INVESTIGATION OF SURVIVAL MODELING IN SUPPLY CHAIN FINANCIAL DISTRESS ANALYSIS

Scott Dellana, East Carolina University, Greenville, NC 27858, dellanas@ecu.edu, (252) 328-4893
David West, East Carolina University, Greenville, NC 27858, westd@ecu.edu, (252) 328-6370
Gary Gammon, East Carolina University, Greenville, NC 27858 gammong10@students.ecu.edu
(704) 689-1571

ABSTRACT

Supply chain risk management has received considerable attention in recent years. Financial models that can help predict the probability and timing of supply chain partner failure can bring great benefit to mitigating risks associated with economic instability and financial distress. In this study we compare the predictive bankruptcy classification accuracy of the Cox Proportional Hazards Regression (CPHR) approach to the statistical approaches of Logistic Regression and Discriminant Analysis. Analysis of CPHR versus other models shows competitive results worthy of further consideration.

Keywords: Bankruptcy, Financial Distress, Risk, Supply.

INTRODUCTION

As supply chains continue to stretch further afield in search of the most efficient and cost effective arrangements, the risks to the supply chain increase. The widespread pressures to adopt lean systems and their reliance on customer-driven manufacturing and delivery in conjunction with the disruptive threats of global terrorism, natural disasters, and economic instability increase the risks of supply chain failures and even collapse.

An example is found in the Land Rover UPF-Thompson case of 2001 (Waters, 2011). Land Rover learned of UPF's bankruptcy only after the sole supplier's chassis frame deliveries stopped. Loss of the UPF supply of frames threatened over 10,000 jobs at Land Rover and its other suppliers. In the end, Land Rover had to agree to take on some of its supplier's debt in order to continue operations. In this case, it is clear that Land Rover was caught off guard and would have benefitted from proactive prediction of this critical supplier failure. While many supply chain risks are not easily mitigated, it is possible to improve on the ability to monitor and predict financial distress of supply chain partners, so that actions may be taken in a timely manner to manage the risks. However, many modeling approaches typically take only a snap shot of financial data at one point in time and make a prediction of failure probability with little information about time frames involved.

The purpose of this research is to investigate whether or not a dynamic survival classification method is able to predict potential bankruptcies more accurately than the static classification methods. An advantage of survival analysis is that it can give a time frame to potential failure.

We study the Cox Proportional Hazards Regression or CPHR model and contrast its decision accuracy with other prominent static classification methods.

LITERATURE REVIEW

Bankruptcy prediction has been an area of interest for many decades. Altman (1968) introduced effective bankruptcy prediction modeling in his seminal work on discriminant analysis using financial ratios. Traditional methods of financial decision support include scorecards for consumer credit (Brill, 1998; Henley, 1995; Mester, 1997; Reichert et al, 1983; Rosenberg & Gleit, 1994) and discriminant models for assessing corporate financial health (Altman et al, 1994; Reichert et al, 1983).

Newer models have emerged in bankruptcy classification modeling, which include neural networks and survival analysis. The reader is referred to Smith and Gupta (2002) and Kumar and Ravi (2007) for a survey of the application of neural networks in a diverse range of research problems that include financial forecasting. In many of the cases the neural networks are reported to provide more accurate bankruptcy prediction capability than the traditional parametric statistical approaches. However, the results are also often mixed (Tam & Kiang, 1992; Coats & Fant, 1993; Altman et al, 1994; Kiviluoto, 1998)

In 1996, Dimitras, et al conducted a comprehensive survey of articles related to business failure and reported only one journal article at that time dealing with survival analysis and bankruptcy prediction. They concluded that survival analysis, although a viable technique, has not been often applied for the prediction of business failure. Balcaen and Ooghe (2006) recently reported on 35 years of studies in business failure. They omitted discussion altogether of survival analysis, stating that it was beyond the scope of their study and needed further research. Thus, very little attention has been paid in the literature to survival analysis as it relates to bankruptcy prediction.

Luoma and Laitinen (1991) studied survival analysis as a tool for predicting company failure. Using a small sample of diverse Finnish failed and non-failed business firms (36 total), they compared their results to those of Lane et al (1986), for discriminant analysis (DA) and logistic regression models. Survival analysis classified both failed firms and non-failed firms correctly 62% of the time. DA classified failed firms correctly 65% of time, while non-failed firms were classified correctly 77% of the time. Logit analysis classified failed firms correctly 74% of time, while non-failed firms were classified correctly 71% of the time. Although survival analysis was not found to outperform DA and Logit, the authors point out that it modeled the dynamic aspects, treated firms as the same population, used more information, and better represented the nature of the failure process. Honjo (2000) examined business failure specifically in new firms using a multiplicative hazards model. The study examined business failure among 2488 new manufacturing firms in Tokyo during the period 1986 to 1994. The hazards regression model was based on age and also on calendar time. By using a regression model based on calendar time, it was found that the firm's age is related to business failure. No other models were used for comparison in this study. Cochran, Darrat and Elkhal (2006) report on the bankruptcy of internet companies. Using a calendar-time model, they identified three key predictors of company failure; net income to total assets, cash flow to total liabilities, and total assets. In addition, they used an event-time model and found that liquidity became more important as a predictor than profit potential about one year prior to the failure, but that this finding is reversed (i.e. liquidity is less important than profit potential) three years prior to bankruptcy. Their results suggest that for three years prior to bankruptcy, a higher ratio of total liabilities to total assets is associated with lower odds of survival. No other models were compared to CPHR in this study.

The literature on survival analysis is fairly limited and reports apparently little success in its application to business failure modeling. However, we believe there is room for improvement and a place for survival analysis in supply chain risk mitigation. Studies in the past have suffered from the use of only static models, with no inclusion of dynamic or time series effects. Metrics employed appear to be limited and experimental methodology sometimes poorly defined. Our study will employ a rigorous methodology of model development, cross-validation, and robust metrics. Our use of quarterly data spanning two years for each company in our dataset will allow the inclusion of dynamic time series effects. Finally, we will employ a company data set large enough to gain the power needed for analysis of a variety of financial and related variables in the predictive model.

DATA AND MODELS

Description of the Dataset

A bankruptcy dataset was constructed by the authors from Standard and Poor's Compustat financial files. Explanatory variables for this part of the study were limited to the five key financial ratios from research by Altman et al (1994). These ratios, constructed from quarterly financial statement information spanning approximately a decade from 1996 to 2005, include: working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value of equity/book value total liability, and sales/total assets. We also included information on company age and a monthly index of US economic health in the PMI index published by the Institute for Supply Management. The dataset has a total of 146 companies, consisting of 75 bankrupt companies and 71 healthy companies. The original data set was larger with an equal number of bankrupt and non-bankrupt companies. However, outlier values were detected in some of the original data and a number of bankrupt and non-bankrupt companies were dropped from the study. The original non-bankrupt company data were matched as closely in time as possible to the bankrupt companies, so that the two groups are close to equally represented by time period. The financial information is available as quarterly data for each company. Ratios were obtained going back eight quarters starting one year prior to bankruptcy for the bankrupt companies. The same ratios were obtained for the healthy companies going back eight quarters during a similar time frame as for matched bankrupt companies.

Discriminant Analysis

Multiple Discriminant Analysis can be used to classify an observation into one of several *a priori* groupings. In this research we limit this analysis to the two binary classifiers of bankrupt and non-bankrupt, thus employing simple discriminant analysis or DA. This classification is based on the individual characteristics of the observation. DA attempts to derive a linear combination of the characteristics which best discriminates among groups (Altman, 1968). In our study, the groupings are bankrupt and non-bankrupt.

The discriminate function of the form,

$$Z = v_1 x_1 + v_2 x_2 + \dots + v_n x_n \tag{1}$$

where $v_1, v_2, \dots v_n$= Discriminant coefficients
and $x_1, x_2, \dots x_n$ = Independents variables

transforms individual variable values to a single discriminant score of $\boldsymbol{Z}$ value which is then used to classify the object. The DA computes the discriminate coefficients, $\boldsymbol{v_j}$, while the independent variables $\boldsymbol{x_j}$ are the actual values, where, $j = 1, 2,\dots \boldsymbol{n}$**.** DA allows the analysis of the entire variable profile of the object simultaneously rather than sequentially examining its individual characteristics. In a financial health application, this allows combinations of financial ratios to be considered together (Altman, 1968).

Logistic Regression (Logit)

Logit is a statistical model based on a discrete probability distribution which takes the value 1 with success probability p and the value 0 with failure probability $q = 1 - p$. It is a generalized linear model that uses the logit as its link function. The model takes the following form (Agresti, 2002):

$$Logit\ (p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \tag{2}$$

$i = 1, \dots, n$, where there are n units with covariates $\boldsymbol{X}$ and

$$p_i = E(Y \mid X_i) = \Pr(Y_i = 1) \tag{3}$$

The logarithm of the odds (the probability divided by one minus the probability) of the outcome is modeled as a linear function of the explanatory variable, $\boldsymbol{X_i}$**.** This can be written equivalently as

$$p_i = \Pr(Y_i = 1 \mid X) = \frac{e^{\alpha+\beta_1 x_{1,i}+\dots+\beta_k x_{k,i}}}{1+e^{\alpha+\beta_1 x_{1,i}+\dots+\beta_k x_{k,i}}} \tag{4}$$

The interpretation of the $\boldsymbol{\beta}$ parameter estimates is as a multiplicative effect on the odds ratio. In the case of a dichotomous explanatory variable, for instance bankrupt versus non-bankrupt, $\boldsymbol{e^{\beta}}$is the estimate of the odds-ratio of having the outcome for bankrupt compared with non-bankrupt. The parameters $\boldsymbol{\alpha, \beta_1, \dots, \beta_k}$ are usually estimated by maximum likelihood.

Cox Proportional Hazards Regression Model (CPHR)

Survival analysis models the time it takes for events to occur, focusing on the distribution of survival times. While DA and Logit can be used to classify firms according to financial health, the CPHR model allows estimation of a firm's probability of failure. DA and Logit are also limited to static or cross-sectional data, while CPHR can be used to study time-varying data as well. The CPHR approach enjoys fewer basic assumptions and can be performed without having to define the baseline hazard function. A major distinction of CPHR is its ability to include censored data in the model. These data represent firms that might not fail during the sample period. Thus, companies that remain financially healthy during the sample period can be included and modeled in the regression.

Survival analysis, as applied to analysis of organizational financial health, allows the estimation of time to business firm failure (i.e., bankruptcy). The survival is measured by the hazard rate using a hazard function (Zhang & Lyn, 2012). The hazard function $h(t)$, giving the intensity to fail, is defined as the limit of the conditional probability

$$h(t) = \lim_{\Delta t \to 0+} \left[\frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t} \right] \tag{5}$$

where T is the random variable describing the moment of failure for a firm and $P(t \le T \le t + \Delta t | T \ge t)$ the survival function (Cochran et al, 2006). This expression gives the instantaneous conditional probability of failure provided that the firm's survival is longer than time t. This can be estimated by

$$h(t, X_i) = h_0(t) \exp \left[\beta^T X_i \right] \tag{6}$$

where $h_0(t)$ is referred to as the baseline hazard function, which is a nonparametric function of time alone and is assumed to be the same for all firms (Noh et al, 2005). The *exp* part of the equation represents the parametric portion of the Cox model. The term X denotes a vector of p explanatory variables. The vector X contains individual firm characteristics and helps predict the survival of the firm in the index.

The CPHR model is semi-parametric (Simonoff & Ma, 2003) since the hazard function is not pre-specified but allows for the incorporation of any chosen survival function. According to Efron (1977) the effect of the baseline hazard function loss is negligible in almost all instances. While the baseline hazard can take any form, the covariates enter the model linearly. For example, consider two observations i and i' that differ in the x-values, with corresponding linear predictors

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{ik} + \cdot \cdot \cdot + \beta_k x_{ik} \tag{7}$$

and

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'k} + \cdot \cdot \cdot + \beta_k x_{i'k} \tag{8}$$

The hazard ratio for these two observations

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_0(t)e^{n_i}}{h_0(t)e^{n_{i'}}} = \frac{e^{n_i}}{e^{n_{i'}}} \tag{9}$$

is independent of time t. Consequently, the Cox model is a proportional-hazards model (Fox & Weisberg, 2011). Surprisingly, even though the baseline hazard is unspecified, after fitting the model, it is then possible to extract an estimate of the baseline hazard.

EXPERIMENTAL DESIGN

The experimental design for the full study involves the analysis and comparison of the three different models of DA, Logit and CPHR from the perspective of data that are static, static with time trend, and dynamic (i.e., capturing a time series effect) as appropriate. In addition, the models will be run for cases of 4, 6 and 8 quarters prior to bankruptcy. However, the preliminary investigation presented in the current paper is limited to static analysis only for all three models with only 4 quarters prior to bankruptcy. If CPHR is competitive to DA and Logit under this basic analysis, then it should provide better predictive power with trend and dynamic modeling included. In addition, interaction of model terms was not included in this study -- only main effects were considered.

Appropriate statistical tests were conducted to assess compliance with underlying parametric model assumptions, as appropriate. A cross-validation "leave-one-out" approach was used to assess the errors in each model. This means that the model was run 146 times, with each model building run leaving out one of the companies from the dataset. This one company was then used to test the model's ability to predict its outcome as either bankrupt or non-bankrupt. The percentage of correct outcomes was then tallied in the case of both bankrupt and non-bankrupt companies.

RESULTS AND ANALYSIS

All models were run in SAS 9.2 statistical software with model parameters for individual runs exported in excel files. These files were then consolidated into a single excel file for construction of the predictive models from each cross-validation run. Each model was then tested using the corresponding hold-out company data to compare the model prediction to the actual outcome for each hold-out company. Summary results of these hold-out tests are shown in Table 1 for the three modeling approaches.

Table 1 shows each model's predictive ability as the percentage of bankrupt and non-bankrupt companies correctly classified. All three models did very well in general. Discriminant analysis was the best of the three models at correctly classifying companies that went bankrupt during the sample period. At 93.24% correctly classified bankrupt firms, it was slightly better than Logit at 91.9% , which was slightly better than CPHR at 90.7%. For classification of non-bankrupt

TABLE 1 – RESULTS OF PRELIMINARY STATIC CLASSIFICATION RUNS

Model	Correct % Bankrupt	Correct % Non-bankrupt	Overall Average % Correct
Discriminant Analysis	93.24%	86.96%	90.21%
Logistic Regression	91.90%	92.80%	92.30%
Cox Proportional Hazards Regression	90.70%	97.00%	93.60%

companies, CPHR was the best of the three models. At 97% correctly classified non-bankrupt firms, it was notably better than Logit at 92.8%, and substantially better than DA at about 87%. Overall, CPHR was the better predictor of company outcome during the sample period, classifying correctly on average 93.6% of the time. This was slightly better than Logit and considerably better than DA.

DISCUSSION AND CONCLUSIONS

We have shown, using cross-sectional data only, that Cox Proportional Hazards Regression modeling performs better overall than both Logistic Regression and Discriminant Analysis in the outcome prediction related to company bankruptcy. CPHR was able to correctly classify company financial state almost 94% of the time as bankrupt or non-bankrupt based on data one year ahead of potential bankruptcy. This finding suggests that the CPHR model is well-suited to predictions regarding future financial health of a company and may be useful in providing a time-frame for potential company failure. Thus, survival analysis using CPHR may provide substantial benefit to supply chain members in the assessment of financial risk within the supply chain. This may allow supply chain partners to take early action to minimize costly disruptions to their operations and possibly even avoid being critically damaged by the unexpected loss of a supply chain partner.

As mentioned previously, the experimental design for the full study will add a comparison of the models for data that are static with time trend, and also dynamic, as appropriate. In addition, runs will be conducted for 6 and 8 quarters prior to bankruptcy. This should extend the available lead time for successfully predicting potential financial instability in the supply chain. Interaction terms will also be considered in the full study. Future research, beyond this basic study, might focus on extending this work to include comparison to various neural network models. However, we do not expect these to perform significantly better given the data are not characterized as highly non-linear or dynamic.

REFERENCES

Agresti, I. (2002). *Categorical Data Analysis*, 2nd ed., New York: Wiley-Interscience.

Allison, P.D. (1984). *Event History Analysis. Regression for Longitudinal Event Data*. Sage Publications, Beverly Hills, CA.

Altman, E., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance* 18, 505-529.

Altman, E.I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23(4), 589-609.

Arsene, C. T. C., & Lisboa, P. J. G. (2007). Artificial neural networks used in the survival analysis of breast cancer patients: A node-negative study. In Azzam F.G. Taktak, & Anthony C. Fisher (Eds.), *Outcome prediction in cancer* (pp. 191-239). Amsterdam: Elsevier.

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. [Electronic version]. *The British Accounting Review, 38*(1), 63-93.

Beran, J., & Djaïdja, A. K. (2007). Credit risk modeling based on survival analysis with immunes. [Electronic version]. *Statistical Methodology, 4*(3), 251-276.

Biganzoli, E., Boracchi, P., & Marubini, E. (2002). A general framework for neural network models on censored survival data. [Electronic version]. *Neural Networks, 15*(2), 209-218.

Brill, J. (1998). The importance of credit scoring models in improving cash flow and collections. *Business Credit* 100, 16-17.

Coats P. & Fant, L. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management* 22, 142-155.

Cochran, J. J., Darrat, A. F., & Elkhal, K. (2006). On the bankruptcy of internet companies: An empirical inquiry. [Electronic version]. *Journal of Business Research, 59*(10-11), 1193-1200.

Cox, D. R. (1972) Regression models and life tables. *Journal of the Royal Statistical Society, Series B (Methodological)* , 34,187-220.

Cox, D. R. & Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London.

De Laurentiis, M., & Ravdin, P. M. (1994). A technique for using neural network analysis to perform survival analysis of censored data. [Electronic version]. *Cancer letters, 77*(2-3), 127-138.

Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. [Electronic version]. *European Journal of Operational Research, 90*(3), 487-513.

Effron, B. (Sep. 1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72, 557-565.

Eleuteri, A., Tagliaferri, R., Milano, L., De Placido, S., & De Laurentiis, M. (2003). A novel neural network-based survival analysis model. [Electronic version]. *Neural Networks, 16*(5-6), 855-864.

Fox, J. & Weisberg, S. (2011). *An Appendix to An R Companion to Applied Regression, 2nd Edition, Cox Proportional-Hazards Regression for Survival Data in R,* Sage Publications.

Geppert, J. M., Ivanov, S. I., & Karels, G. V. (2010). Analysis of the probability of deletion of S&P 500 companies: Survival analysis and neural networks approach. [Electronic version]. *The Quarterly Review of Economics and Finance, 50*(2), 191-201.

Henley, WE. (1995). *Statistical aspects of credit scoring*. Dissertation. The Open University, Milton Keynes, UK.

Honjo, Y. (2000). Business failure of new firms: An empirical analysis using a multiplicative hazards model. [Electronic version]. *International Journal of Industrial Organization, 18*(4), 557-574.

Kiviluoto, K. (1998). Predicting bankruptcies with the self-organizing map. *Neurocomputing* 21, 191-201.

Kumar, P.R. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180, 1-28.

Lisboa, P. J. G. (2002). A review of evidence of health benefit from artificial neural networks in medical intervention. [Electronic version]. *Neural Networks, 15*(1), 11-39.

Luoma, M., & Laitinen, E. (1991). Survival analysis as a tool for company failure prediction. [Electronic version]. *Omega, 19*(6), 673-678.

Mester, L.J. (1997). What's the point of credit scoring? *Business Review - Federal Reserve Bank of Philadelphia*, 3-16.

Noh, H. J., Roh, T. H., & Han, I. (2005). Prognostic personal credit risk model considering censored information. [Electronic version]. *Expert Systems with Applications, 28*(4), 753-762.

Ohno-Machado, L. (2001). Modeling medical prognosis: Survival analysis techniques. [Electronic version]. *Journal of Biomedical Informatics, 34*(6), 428-439.

Ohno-Machado, L. (1997). A comparison of cox proportional hazards and artificial neural network models for medical prognosis. [Electronic version]. *Computers in biology and medicine, 27*(1), 55-65.

Reichert, A.K., Cho, C.C. and Wagner, G.M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business and Economic Statistics* 1, 101-114.

Rosenberg, E. & Gleit, A. (1994). Quantitative methods in credit management: a survey. *Operations Research*, 42, 589-613.

Siegel, S. *Nonparametric Statistics for the Behavioral Sciences*. McGraw Hill Book Company, New York: 1956.

Simonoff & Ma (2003). An empirical study of factors relating to the success of broadway shows. *The Journal of Business*, 76, 135-150.

Tam K. & Kiang, M. (1992). Managerial applications of the neural networks: The case of bank failure predictions. *Management Science*, 38, 416-430.

Therneau, T. M. & Grambsch, P. M.. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Waters, D. (2011). *Supply Chain Risk Management: Vulnerability and Resilience in Logistics, 2nd edition*, Kogan Page Publishers, London. 60-61.

Xiang, A., Lapuerta, P., Ryutov, A., Buckley, J., & Azen, S. (2000). Comparison of the performance of neural network methods and cox regression for censored survival data. [Electronic version]. *Computational Statistics & Data Analysis, 34*(2), 243-257.

Zhang, J. & Lyn, T. (January–March 2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, *28(1)*, *204-215.*

Zupan, B., Demšar, J., Kattan, M. W., Beck, J. R., & Bratko, I. (2000). Machine learning for survival analysis: A case study on recurrence of prostate cancer. [Electronic version]. *Artificial Intelligence in Medicine, 20*(1), 59-75.