Derrig, 2006) of Massachusetts. In order to address automotive fraud, insurers use four approaches. First, insurers investigate fraud and buildup. An early study (Derrig, 1994) indicated that investigation of suspected fraudulent claims reduced claim payments by about 18 percent. Insurers in the automotive industry often outsource their claims to third party service providers and as such this creates another avenue for potential fraud. Caldeira et al (2015) studied strategies of how to select claims that need auditing, with an aim at detecting frauds in third party claims, by using strategies like logistic models and neural network to estimate the probability of a fraud detection. Strategies such as Iterative Assessment Algorithm (Šubelj et al, 2011) and discrete choice models (Artis et al, 1999; Artis et al, 2002) were shown to be useful as an expert system for detection and subsequent investigation, of groups of collaborating automobile insurance fraudsters. Nian et al (2016) showed how fraud detection can be approached as an anomaly ranking problem. It is often assumed that when a claim is examined by an inspector and found to be true, that judgment is assumed to be accurate all the time. Caudill et al (2005) used a modified logit model that allows for the possibility that some claims classified as honest may actually, be fraudulent. This model was applied to the auto industry in Spain, and the results showed its ability to identify misclassified claims.

FRAUD AND REMANUFACTURING

Based on reviewing literature on how fraud is dealt with in other industries, a number of conclusions and parallels may be drawn. The warranty service chain for remanufactured products in particular bears a great resemblance to the service chains for the other industries examined in the review.

Figure 1: Information flow between key parties involved in a warranty servicing chain

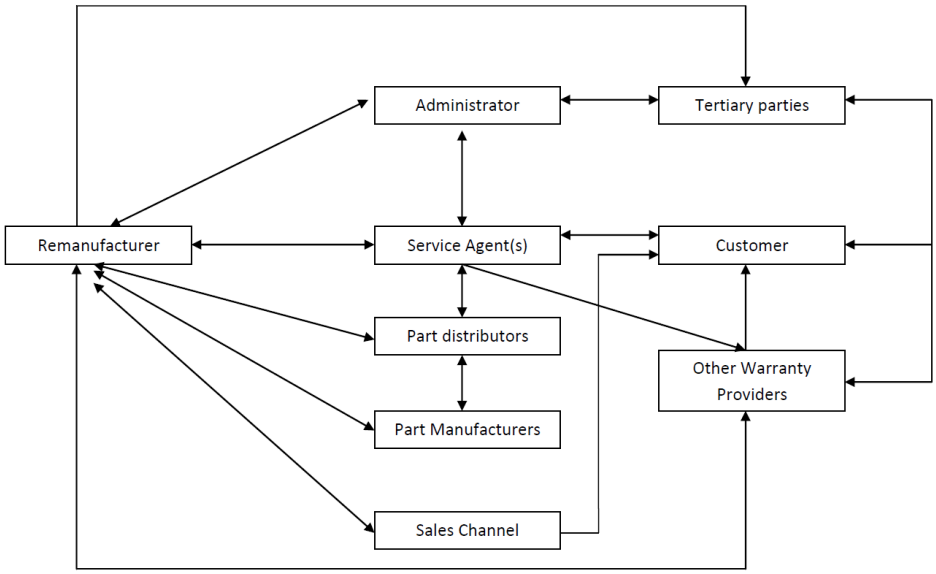


Figure 1 shows the warranty supply chain for remanufactured products (Kurvinen, Töyrylä & Murthy, 2016). When a warranty provider offers a warranty on a remanufactured product to a customer, there are a host of other parties that are also involved in the service and thus may also be involved in any potential fraud. In a typical warranty service system, when a product is rendered nonfunctional, it is inspected to determine the cause of failure. The information about any such failure is transmitted to the service personnel (third party) who conducts the required service operations; for example, replacing the failed component or components. After this process, failed products are transferred to the service facility. After the maintenance process takes place, the products are brought back to working condition. Once the maintenance service operations are complete, the products are returned to the customers.

Dealing with frauds affecting the primary parties (Remanufacturer, service agent and customer) presents varying challenges. If the fraud source is the service provider, the audit of claims mirrors the same steps as for the review of claims in the health care and automotive field and as such data mining and neural network methods would present the best way of dealing with such frauds. Similarly if a service agent overcharges a claim, methodologies (anomaly detection) that are employed in the financial sector would assist in monitoring such frauds. If the source of fraud is the customer, the process of assessing the validity of a customer claims is comparable to that of health care and automotive industry and it follows that similar techniques would be effective.

The study considered numerous fraud scenarios and checked compatibility against various fraud modeling and detection techniques. It was determined that there was no “one size fits all” methodology that would properly encapsulate the problem. The merits of certain techniques, some that were not covered in the review were also noted. With regards to fraud modeling, game theory was shown to be useful in determining a parties (players) optimum decision given the pay off (fraud amount) (Jack & Murthy, 2017). One study (Pandit & Gupta, 2018a) used Nash equilibrium to contrast the optimum decision between the remanufacturer and the service

agent for an overcharging warranty fraud scenario, and examined the relationship between fraud amount, penalty value, inspection cost and player risk. Discrete event simulation was also shown to achieve some measure of success in properly recreating the fraud scenarios and was also the most easily adaptable to fit new scenarios. This was shown to be the case when a fraud model developed in order to model service agent fraud (Pandit & Gupta, 2019a), was able to be adapted to simulate customer driven fraud as well (Pandit & Gupta, 2018b). As with many other industries, the Internet of things shows the greatest promise in both fraud detection and prevention. Past studies have shown the usefulness of sensor implementation (Ondemir & Gupta, 2014) in dealing with disassembly line quality issues and this would also extend to fraud detection (location data, temperature data etc., would assist in determining if a product was actually serviced and or used within recommended guidelines). A theoretical model was proposed by Pandit & Gupta (2019b), which considered the benefits of incorporating sensors into products would have with regards to reducing maintenance service times and inspection costs. A study by Pandit & Gupta (2020), which attempted to use a neural network model in a remanufactured product warranty scenario, noted that the main hindrance to researching remanufactured product fraud is the lack of readily available data sets to conduct said research.

CONCLUSION

This paper discussed the concept of fraud and the forms it appears across many industries. Concepts and trends in fraud research were reviewed, with a few cases being highlighted. This aided in outlining the available tools that could be employed in tackling fraud in the remanufacturing sector. Game theory, discrete event simulation and neural networks models are some of the more promising methodologies that were deemed to be of use in detecting and dealing with these frauds.

REFERENCES

Artis, M., Ayuso, M., & Guillen, M. (2002). Detection of automobile insurance fraud with discrete choice models and misclassified claims. *Journal of Risk and Insurance*, 69(3), 325-341.

Artís, M., Ayuso, M., & Guillén, M. (1999). Modeling different types of automobile insurance fraud behaviour in the Spanish market**.** *Insurance Mathematics and Economics*, 24(1), 67-81.

Barber, K. S., Fullam, K., Kim, J. (2003). Challenges for Trust, Fraud and Deception Research in Multi-agent Systems. The Laboratory for Intelligent Processes and Systems Electrical and Computer Engineering, the University of Texas at Austin.

Bayerstadler, A., Dijk, L. V., & Winter, F. (2016). Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. *Insurance: Mathematics and Economics*, 71, 244–252.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J.C. **(**2011). Data mining for credit card fraud: A comparative study**.** *Decision Support Systems*, 50(3), 602-613.

Busch, R.S. (2012). Healthcare fraud auditing and detection guide. Hoboken, N.J. John Wiley & Sons.

Caldeira, A. M., Gassenferth, W., Machado, M.A.S., & Santos, D. J. (2015). Auditing Vehicles Claims Using Neural Networks. *Procedia Computer Science*, 55, 62-71.

Caudill, S.B., Ayuso, M., & Guillén, M. (2005). Fraud detection using a multinomial logit model with missing information. *Journal of Risk and Insurance*, 72(4), 539-550.

Derrig, R. A., Johnston, D. J., & Sprinkel, E. A., 2006, "Auto Insurance Fraud: Measurements and Efforts to Combat It.(Author abstract)", *Risk Management & Insurance Review*, Vol.9(2), pp.109(22).

Derrig, R. A., Weisberg H. I., & Chen, X., 1994, "Behavioral Factors and Lotteries Under No-Fault With a Monetary Threshold: A Study of Massachusetts Automobile Insurance Claims", *Journal of Risk and Insurance*, Vol.61, pp. 245-275

Desjardins Insurance, (2019).Understanding the effects of fraud and abuse on your group benefits plan. 360 Health Insurance.

Green, B. P., & Choi, J. H. (1997).Assessing the risk of management fraud through neural network technology. *Auditing-A Journal Of Practice & Theory*, *16*(1), 14-28.

Hanna, M. D., Newman, R. W., & Johnson, P. (2000). Linking operational improvement through employee involvement. *International Journal of Operations and Production Management*, *20* (2), 148–165.

Hawkins, T., Arnum, E., & Froning, D. (2011). Using Analytics to Uncover Claims Errors and Fraud. *Insights from the third webcast in the 2011 Supply Chain Leadership Series*, SAS Conclusions Paper.

Ilgin, M.A. & Gupta, S.M. (2010). Environmentally conscious manufacturing and product recovery (ECMPRO): A review of the state of the art. *Journal of Environmental Management,* 91 (3), 563-591.

Jack, N., & Murthy, D. N. P. (2017). Game theoretic modelling of service agent warranty fraud. *Journal of the Operational Research Society 68*(11), 1399-1409.

Kirlidog, M., & Asuk, C. (2012). A Fraud Detection Approach with Data Mining in Health Insurance. *Procedia - Social and Behavioral Sciences* , 62, 989-994.

Kurvinen, M., Töyrylä, I., & Murthy, D. N. P. (2016). *Warranty fraud management: reducing fraud and other excess costs in warranty and service operations*. Hoboken, New Jersey: Wiley.

Leap, T. L. (1948). Phantom billing, fake prescriptions, and the high cost of medicine health care fraud and what to do about it. ProQuest (Firm) Ithaca: ILR Press 2011.

Lin, C.C., Chiu, A. A., Huang, S. Y., & Yen, D. C. (2015) .Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89, 459-470.

Mitka, M. (2012). Health Care Fraud. *The journal of the American Medical Association*, 307 (6), 553-553.

Mu, E., & Carroll, J. (2016). Development of a fraud risk decision model for prioritizing fraud risk cases in manufacturing firms. *International Journal of Production Economics*, 173, 30-42.

Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly**.** *The Journal of Finance and Data Science*, 2(1), 58-75.

Ondemir, O., & Gupta, S. M. (2014). Quality management in product recovery using the Internet of Things: An optimization approach. *Computers in Industry*, 65(3), 491-504.

Pandit, A., & Gupta, S. M. (2018a). Warranty fraud in remanufacturing. Proceedings of the Global Interdisciplinary Conference: Green Cities, June 27- 30, Nancy, France.

Pandit, A. & Gupta, S. M. (2019a).Warranty Fraud in a Remanufacturing Environment, in *Responsible Manufacturing - Issues Pertaining to Sustainability*, Edited by A. Y. Alqahtani, E. Kongar, K. K. Pochampally and S. M. Gupta, CRC Press, 11, 241-261, ISBN: 978-0815375074

Pandit, A., & Gupta S. M. (2018b). Mitigating customer driven fraud for remanufactured products in reverse supply chain. *Proceedings of 16th International Logistics and Supply Chain Congress*, Denizli, Turkey.

Pandit, A., & Gupta, S. M. (2019b).Impact of warranty fraud on remanufactured products. *Proceedings of International conference on remanufacturing*, Amsterdam, Netherlands

Pandit, A. & Gupta, S. M., (2020). Predicting Remanufactured Product Fraud using Neural Networks. *Proceedings of the 2020 Annual Meeting of the Northeast Decision Sciences Institute*, Cambridge, Massachusetts, 712-719.

Rudman, W. J., Eberhardt, J. S., Pierce, W., & Hart-Hester, S. (2009). Healthcare fraud and abuse. Perspectives in health information management, 6, 1-1.

Sparrow, M. K. (1997). A criminal's dream. Automated payment systems, designed for honest providers, are easy targets for fraud. Modern Healthcare.

Šubelj, L., Furlan, Š., & Bajec, M., (2011)**.** An expert system for detecting automobile insurance fraud using social network analysis**.** *Expert Systems With Applications*, 38(1), 1039-1052.

Thornton, D., Mueller, R, M., Schoutsen, P., & van Hillegersberg, J. (2013). Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection. *Procedia Technology,* 9, 1252-1264.

Vevera, A.V., (Manager), (2015). Healthcare Fraud Investigations. 22 (1), 165-174.

West, J., & Bhattacharya, M. (2016a). Intelligent financial fraud detection: A comprehensive review. *Computers & Security,* 57, 47-66.

West, J., & Bhattacharya, M. (2016b). Some Experimental Issues in Financial Fraud Mining. *Procedia Computer Science*, 80, 1734-1744.

Yang, W. S., & Hwang, S. Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, 31, 56–68.

Zareapoor, M., & Shamsolmoali, P. (2015). Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier**.** *Procedia Computer Science*, 48, 679-685.

Zhang, H., Hu, Y., & Zhou, Z. (2010). Prevention of resource trading fraud in manufacturing grid: a signalling games approach. *International Journal of Computer Integrated Manufacturing*, *23* (5), 391-401.

DECISION SCIENCES INSTITUTE
Application of Two-Stage Network DEA with Transformed Decision Making Units
To Biofuel Supply Chain Network Design

Jae-Dong Hong
Industrial Engineering
South Carolina State University
Orangeburg, SC 29117
Email: jhong@scsu.edu

ABSTRACT

This paper proposes an innovative two (2)-stage network data envelopment analysis (DEA) with transformed decision-making units for designing efficient and robust biofuel supply chain network system (BSCN) under the risk of disruptions. The applicability of the proposed method is demonstrated through a case study. It is observed that the proposed method performs well in identifying efficient and robust biofuel supply chain network schemes.

KEYWORDS**:** Two (2)-Stage Data Envelopment Analysis, Transformed Decision Making Unit, Biofuel Supply Chain Network

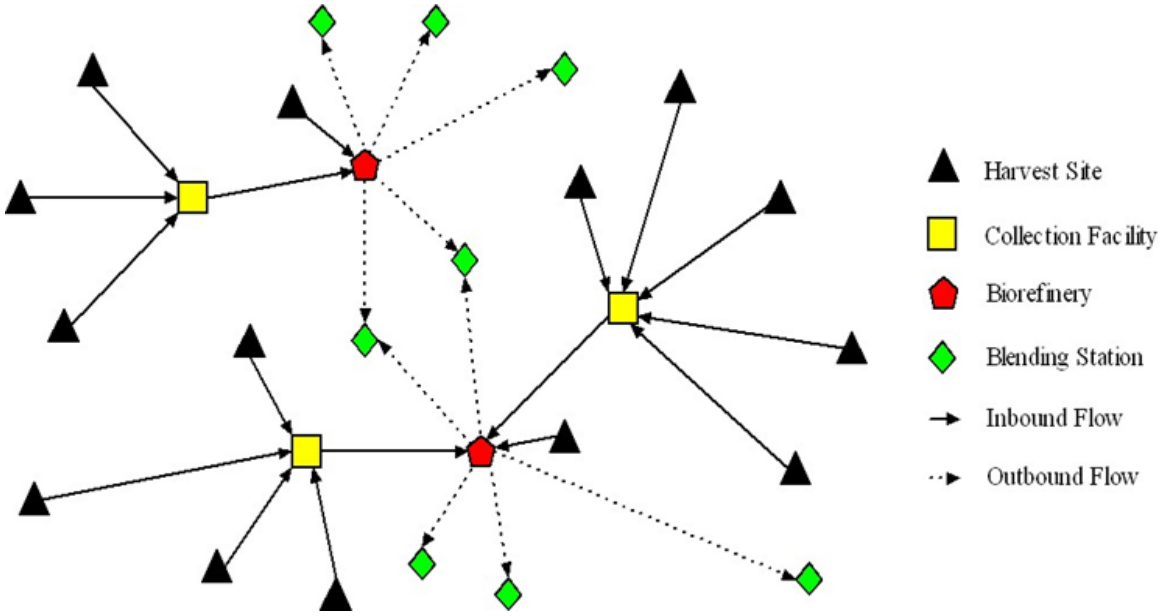
INTRODUCTION

Biofuel is a kind of renewable energy made from biomass. As world demand for energy increases, biomass-based biofuel production has been recognized as an essential alternative that will reduce the nation's dependency on the limited fossil fuel source. Besides, biofuel production has a positive impact on the economy, environment, and society. Biofuel, such as biodiesel and bioethanol, refers to liquid and gas fuels used for transportation and industrial processes. Biofuel has been considered as a potential energy source and has gathered much attention. In this vein, Balaman and Selim (2015) emphasize that design, operation, and management issues in renewable energy supply chains are increasingly gaining importance in recent years in parallel with the rising interest in renewable energy sources. Balaman (2019) also stresses that developing integrated frameworks to design resilient and efficient supply chains should be one of the core priorities to boost the research on this topic. This paper attempts to propose such an integrated framework to design biofuel supply chain network (*BSCN)* systems by applying two-stage network (TSN) data envelopment analysis (DEA) methods with transformed decision-making units (DMUs).

Most of the *BSCN* models in the references reviewed by Atashbar et al. (2018) have considered four different types of biofuel facilities. They are a supply point - farm or harvest site (HS), a storage point – storage facility (SF), and a production point - biorefinery (BRF), and a demand point – blending station (BS). Eksioglu et al. (2009) consider the *BSCN* structure, which is depicted in Figure 1. An SF is a potential site to store and preprocess (e.g., compress) biomass to a more valuable density and/or to pre-treat biomass to make a better-quality biomass feedstock, so that they can be transported more cost-effectively. Direct transportation of biomass from an HS to a BRF is allowed. The resulting transportation cost is usually higher than going through the SF, since the direct shipping of biomass from an HS to the BRF requires more space (due to the low biomass density) and more operations and preparation to be processed into biofuel (see van Dyken et al., 2010). Thus, the conversion rates to biofuel of biomass feedstocks shipped from SF to BRF are usually higher than those for biomass

feedstocks shipped from HS to BRF. Biofuel is transported from BRFs to BSs to be blended with gasoline before being distributed to gas stations.

Figure 1: Biofuel Supply Chain Network.



Various mathematical programming models in the literature related to the *BSCN* system have a single objective of minimizing total costs or of maximizing total profit. In terms of optimization criteria, minimizing the total relevant cost is the most preferred optimization criterion, followed by maximizing the total profit. Parker et al. (2008) develop a profit-maximizing model of a biomass hydrogen industry from field to vehicle tank, estimating the economic potential for hydrogen production from two waste biomass resources in Northern California. Later, Parker et al. (2010) develop a mixed-integer nonlinear programming model to maximize profit for finding the most efficient and economical configuration of the whole pathway. Lim and Lam (2016) introduce the biomass element life cycle analysis (BELCA) via biomass element characteristics and demonstrate that the BELCA approach would improve the existing biomass to biofuel supply chain system.

The majority of models on the topic of *BSCN* design problems are modeled as the single-stage network (SSN) model with single or multiple objectives (see Chanes et al., 1978). Recently, Hong (2020) models the biofuel supply chain network system as the two-stage network processes and analyzes it using the two-stage network (TSN) DEA. This paper attempts to propose a procedure of the TSN DEA with transformed DMUs through the multiple-criteria DEA.

BACKGROUND

It is assumed that SFs can be located at any HS and that a BRF can only be built at candidate BRF location since BRF locations must satisfy some realistic requirements. This is a reasonable assumption at the strategic planning stage for the biofuel logistics model. It may be difficult to decide potential SF locations, which are not HSs since the assignment of HSs to an SF is not known. It is assumed that if a facility is disrupted, biomass would not be used to produce biofuel. The five goals are considered as major performance measures. To

accommodate the five goals in one objective function, we use a goal programming (GP) approach as a tool for designing the effective *BSCN*. The typical GP model allows the decision-maker to assign weights to the deviational variables in the objective function to better reflect the importance and desirability of deviations from the various goals.

DEA is a performance evaluation technique based on a non-parametric linear programming (LP) approach for frontier analysis. DEA is an approach for identifying best practices of peer decision-making units (DMUs) in the presence of multiple inputs and outputs. Cook and Zhu (2014) state that the network DEA approach is an essential area of development in recent years. All the models developed so far on the topic of the *BSCN* design problem can be of the single-stage model with single or multiple objectives. Monfared and Safi (2013) state that the single-stage DEA model considers a DMU as a 'black box' and neglects intervening processes, i.e., different series or parallel functions. They continue to say that the black box approach provides no insights regarding the inter-relationships among the components' inefficiencies and cannot provide specific process guidance to DMU managers to help them improve the DMU's efficiency. These single-stage models neglect internal linking activities.

The *BSCN* design could be considered as the two-stage network problem, where the total cost and demand-weighted covered distance are the inputs to the first stage, which generates the amount of collected biomass feedstocks as an output. Then, the output from the first stage becomes an input to the second stage to produce the two outputs, the amount of biofuel production and a population that is free of air pollution generated by the biorefineries. In other words, the amount of collected biomass feedstocks is measured using the total logistics cost, maximum demand-weighted distance, and the outputs are the expected amount of biomass feedstocks. In the second stage for biofuel production and population free from air pollution, the expected amount of biomass feedstocks delivered to the BRFs are then used as an input, while the amount of biofuel production and a population who are free of air pollution generated by the biorefineries are used as outputs.

Each biofuel network configuration, which is obtained by solving a GP model and can be considered as a DMU, are converted from DMU in TSN into SSN DMU through a multiple-criteria DEA. After that, SSN-DEA to evaluate and identify efficient ones among various network schemes. The proposed method enables the efficiency of the supply chain network schemes to be more accurately evaluated. It would provide a robust ranking system for identifying more efficient network schemes.

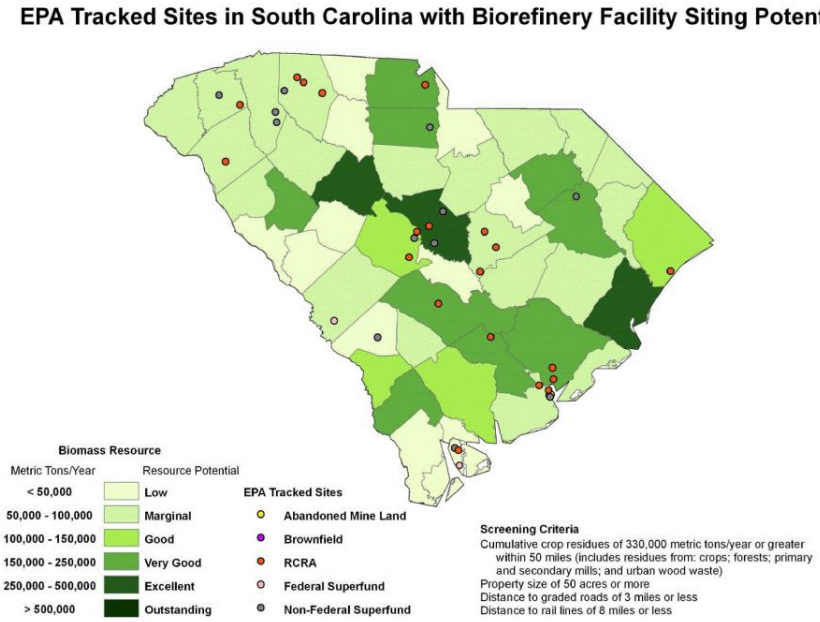
CASE STUDY

This study modifies and conducts the case study considered by Hong et al. (2014), using the proposed procedure. Based on *EPA tracked sites in South Carolina (SC) with biorefinery facility siting potential* (2013) as shown in Figure 2, sixteen (16) counties, whose biomass resources are classified 'good' or better, are selected as the harvesting sites (HSs). Then, we choose one city from each county using a centroid approach and consider it as a storage facility (SF) location potential. Five (5) locations and ten locations (10) are considered throughout SC as candidate sites for BRFs and BSs, respectively. The potential locations for BRFs are selected based upon low population density, easy access to interstate highways, etc.

The GP model with five goals is formulated and solved for various values of the weight. The centralized model for the TSN for the entire DMUs, which are generated by solving the GP model, is applied, and the twenty (20) most efficient DMUs are selected based on centralized

efficiency score. We observe that the rankings by the efficiency score by the proposed method are much less affected, yielding more robust and stabilized rankings for the top DMUs.

Figure 2: EPA Tracked Sites and Candidate Sites for Biofuel Facilities.



SUMMARY AND CONCLUSIONS

Contrary to the traditional logistics network models focusing primarily on cost-efficiency, this paper considers five (5) performance measures simultaneously to design more balanced biofuel supply chain (*BSCN*) networks. A goal programming (GP) technique is applied to handle these five performance measures. Solving the GP model would provide decision-makers with various *BSCN* alternatives or schemes for multiple values of the weights assigned to each performance measure. This paper proposes an innovative procedure of designing efficient and robust *BSCNs* generated by solving the GP. To compensate for the weakness of the TSN DEA model, the proposed method is developed to transform the TSN DEA into a single-stage network (SSN) DEA model. The cross efficiency DEA is applied to evaluate the transformed DMUs with two inputs and one output and rank the various alternatives to compensate for the critical weakness of the classical DEA method. For designing a supply chain network, the proposed DEA method might reveal different network schemes, which could be more efficient and robust alternatives than the TSN DEA alone.

ACKNOWLEDGMENTS

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Evans-Allen project number SCX-313-04-18.

REFERENCES

References available upon request

DECISION SCIENCES INSTITUTE

Business Analytics and Firm Performance: The mediating role of Resource Reconfiguration

Phuoc Pham
Department of Management,
College of Business and Public Management, West Chester University, Pennsylvania, USA
ppham@wcupa.edu

Jeen-Su Lim
Department of Marketing & International Business,
College of Business and Innovation, University of Toledo, Toledo, Ohio, USA
jeen.lim@utoledo.edu

ABSTRACT

Business analytics receives increased attention from practitioner and scholars recently. However, there is limited knowledge of the mechanism by which business analytics help firms manage their operational risks to stay competitive. Through the lens of resource-based view and dynamic capability view, this study explores the relationships among business analytics, resource reconfiguration and firm performance. This study posits that the resource reconfiguration mediates the relationship between business analytics and firm performance. The proposed relationships are tested using survey data. The results from structural equation modeling support the proposed model.

KEYWORDS: Business Analytics, Resource Reconfiguration, Firm Performance, and Dynamic Capability

INTRODUCTION

Business analytics receives increased attention from practitioner and scholars in the past few years. Especially in the areas of Industry 4.0, smart devices produce a huge volume of industrial data in every second. As a result, firms are under stress of converting these big data into valuable insights to improve their decision-making processes and to generate business values. It is expected that firms investing in business analytic system would reduce the stress and improve firm performance.

Firms' growing interests at the moment is how business analytics can support them to monitor or identify real-time events (e.g. disruption) and to suggest proper actions during and after events (Ivanov & Dolgui, 2020; Wamba, Ngai, Riggins, & Akter, 2017). One of the most recent events causing supply chain disruptions leading to lower firms' competitive position in their markets is the global outbreak by Coronavirus disease (COVID-19). While digital technologies and business analytics have been evolving, there exists limited understanding of how business analytics help firms utilize this technology to manage their operational risk in theory and applications (Araz et al., 2020; Ivanov & Dolgui, 2020).

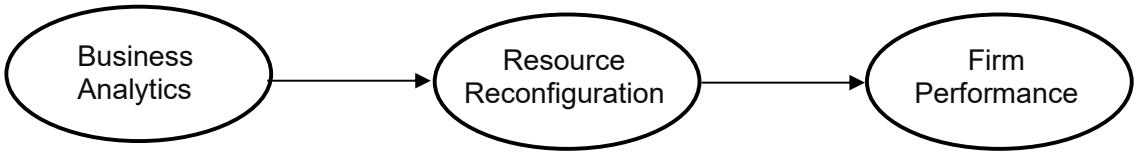
This study addresses this gap by empirically examining how business analytics improve firms' business performance through their resource reconfiguration. Drawing on the theory of resource base view and dynamic capability, this study theorizes that business analytics can be

viewed as a firm resource that can build up its resource reconfiguration capability. Furthermore, in line with extant research which suggests a firm's ability to reconfigure its resources is critical to its resilience (Ambulkar, Blackhurst, & Grawe, 2015) as well as its maintenance of competitive position (Vanpoucke et al., 2014), we postulate this ability of resource configuration would improve firm performance.

LITERATURE REVIEW

The theoretical framework of this study is built on Resource-Based View (RBV) and Dynamic Capability View (DCV). To stay competitive in today's turbulent and uncertain environment, firms have to develop dynamic capabilities (Schilke, 2014), which help them to build, incorporate, and reconfigure their resources and capabilities to cope with changing conditions (Teece et al., 1997). But to build up dynamic capabilities, firms need resources including tangibles (e.g. infrastructure, systems), intangibles (e.g. culture, structure, knowledge) (Grant 1991). Following that logic, this study considers business analytics as a resource for firms to manage their resource reconfiguration capability needed to achieve firm performance. Thus, the research model is constructed as follows: business analytics is an antecedent of resource reconfiguration, which then enhances firm performance (Figure 1).

Figure 1: Research model



Business Analytics

The concept of business analytics (BA) has been around in the literature (Holsapple et al., 2014), yet it gains more and more scholarly attention thanks to the development of its abilities to extract values from big data. While the term BA has been used interchangeably with big data analytics, data analytics, the common focus is the capability to allow firms to collect, store, and analyze big data to make better business decisions (Agarwal and Dhar, 2014; Chen et al., 2012; Souza, 2014).

BA in this study is viewed as firms' abilities to use tools, techniques, and processes to collect, organize, process, and analyze data to drive or assist business decisions (Srinivasan & Swink, 2018). Indeed, working with the accumulated big data characterized by ever-increasing volume, variety, and velocity of data (Sharda et al., 2016), firms have to rely on a combination of information tools/technologies and utilize advanced analytical techniques (Davenport and Harris, 2007).

BA can be categorized into descriptive, predictive, and prescriptive (Robinson, Levis, & Bennett, 2010). Descriptive analytics tries to answer such questions as what has happened and what is happening by using business intelligence, data mining, web analytics, etc. to get context and trending information. Predictive analytics focuses on question of what could happen by utilizing statistical models, machine learning, and forecasts, etc. to suggest an estimation of future occurrences and the corresponding causes. Prescriptive analytics aims to address the question

of what firms should do by involving optimization, simulation, artificial intelligence, and case-based reasoning, etc. to suggest courses of actions and indicates the possible outcome of each decision. Firms utilize these tools and processes to gain usable information and insights for their decision-making process.

Many studies have shown that firms equipped with BA capabilities and applications achieve better business value and organizational performance (Torres, Sidorova, & Jones, 2018; Wang, Yeoh, Richards, Wong, & Chang, 2019) and supply chain performance (Trkman, McCormack, De Oliveira, & Ladeira, 2010). The adoption of business analytics (BA) is proved to positively influence business process performance, which in turn positively impacts firm performance (Aydiner, Tatoglu, Bayraktar, Zaim, & Delen, 2019). BA is also found to strengthen firms' innovation by directly enhancing the firms' environmental scanning ability (Duan, Cao, & Edwards, 2020). BA supports firms to bolster its rational strategic decision making and to rely less on intuitive one (Cao et al., 2019).

Resource Reconfiguration

As referred to one of the three main pillars of dynamic capabilities, reconfiguration is the ability to continuously align and realign operational practices (Teece, 2007). In this study we adopt the definition of resource reconfiguration that is viewed as a firm' capabilities to "reconfigure, realign and reorganize their resources in response to changes in the firm's external environment" (Ambulkar, Blackhurst, & Grawe, 2015; p. 112). Working in an ever-changing, volatile environment, firms have to continuously reconfigure resources to fit changing situations to maintain their competitive position (Vanpoucke et al., 2014). By constantly acquiring, dropping, and rearranging their resource base, firms can increase their innovation and other necessary capabilities (Marsh and Stock, 2006; Helfat et al., 2007). In addition, under high impact supply chain disruptions, resource reconfiguration is found to be intermediary mechanism that enable firms to mitigate supply chain disruptions and improve their resilience (Ambulkar et al., 2015). Resource reconfiguration works best in market development situations, helping strong firms to thrive and weak firms to retrench (Chakrabarti, Vidal, & Mitchell, 2011).

Hypotheses

Business Analytics' Impact on Resource Reconfiguration

It is imperative that firms have timely and relevant information to be able to renew and reconfigure their resource base (Côrte-Real et al., 2020). BA can provide business management with visualized reports in the form of dashboards or charts (e.g. pie charts, heat maps, geo-maps, and other charts) to promote quick understanding and sharing (Davenport, 2014). In addition, information presented in a complex analysis report is not preferable. BA can also reduce the information lead time and improve the reliability of the information. When markets are volatile, it is critical that firms can quickly access valuable insights to reduce uncertainty (Srinivasan and Swink, 2018).

A BA system can help firms discover new threats and opportunities to renew, reconfigure, or realign their resources. Descriptive analytics would offer a comprehensive picture of market situation. This information, then, becomes data source for predictive analytics to project future situation. Finally, prescriptive analytics would suggest firms to take alternative actions to achieve business goals under limited resources. Text mining techniques is used to analyze

social media data (e.g. twitter feeds) to forecast information related to product features, product competition and market adoption (Tuarob and Tucker, 2013). Firms then can estimate optimal marketing budget and redirect resources in the right market segment (Lim, Pham, & Heinrichs, 2020). In another example, utilizing optimization models, a manufacturing company is able to choose among different suppliers that could decrease cost and increase profits (Taleizadeh et al., 2015). Therefore, the following hypothesis is presented.

Hypothesis 1: Business Analytics is positively associated with resource reconfiguration.

Resource Reconfiguration's Impact on Firm Performance

According to RBV, while firms achieve competitive advantage by holding differentiated resources, they can only maintain competitive advantage by continuously reconfiguring their resources (Vidal & Mitchell, 2015). Davis et al. (2009) also argue that firms possess the ability to manage resources and reconfigure them along with their environmental setting would achieve and sustain their superior performance. This happens because capabilities that enable firms to adapt to the changing environment are developed by firms' acquiring, shedding, and reorganizing their current resource base, which open doors for new business applications (Eisenhardt and Martin, 2000). For example, small firms are found to use their resources to sense underserved market niches, then reconfigure and redeploy the delivery of value to these customers to stay competitive (Davis & Bendickson, 2020).

Firm's ability to reconfigure their resources would help them quickly develop new product applications and innovations. Software industry is a typical example. Effective software designers usually build modules of routines first. Then these modules can be combined and recombined in different configurations to create different software versions (Sanchez and Mahoney, 1996). This reflects Schumpeter (1934)'s view of innovation which is the "carrying out of new combinations" or the product of recombining current systems, resources and technologies in new ways. These unprecedented combinations are to solve the current problems. Thus, through reconfiguring current resources and processes, firms improve their innovation capability (Helfat et al., 2007), which, in turn, help to enhance firms' performance (Anderson et al., 2014). Therefore, we hypothesize that:

Hypothesis 2: Resource reconfiguration is positively associated with firm performance.

The mediating role of Resource Reconfiguration

As suggested in RBV and DCV, there should be a path of resource – capability – performance. BA can be considered as organizational resource. BA provides firms with valuable insights to help them detect threats and opportunities from their environment. Thus, BA would improve firms' ability to reconfigure their resource base to mitigate the threats or pursue opportunities. On the other hand, the resource reconfiguration, in turn, would help firms to recover from disruption and improve organizational performance. Also, there are several previous studies confirming the direct link between BA and firm performance in term of delivery performance (Srinivasan & Swink, 2018) and competitive advantage (Dubey et al., 2019). Therefore, we present the following hypothesis.

Hypothesis 3: Resource reconfiguration mediates the relationship between BA and firm performance.

METHODS

Measurement

The measurement items for the construct BA is adapted from existing scales in previous studies (Dubey et al., 2019; Srinivasan & Swink, 2018). The items measuring the construct of resource reconfiguration (RR) are adapted from Ambulkar et al. (2015). And Firm Performance (FP)'s measurement scale is adapted from literature (Vonderembse, Tracey & Lim, 1999; Tracey, Lim, & Vonderembse, 2005). A seven-point Likert scale is used to indicate this construct ranging from 'strongly disagree' to 'strongly agree'.

Sample and Procedure

This research is conducted using online-based survey created on Qualtrics platform. Also, we employed Qualtrics Research Core service to collect 250 respondents. Qualtrics is a popular online survey platform that is increasingly used in OM research (Gligor, Holcomb, & Stank, 2013; Thornton, Esper, & Autry, 2016). Participant panel was contacted by Qualtrics base on the requested characteristics and knowledge related to the research topic. Email invitations were sent to its panel members by Qualitrics.com with the statement of the purpose of the study.

Considering the nature of the current study, qualified respondents are screened through a careful procedure. The target respondents have a job title of "manager" and above including "director", "president", "vice president", "CTO" or "CEO", etc. and have enough knowledge of their organization's technology adoption. After 250 completed surveys are collected, data screening was conducted to make sure that the data are clean, usable, and suitable for statistical analysis. After this process, 32 responses were removed because of missing data and inattentive respondents. Thus, the final sample of this study was 218 responses.

Measurement Model

Confirmatory methods by AMOS 26 are utilized in this study to examine the validity of the first order measurement models. All these constructs were modeled with reflective measurement items. Item loadings, reliability, mean, and standard deviation are shown in Table 1. The convergent validity was checked by looking at the item loadings (Anderson & Gerbing, 1988). All the items were loaded well (>0.63) with their corresponding factors as shown in Table 1.

Table 1. Confirmatory Factor Analysis: Construct Descriptive Statistics, Factor Loadings, and Cronbach α Values

Constructs	Mean	Standard deviation	Loading
Business Analytics Tools (BA) (α = 0.818)			
BA1	5.82	1.121	0.66417
BA2	5.63	1.181	0.81929
BA3	5.66	1.201	0.70509
BA4	5.69	1.169	0.73322
Resource Reconfigurability (RR) (α = 0.760)			
RR1	5.67	1.051	0.69841

RR2	5.78	1.062	0.74027
RR3	5.80	1.005	0.71269
Firm Performance (FP) (α = 0.757)			
FP1	5.62	1.059	0.72733
FP2	5.62	1.023	0.79424
FP3	5.71	1.004	0.63028

Also, as shown in Table 2, the discriminant validity of the constructs was established because the square root of AVE value of each construct was higher than that construct's correlation values with other constructs (Fornell & Larcker, 1981). Regarding model fit, CFI (=0.986), SRMR (=0.043), RMSEA (=0.039), and PClose (=0.700) indicate great fit (Hair et al., 2010; Hu & Bentler, 1999).

Table 2. Correlation Matrix and Validity Measures

	CR	AVE	MSV	MaxR(H)	BA	RR	FP
BA	0.822	0.537	0.514	0.833	**0.733**		
RR	0.761	0.515	0.514	0.762	0.717***	**0.717**	
FP	0.762	0.519	0.376	0.777	0.484***	0.614***	**0.720**

*** p < 0.001

Structural Model

All the constructs and their pre-specified relationships were entered in the model and the model fit was checked. The result showed that CFI (=0.987), SRMR (=0.045), RMSEA (=0.038), and PClose (=0.732) suggest excellent fit (Hair et al., 2010; Hu & Bentler, 1999).

After that, the results of all hypotheses were examined. Hypothesis 1 suggests that Business Analytics is positively associated with Resource Reconfiguration. The test result shows that Business Analytics is statistically significant and has a positive impact (β =0.73, p<0.001). This result supports Hypothesis 1. Hypothesis 2, which states that Resource Reconfiguration is positively associated with Firm Performance, received strong support (β =0.63, p<0.001). Finally, Hypothesis 3, which posits that Resource Reconfiguration mediates the relationship between Business Analytics and Firm Performance, was tested. Mediation analysis indicated that BA has a positive effect on FP through the mediation role of RR (Table 3). Thus, Hypothesis 3 is supported.

Table 3: Resource Reconfiguration Mediating effect result

Indirect Path	Unstandardized Estimate	Lower	Upper	P-Value	Standardized Estimate
BA --> RR --> FP	0.447	0.362	0.545	0.001	0.618***

*** p < 0.001

DISCUSSION AND CONCLUSIONS

This paper contributes to existing literature in the following ways. First, it examines the mechanism that business analytics helps firms improve firm performance in term of sales growth. Second, this research is one of the few empirical studies that looks at the resource reconfiguration and examines its mediating role on the relationship between business analytics and firm performance. Finally, it enriches the knowledge of RBV and DCV by viewing BA as a resource to build up resource reconfiguration, which, in turn, serves as a dynamic capability to generate firm performance. The findings of the study indicate that there are direct relationships between BA and RR, and between RR and FP. The indirect relationship between BA and FP through the mediating role of RR is also confirmed.

The direct relationship between BA and RR suggests that BA applications and technologies can be used to support the effort to build up firms' resource reconfiguration ability. This connection helps to extend our knowledge of resource reconfiguration' antecedents and the benefits of business analytics. Moreover, the direct relationship between RR and FP confirms the business value of building dynamic capability. In this case, static resources, such as business analytics, are utilized to build up a dynamic capability, such as reconfiguration. This capability then helps firms to proactively realign, reorganize, and reconfigure their resource base, which allow them to generate business value.

The findings of the study have several managerial implications. The results can help firms gain a better understanding of business analytics' benefits. Investments on BA applications and technologies may speed up reconfiguration processes, which support firms staying competitive.

Limitations

While this study provides support for the positive relationships between BA and RR, between RR and FP, and RR's role in meditating the relationship between BA and FP, care should be taken in interpreting the findings. One of the limitations of this study is that it was validated using U.S. firms. It is interesting to see how the findings would be different in other countries or culture. With different cultural or business environments, the results from the model would be different. Second, we tested our research model using cross-sectional data.

Future research should validate these findings using longitudinal data to confirm the stability. Finally, future research can extend this research model by examining the moderating roles of organizational culture and environmental factors.

REFERENCES

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin, 103(3), 411.

Anderson, N., Potocnik, K. and Zhou, J. (2014), "Innovation and creativity in organizations: a state-of-the-science review, prospective commentary, and guiding framework", Journal of Management, Vol. 40 No. 5, pp. 1297-1333.

Ambulkar, S., Blackhurst, J., & Grawe, S. (2015). Firm's resilience to supply chain disruptions: Scale development and empirical examination. Journal of Operations Management, 33–34, 111–122. https://doi.org/10.1016/j.jom.2014.11.002

Araz, O. M., Choi, T. M., Olson, D., & Salman, F. S. (2020). Data analytics for operational risk management. Decision Sciences. Doi:10.1111/deci.12443.

Aydiner, A. S., Tatoglu, E., Bayraktar, E., Zaim, S., & Delen, D. (2019). Business analytics and firm performance: The mediating role of business process performance. Journal of Business Research, 96(October 2018), 228–237. https://doi.org/10.1016/j.jbusres.2018.11.028

Cao, Guangming; Duan, Yanqing; and Tian, Nan, "Improving Strategic Decision-Making through the Use of Business Analytics: A Resource-Based View" (2019). CONF-IRM 2019 Proceedings. 4.

Chakrabarti, A., Vidal, E., & Mitchell, W. (2011). Business transformation in heterogeneous environments: The impact of market development and firm strength on retrenchment and growth reconfiguration. Global Strategy Journal, 1(1–2), 6–26. https://doi.org/10.1002/gsj.3

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? Information & Management, 57(1), 103141.

Davenport, T. (2014). Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press.

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics. Boston: Harvard Business School Press, 46.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal structure, market dynamism, and the strategy of simple rules. Administrative Science Quarterly, 54(3), 413-452.

Davis, P. E., & Bendickson, J. S. (2020). Strategic antecedents of innovation: Variance between small and large firms. Journal of Small Business Management, 1-26.

Duan, Y., Cao, G., & Edwards, J. S. (2020). Understanding the impact of business analytics on innovation. European Journal of Operational Research, 281(3), 673–686. https://doi.org/10.1016/j.ejor.2018.06.021

Dubey, R., Gunasekaran, A., Childe, S. J., Fosso Wamba, S., Roubaud, D., & Foropon, C. (2019). Empirical investigation of data analytics capability and organizational flexibility as complements to supply chain resilience. International Journal of Production Research. https://doi.org/10.1080/00207543.2019.1582820

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? Strategic management journal, 21(10‐11), 1105-1121.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research, 18(1), 39-50.

Gligor, D. M., Holcomb, M. C., & Stank, T. P. (2013). A multidisciplinary approach to supply chain agility: Conceptualization and scale development. Journal of Business Logistics, 34(2), 94–108. https://doi.org/10.1111/jbl.12012

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. California management review, 33(3), 114-135.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010).Multivariate Data Analysis: A Global Perspective, 7th Edition. New Jersey: Pearson Educational Inc.

Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M., Singh, H., Teece, D., & Winter, S. G. (2009). Dynamic capabilities: Understanding strategic change in organizations. John Wiley & Sons.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision Support Systems, 64, 130-141.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal, 6(1), 1-55.

Ivanov, D., & Dolgui, A. (2020). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. Production Planning & Control, 0(0), 1–14. https://doi.org/10.1080/09537287.2020.1768450

Knemeyer, a. M., & Walker Naylor, R. (2011). Using Behavioral Experiments to Expand Our Horizon and Deepen Our Understanding of Logistics and Supply Chain Decision Making. Journal of Business Logistics, 32(4), 296–302. https://doi.org/10.1111/j.0000-0000.2011.01025.x

Lee, S. M., Rha, J. S., Choi, D., & Noh, Y. (2013). Pressures affecting green supply chain performance. Management Decision, 51(8), 1753–1768. https://doi.org/10.1108/MD-12-2012-0841

Lim, J.-S., Pham, P., & Heinrichs, J. H. (2020). Impact of social media activity outcomes on brand equity. Journal of Product and Brand Management. https://doi.org/10.1108/JPBM-03-2019-2298

Marsh, S. J., & Stock, G. N. (2006). Creating dynamic capability: The role of intertemporal integration, knowledge retention, and interpretation. Journal of Product Innovation Management, 23(5), 422-436.

Peng, D. X., Heim, G. R., & Mallick, D. N. (2014). Collaborative product development: The effect of project complexity on the use of information technology tools and new product development practices. Production and Operations Management, 23(8), 1421-1438.

Robinson, A., Levis, J., & Bennett, G. (2010). INFORMS to officially join analytics movement. OR/MS Today, 37(5), 59.

Sanchez, R. and Mahoney, J.T. (1996), “Modularity, flexibility, and knowledge management in product and organization design”, Strategic Management Journal, Vol. 17, pp. 63-76.

Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. Strategic management journal, 35(2), 179-203.

Schumpeter, J. (1934). The theory of economic development Harvard University Press. Cambridge, MA.

Sharda, R., Delen, D., & Turban, E. (2016). Business intelligence, analytics, and data science: a managerial perspective. Pearson.

Souza, G. C. (2014). Supply chain analytics. Business Horizons, 57(5), 595-605.

Srinivasan, R., & Swink, M. (2018). An Investigation of Visibility and Flexibility as Complements to Supply Chain Analytics: An Organizational Information Processing Theory Perspective. Production and Operations Management, 27(10), 1849–1867. https://doi.org/10.1111/poms.12746

Taleizadeh, A. A., Noori-daryan, M., & Cárdenas-Barrón, L. E. (2015). Joint optimization of price, replenishment frequency, replenishment cycle and production rate in vendor managed inventory system with deteriorating items. International Journal of Production Economics, 159, 285-295.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic management journal, 18(7), 509-533.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj

Thornton, L. D. M., Esper, T. L., & Autry, C. W. (2016). Leader or Lobbyist? How Organizational Politics and Top Supply Chain Manager Political Skill Impacts Supply Chain Orientation and Internal Integration. Journal of Supply Chain Management, 52(4), 42–62. https://doi.org/10.1111/jscm.12119

Torres, R., Sidorova, A., & Jones, M. C. (2018). Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective. Information and Management, 55(7), 822–839. https://doi.org/10.1016/j.im.2018.03.010

Tracey, M., Lim, J. S., & Vonderembse, M. A. (2005). The impact of supply‐chain management capabilities on business performance. Supply Chain Management: An International Journal.

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. Decision Support Systems, 49(3), 318–327. https://doi.org/10.1016/j.dss.2010.03.007

Tuarob, S., & Tucker, C. S. (2013, August). Fad or here to stay: Predicting product market adoption and longevity using large scale, social media data. In ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers Digital Collection.

Vanpoucke, E., Vereecke, A., & Wetzels, M. (2014). Developing supplier integration capabilities for sustainable competitive advantage: A dynamic capabilities approach. Journal of Operations Management, 32(7–8), 446–461.

Vidal, E., & Mitchell, W. (2015). Adding by subtracting: The relationship between performance feedback and resource reconfiguration through divestitures. Organization Science, 26(4), 1101-1118.

Tracey, M., Vonderembse, M. A., & Lim, J. S. (1999). Manufacturing technology and strategy formulation: keys to enhancing competitiveness and improving performance. Journal of operations management, 17(4), 411-428.

Wamba, S. F., Ngai, E. W. T., Riggins, F., & Akter, S. (2017). Transforming operations and production management using big data and business analytics: future research directions. International Journal of Operations & Production Management, 37(1), 2–9. https://doi.org/10.1108/IJOPM-07-2016-0414

Wang, S., Yeoh, W., Richards, G., Wong, S. F., & Chang, Y. (2019). Harnessing business analytics value through organizational absorptive capacity. Information and Management, 56(7). https://doi.org/10.1016/j.im.2019.02.007

DECISION SCIENCES INSTITUTE

ROLE OF INFORMATION MANAGEMENT SYSTEM IN LABORATORY AND PETROCHEMICAL QUALITY ASSURANCE

Shah Nazir
Department of Computer Science, University of Swabi, Swabi, Pakistan
Email: snshahnzr@gmail.com / shahnazir@uoswabi.edu.pk

Raiha Arshad
College of Engineering, Qatar University, Doha, Qatar
Email: ra1800228@student.qu.edu.qa

Habib Ullah Khan* (Corresponding Author)
College of Business & Economics, Qatar University, Doha, Qatar
Email: habib.khan@qu.edu.qa

ABSTRACT

Since its introduction in 1999, many laboratories have established their quality management systems according to ISO/IEC 17025:1999 standard and obtained the accreditation to it by accreditation bodies as competent laboratories. Accreditation to ISO/IEC 17025:2005 is an international recognition of laboratory competence. However, this standard is very comprehensive. Clause 4 Management requirements & Clause 5 Technical requirements of this international standard cover all quality management activities in the laboratory. On the other hand, petroleum and petrochemical laboratories use computer mediated software such as information management system "LIMS." Yet, it does not meet all the requirements set forth by ISO/IEC 17025:2005. As LIMS local administrator I found this application an excellent tool for laboratory technical activities such as samples scheduling, tracking and reporting, but it lacks some management capability such customer complaint, subcontracting, preventive action, corrective actions, documentation of quality control activities, taskforce training and competency records.As well, LIMS providers may not follow the standard guide for LIMS ASTM E1578-06 the technical guidelines for developing LIMS.At this point, it is obvious that LIMS software approaches quality different from ISO/IEC 17025:2005 standard and accreditation bodies' requirements. Consequently, this invariably revealed major deficiency with current LIMS software that lacks many aspects required by ISO 17025:2005 accreditation bodies. Thus, these deficiencies made laboratories run another documentation system beside the LIMS in order to have comprehensive QA/QC system and satisfy ISO/IEC 17025:2005 standard requirements. Moreover, LIMS software comes in separated modules, each provides a specific task(s), and it is up to each laboratory to select the module(s) that suite their needs. LIMS providers focus on the wider beneficiary range, "the laboratories' customers," such as the plants processing units & entity management rather than laboratory internal activity, for example testing equipment and associated taskforces.This focus may enforced by the assumption that laboratories have their own QA/QC in place, freeing LIMS developers to support & appeal the wider range of customers with the timely professional solutions such as the state of the art reports.

KEYWORDS. Information System Management, Quality Assurance, Petrochemical, Technology Adoption.

INTRODUCTION

The aim of the project is to highlight the key requirements for laboratory computer mediated software, LIMS, to meet the ISO/IEC 17025:2005 requirement and become an effective quality management system.The project examined LIMS features and capabilities with respect to petroleum & petrochemical laboratories QA/QC requirements and the benefits of incorporating ISO/IEC 17025:2005 elements into LIMS. Furthermore, the project answered the following questions: What should LIMS contain? How laboratory setup, manages, and monitor? What should QA/QC contain? How laboratory establish, implement, & control? How laboratory ensures the quality of its internal devices & taskforces? How laboratory assists the quality improvement of its external customers?

It is true that everyone is seeking for quality. However, the quality in one perception today may not be the same as yesterday. In that case, what is quality? Does quality have an absolute or relative definition? It was surprising to researchers when they have started researching for the meaning of quality; they found that defining quality is not an easy task, maybe due to the relative relationship between quality level in ourselves and its importance at that time. For example, if one needs to see what the time now is, a traditional clock has enough quality here, but if one needs to measure a distance with a laser beam, a precise scientific clock will be needed, and the traditional one will not have a good or enough quality here.

Research team examines some others descriptions, ISO/IEC 9000 (2005, p. 7) the international standard defines quality as 'level to which a set of inherent attributes meet requirements'. At this point, they should notice that the word inherent indicating an originality in elements. Furthermore, Notes 1 & 2 on the same page illustrate that the word attributes or 'characteristics' encompasses quantitative as well qualitative issues such as physical, sensory, behavioural, temporal, ergonomic or functional. Therefore, quality here means the elements have unique features in itself that precisely fit and right to the use.

What's more, BusinessDictionary.com (2011) defines quality in manufacturing as a measure of excellence indicating that the items are free from defects, deficiencies, or significant variations in order to meet end-user satisfaction and / or need.In addition, the Free On-line Dictionary of Computing defines quality as the entire aspects and attributes of a product or service that bear on its ability to satisfy user needs implied or stated. Here, researchers should notice that the word entire or 'totality' does not meet the words "degree of excellence" or "fitness for use" as these indicates only part of the definition. (Dictionary.com, 1995).

The Global Voice of Quality 'ASQ' excerpted from (Benbow et al., 2003, pp. 1 -2) that users' satisfaction may not be clear; thus way, many organizations spend a lot of efforts, time and expenses to grasp the voice of targeted customers. At this point, customers' satisfaction includes providing what is needed when it is needed. (ASQ, n.d.). In addition, ASQ has listed many basic concepts of quality such as supplier quality, continuous advancing, quality price, customers support for problem solving, activities monitoring for a practice, and deviation from targeted limits.

ISO/IEC 9000 (2005, p. 9) defines management as the "coordinated activities" in order to direct and control an organization with reference to quality. What's more, the note on the same page illustrates that quality managements includes quality policy, quality objectives, quality planning, quality assurance (QA), quality control (QC) and quality improvement.

At this point, QA and QC is what matters most to this project. ISO/IEC 9000 (2005, p. 9) restricts assurance of quality within QM as the activities that pay particular attention to provide confidence on fulfilling quality requirements. Here, confidence has dual viewpoints, first from the organization side that a business is able to provide consistent products, services, or process; and second from the customers' side, those are happy with what they get. Alternatively, control within QM concentrates on fulfilling what is required. At this instant, control indicates evaluation activities & practices to fulfill quality requirements or maintain assurance.

ISO Web (2011) stated that ISO enables an agreement to be reached on solutions that meet both the requirements of a business and the need of their customers or the society. Accordingly, the core business of ISO is assisting organizations to implement and to operate effective QMS in its broader concepts and especially the QA/QC by providing standards that are satisfactory to litigant parties, facilitating mutual understanding in national and international trade(ISO/IEC 9000, 2005, p. v). Therefore, this project will follow the definitions and principles provided by the following international standards ISO/IEC 9000:2005, ISO/IEC 17000:2004 and ISO/IEC 17025:2005.

Research still in progress.

"This publication was supported by Qatar university Internal Grant No. IRCC-2020-009. The findings achieved herein are solely the responsibility of the authors"

REFERENCES

Askoul, R., Khan, H.U. and Madhavi Lalitha, V.V. (2016) 'Cross-functional integration of marketing and information services in banking: a cross-industry comparison', International. Journal of Process Management and Benchmarking, Vol. 6. No. 1. pp. 57-78

Bankole, O. A., Lalitha, M., Khan, H.U., Jinugu, A. (2017), "Information Technology In The Maritime Industry Past, Present And Future: Focus On Lng Carriers", 7th IEEE International Advance Computing Conference, Hyderabad, India, January 5 – 7, 2017. (Conference Proceeding).

Bashir, G. M., Khan, H.U (2016b), "Factors Affecting Learning Capacity Of Information Technology Concepts In A Classroom Environment Of Adult Learner", 15th International Conference on Information Technology Based Higher Education and Training (IEEE Conference), Istanbul, Turkey, September 8th – September 10, 2016. (Conference Proceeding).

Brock, V. F. and Khan H.U. (2017), "Are Enterprises Ready For Big Data Analytics? A Survey Based Approach", Int. J. of Business Information Systems, Vol.25, No.2, pp.256:277.

Das, A. and Khan, H.U. (2016) "Security behaviors of smartphone users", Information and Computer Security, Information and Computer Security, Vol. 24, No.1, pp. 116-134

Ejike, A. C. , Khan, H.U., Fournier-Bonilla, S. D. (2016), "Possible Impact of Mobile Banking on Traditional Banking: A Case Study of Nigeria", Northeast Decision Sciences Institute Conference, Alexandria, Virginia, USA, March 31st – April 2nd, 2016. (Conference Proceeding).

Erfan Najmi, E., Hashmi, K., Malik, Z., Rezgui, A., Khan, H.U. (2015), "CAPRA: a comprehensive approach to product ranking using customer reviews", Computing, DOI 10.1007/s00607-015-0439-8.

Halabi A. E., Hachem A., Al-Akhrass L., Artail H., Khan H.U. (2014), "Identifying the linkability between Web servers for Enhanced Internet Computing", 17th IEEE Mediterranean Electrotechnical Conference MELECON 2014, 13-16 April, Beirut, Lebanon. (Conference Proceeding)

Hassan, I. M., Khan, H. U., and Lalitha, M. (2016), 'Pedagogical Potentials of IEEE 802.11 WLAN to Nigerian Universities: A Case Study of the University of Uyo', International Journal of Information and Education Technology, vol. 6, no. 4, pp. 256-261

Hassan, I. M., Khan, H. U., Zaitun, R., Mardini, G. (2015), "Pedagogical Potentials of IEEE 802.11 WLAN to Higher Educational Institutions: A Case Study of Nigerian based University", IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Anhaim, CA, USA, Feb.7 - 9, 2015. (Conference Proceeding).

Heang, J.F., and Khan, H.U. (2015), "The Role of Internet Marketing in the Development of Agricultural Industry: A Case Study of China",Vol. 14, Issue.1 , pp. 1-49.

Ho, H. C., Awan, M. A., & Khan, H. U. (2016), "Luxury brands and corporate responsibility: A perspective on consumers' preferences", Journal of International Management Studies, 16(1), 77-81.

Khan H.U., Awan M.A.(2017), "Possible Factors Affecting Internet Addiction: A Case Study of Higher Education Students of Qatar", Int. J. of Business Information Systems (IJBIS), Vol.26, No.2 pp.261:276.
Khan, H.U. (2012) 'Computer Mediated Communication, Quality of Learning, and Performance', Journal of GSTF Business Review, Vol. 1, No. 3, pp. 81-88.

Khan, H.U. (2013a) 'Use of e-learning tools to solve group work problems in higher education: A Case study of gulf countries', The Advances in Computer Science: an International Journal, Vol. 2, No. 3, pp.90-96.

Khan, H.U. (2013b) 'Role of Computer Mediated Communication in Affect Empowerment and Performance Improvement', IFRSA's International Journal of Computing, Vol.3, No.3, P.165-171.

Khan, H.U. (2016), "Possible effect of video lecture capture technology on the cognitive Empowerment of higher education students: a case study of gulf-based university", International Journal of Innovation and Learning, Vol.20, No. 1, pp. 68: 84.

Khan, H.U. and Adediji, O.A. (2017) 'Need for RADAR system utilisation for maritime traffic management: a case of Congo River Basin', Int. J. Computational Systems Engineering, forthcoming.

Khan, H.U. and Alhusseini, A. (2015), "Optimized Web Design in the Saudi Culture", IEEE Science and Information Conference 2015, London, UK, July 28 - 30, 2015. pp.906-915 [Co-sponsored by Springer]

Khan, H.U. and Ejike, A.C. (2017) "An assessment of the impact of mobile banking on traditional banking in Nigeria", Int. J. Business Excellence, Vol.11, No.4, pp.446:463.

Khan, H.U. and Fournier-Bonilla, S. D. (2016), "Technological Infrastructure Effects on Export Diversification: A Case Study of Qatar", Northeast Decision Sciences Institute Conference, Alexandria, Virginia, USA, March 31st – April 2nd, 2016. (Conference Proceeding).

Khan, H.U., Ahmed, S., Abdollahian, M. (2013) 'Supply chain technology acceptance, adoption, and possible challenges: A case study of service organizations of Saudi Arabia', 10th International Conference on Information Technology: New Generations (ITNG 2013), Las Vegas, Nevada, USA.

Khan, H.U., Artail, H., Malik, Z., Niazi, M.(2014a) 'Information Technology Adoption, possible challenges, and Framework of Supply Chain Management: A Case Study of a Leading Gulf Economy', 4th International Conference on International Conference on International Conference on Engineering Technology and Technopreneurship, Kuala Lumpur, Malaysia.

Khan, H.U., Awan, M.A., Ho. H.C. (2014), "How do Chinese and Saudi Customers Perceive Online Service Quality? A Comparative Study", The Journal of Business Inquiry, Vol. 13, No. 2.

Khan, H.U., Awan, M.A., Ho. H.C. (2014b), "How do Chinese and Saudi Customers Perceive Online Service Quality? A Comparative Study", The Journal of Business Inquiry, Vol. 13, No. 2.pp.142-157.

Khan, H.U., Bankole, O.A. and Alomari, M.K. (2017b) 'Possible effect of IT introduction into the election process: a case study of Nigeria', Int. J. Business Forecasting and Marketing Intelligence, Vol. 3, No. 2, pp.109-129.

Khan, H.U., Omonaiye, J.F., and Madhavi Lalitha, V.V. (2017a) "Employees' perception as internal customers about online services: A case study of banking sector in Nigeria', International Journal of Business Innovation and Research, Vol.13, No.2, pp.181:202.

Khan, H.U., Uwemi, S. (2018a) "Possible Impact Of E-Commerce Strategies On The Utilization Of E-Commerce In Nigeria', International Journal of Business Innovation and Research, vol. 15, No. 2., Pp. 231-246.
35.

Khan, H.U., Uwemi, S. (2018b) “What are e-commerce possible challenges in developing countries: a case study of Nigeria”, International Journal of Business and Systems Research, Vol. 12, No. 4, pp: 454-486.

Khan H.U. (2020), “The Role Of Smac (Social Media, Mobility, Analytics, Cloud) For Students And Educators In Online Education”, Journal of Theoretical and Applied Information Technology, Vol. 98, No. 6, pp: 915-934.

Musa A., Khan, H.U., Alshare, K. (2015), “Factors influence consumers’ adoption of mobile payment devices in Qatar”, International Journal of Mobile Communications, Vol. 13, No. 6, pp. 670-689..

Najmi E., Hashmi K., Malik Z., Rezgui, A., Khan H.U.(2014), “ConceptOnto: An upper ontology based on Conceptnet”, 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA' 2014), November 10-13, 2014, Doha, Qatar ,PP.366-372 (Conference Proceeding)

Najmi, E., Hashmi, K., Malik, Z., Rezgui, A., Khan, H.U. (2015), “CAPRA: a comprehensive approach to product ranking using customer reviews”, Computing, Vol. 97, No. 8, pp: 843-867.,

O. Bankole and E. Cloete, “Mobile Banking: A Comparative Study of South Africa and Nigeria”, IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia, 13 – 15, September 2011.

Omonaiye, J.F., Madhavi Lalitha., Khan, H.U., Signh, R., Fournier-Bonilla, S. D. (2015), “Ability and hurdle to provide Banking online services: A case study of banking employees in Nigeria”, 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing, New York,USA, November 03 - 05, 2015. (Conference Proceeding)

Saied, D.E.E.L., Khan, H.U. (2017) ‘Implementation of Health Information System-A case study of Magrabi hospitals, KSA’, Journal of Computer Science, Vol. 13. No. 5., pp. 91-104.

Smuts, R.G., Lalitha, M., Khan, H.U. (2017), “Change Management Guidelines That Address Barriers To Technology Adoption In An HEI Context”, 7th IEEE International Advance Computing Conference, Hyderabad, India, January 5 – 7, 2017. (Conference Proceeding).

Uwemi, S., Khan, H.U. (2016), “E-commerce, Challenges, and Developing Countries”, 2016 DSI Annual Meeting in Austin, TX, USA. November 19th – November 22nd, 2016. (Conference Proceeding).

Uwemi, S., Khan, H.U., Fournier-Bonilla, S. D. (2016), “Challenges of E-Commerce in Developing Countries: Nigeria As Case Study”, Northeast Decision Sciences Institute Conference, Alexandria, Virginia, USA, March 31st – April 2nd, 2016. (Conference Proceeding).

Authors Index