CUSTOMER RESPONSE MODELS WITH A DELAY FACTOR IN DIRECT MARKETING

Young H. Chun, E. J. Ourso College of Business
Louisiana State University, Baton Rouge, LA 70803, USA
chun@lsu.edu, (225)578-2506

Yoonhyuk Jung, School of Technology Management
Ulsan National Institute of Science and Technology, Ulsan, Korea
yjung@unist.ac.kr. +82-52-217-3122

ABSTRACT

Many growth-curve models are based on simplifying assumptions that are not valid in many practical situations. In the paper, we propose a probabilistic response model that has many desirable properties. Our geometric response model has three meaningful parameters – an ultimate response rate among the recipients, a resistance rate of respondents, and a delay factor. We show that these parameters can be estimated by the method of maximum likelihood. We finally fitted our response model to a mail survey data to show its outstanding performance, which is attributed to the delay factor that describes the delivery and processing time of responses.

Keywords: Response modeling; Direct marketing; Growth models.

INTRODUCTION

In direct marketing, customers are asked to take a specific action such as returning a questionnaire, placing a catalog order, mailing a prepaid postcard, calling a toll-free telephone number, clicking a link to a specific website, redeeming a discount coupon, or ordering a product online with a promotional code (Bose and Chen 2009). In such a case, the customers’ responses are directly traceable and easily measured by the direct marketer. Using the records of customers’ responses over time, we can predict the customers’ response rate and speed, and use such information in making important marketing decisions.

Suppose, for example, that a direct marketer mailed a catalog to the customers in a target population. After the launch of a direct marketing campaign, the marketer has recorded the number of orders that have been placed in each day. Based on the daily sales record, the marketer needs to estimate the total number of catalog items that will be ordered eventually. If the marketer underestimates the total demand, the catalog item in stock will be run out and the marketer may suffer the loss of customer goodwill or extra ordering and expedite shipping costs. On the other hand, over-stocking the catalog item in the first place may result in higher inventory, maintenance, and salvage costs.

A similar prediction problem can be found when we mailed out a questionnaire to the individuals in a target population and recorded the number of individuals who have responded to the survey.
form in each day. We have the same type of prediction problem with solicitation letters for fund-raising or credit card applications, discount coupons in Sunday newspaper, or e-mail advertisements with a promotional code.

The cumulative number of responses is usually an increasing function of time. Many curve-fitting models have been proposed to describe and predict the “growth curve” of responses. In the paper, we first propose a probabilistic response model with three parameters, one of which is a delay factor that describes the processing and delivery time of responses. Due to the delay factor, our response model has many desirable properties. With a mail survey data, we finally illustrate the outstanding performance of our response model over other conventional curve-fitting models.

PRELIMINARIES

Conventional response models

Suppose that we mailed out a survey form, catalog, or solicitation letter to \( N \) customers in a target population of customers who have been selected based on the recency, frequency, and monetary value of their past purchases (Baumgartner and Hruschka 2005). Let \( y = \{y_1, y_2, \ldots, y_k\} \) denote the number of responses received during each of the past \( k \) days (or weeks) after the launch of the direct marketing campaign. For notational convenience, let \( s_i = y_1 + y_2 + \ldots + y_i \) be the total number of responses accumulated by the end of the \( i \)th day. The cumulative number of responses \( s_i \) is usually a monotonically increasing function of time \( i \).

For a given response data \( s = \{s_1, s_2, \ldots, s_k\} \), many researchers have proposed various types of growth curves and considered different methods of estimating the model parameters. For example, Huxley (1980) made the first formal attempt to model the response pattern to a mail survey:

\[
s_i = N - \alpha \beta^i,
\]

where \( N \) is the number of questionnaires mailed initially and \( \alpha \ (> 0) \) and \( \beta \ (< 1) \) are unknown parameters to be estimated empirically. After some rearrangements, the growth curve in (1) can be written as a simple linear regression model,

\[
\ln[N - s_i] = \ln \alpha + i \ln \beta,
\]

from which we can find the least square estimators of \( \alpha \) and \( \beta \) for a given response data.

However, it is not reasonable to assume the randomness of residuals in the linear regression model. That is, the error terms around the regression line in (2) are not independent, identically distributed random variables with a constant variance. Furthermore, the response model in (1) is based on the assumption that \( s_i \) approaches \( N \) as \( i \) increases to infinity, which implies that all the questionnaires will be returned eventually. When \( i = 0 \), on the other hand, the cumulative number of responses \( s_0 \) in (1) has a non-zero value, \( N-\alpha \), which is not a desirable property.

Numerous researchers have modified Huxley’s original model or proposed alternative ones. (e.g., Hill 1981; McGowan 1986; Bauer 1987, 1991; Wilson and Singer 1991; Basu, Basu, and
Most response models have two or three parameters, whereas McGowan (1986) proposed a logistics curve that has five unknown parameters that have no meaningful interpretations. Those response models are classified into (i) the growth curve model and (ii) the probabilistic response model.

Most of the earlier research has focused on how to find the best growth curve that fits a given response data (Huxley 1980; Hill 1981; McGowan 1986; Bauer 1987, 1991). The method of least squares is usually used to estimate the parameter values. In the probabilistic response model, on the other hand, the daily response of each respondent is modeled as a Bernoulli process so that the total number of responses in each day could be a random variable from a geometric distribution. In such a case, the model parameters are estimated by the method of maximum likelihood (Wilson and Singer 1991; Chun 2012).

Finn (1983) concluded that “more research into the nature of response functions in mail surveys is needed. If a consistently accurate predictive technique can be found, it will be invaluable to users of mail surveys.” In the paper, we propose a new probabilistic response model that has many desirable properties. First, the cumulative number of responses is \( s_i = 0 \) when \( i = 0 \), and has an asymptote \( s_i < N \) when \( i = \infty \). Second, the response model is flexible enough to represent various types of response patterns with different shapes and locations. Third, the response model is parsimonious with a fewer number of parameters. Fourth, each of the model parameters has a meaningful interpretation. Few researchers have proposed response models that have all the desirable properties.

![Typical pattern](image1.png)

![Decreasing pattern](image2.png)

![Symmetrical pattern](image3.png)

**Figure 1.** Frequency distribution of the number of daily responses over time.

**Delay factor**

In most direct marketing campaigns, the number of daily responses \( y_i \) is increasing initially, reaching a peak, and then showing a longer tail dwindling over time as shown in Figure 1 (a). However, many researchers have assumed that the daily number of responses \( y_i \) is a monotonically decreasing function in time as shown in Figure 1 (b), and considered growth curves that look like a banana-shaped concave function. Such growth curves do not fit very well particularly in postal mail surveys, and Bauer (1991) proposed to arbitrarily exclude the first one
or two days (or weeks) of responses to have a better fit. Alternatively, other researchers have assumed that the frequency distribution of $y_i$ is symmetrical as shown in Figure 1 (c), and proposed $S$-shaped logistics or Gompertz curves (Fildes et al. 2008).

Recently, Chun (2012) proposed a geometric response model with two meaningful parameters – (i) an ultimate response rate among the recipients and (ii) a resistance factor of respondents. His response model is appropriate for Web-based or e-mail surveys in which the daily number of responses is geometrically decreasing in time as shown in Figure 1 (b). In the paper, we added a delay factor to his model to effectively represent the typical $S$-shaped response pattern in Figure 1 (a). If the delay factor is negligible, then the response pattern of our model is similar to the banana-shaped concave function in Figure 1 (b).

There are many cases in which the processing and delivery time is non-negligible. In postal mail surveys or catalog sales, for example, it takes longer time to deliver the request to a customer and receive his or her response. In the paper, the delay factor includes the time the postal service takes to deliver a questionnaire (or catalog) to the recipient, the time for a respondent to review and fill out the questionnaire, and the time to bring the response back to the direct marketer. The response model with a delay factor is called a “heterogeneous starting point” model in Basu, Basu, and Batra (1995), who assume that the delay time is a uniform (a.k.a., rectangular) distribution. In addition to the uniform distribution, we consider two more probability distributions and compare their performances in this paper.

**RESPONSE MODEL**

**Geometric model with a delay factor**

In a direct marketing campaign, suppose that we sent out a request to $N$ individuals simultaneously. Among the $N$ individuals, the proportion of the “respondents” who will eventually respond to the request is $\pi$, which is called the “ultimate response rate”. Due to procrastination, even those respondents do not reply immediately. For each respondent, let $p$ be the probability that he or she replies during a given day, and $q = 1-p$ is the daily “resistance rate” of a respondent. Thus, the number of Bernoulli trials for each respondent to react is a geometric distribution with a parameter $q$. 

![Response Model Diagram]
Figure 2. Flow chart of response patterns during the first 3 days.

Chun (2012) considered the geometric response model with the two parameters $\pi$ and $q$, in which the expected number of daily responses is decreasing in time as shown in Figure 1 (b). Now, we assume that each reply will be delivered in $d$ days later ($0 \leq d < \infty$), and the “delay factor” $d$ is a discrete random variable. At the cost of introducing the additional parameter $d$, we can represent various types of response patterns with different locations and shapes. Figure 2 illustrates the flows chart of responses during the first three days.

For a respondent, let $P_i$ be the probability that his or her reply will be received $i$ days after the launch of a direct marketing campaign. As shown in Figure 2, $P_i$ does not depend on $\pi$, but is a function of $q$ and $d$. The probability of receiving a series of responses, $y=\{y_1, y_2, \ldots, y_k\}$, during the first $k$ days can be described as a multinomial distribution with $(k+1)$ classes:

$$P[y \mid \pi, q, d] = \frac{N!}{(N-s_k)!} \prod_{i=1}^{k} y_i! \left[1 - \pi \sum_{i=1}^{k} P_i\right]^{N-s_k} \prod_{i=1}^{k} (\pi P_i)^{y_i}. \quad (3)$$

from which we can find the expected values of $y_i$ and $s_i$ as follows:

$$E[y_i] = N \pi P_i \quad \text{and}$$

$$E[s_i] = N \pi \sum_{j=1}^{i} P_j, \quad \text{for } i=1, 2, \ldots, k. \quad (5)$$

If we have the estimates of the parameters $\pi$, $q$, and $d$, we can predict the expected number of responses by a certain time, anticipated time period to achieve a certain level of responses, and the like. Thus, our primary goal is to estimate $\pi$, $q$, and $d$ based on the sample observations $y=\{y_1, y_2, \ldots, y_k\}$.

Maximum likelihood estimators

Suppose that a response data $y=\{y_1, y_2, \ldots, y_k\}$ is available at time $k$. It follows from the multinomial distribution in (3) that the “likelihood function” of $\pi$ is

$$L_y(\pi) = \left[1 - \pi \sum_{i=1}^{k} P_i\right]^{N-s_k} \prod_{i=1}^{k} (\pi P_i)^{y_i}. \quad (6)$$

The maximum likelihood estimator of $\pi$ is the one that maximizes this likelihood function. Note that the optimal value which maximizes the likelihood function $L_y(\pi)$ also maximizes its log-likelihood function, $\ln L_y(\pi)$. Therefore, it is more convenient to find the maximum likelihood estimator of $\pi$ from the following log-likelihood function:
\[ \ln L_\gamma(\pi) = (N - s_k) \ln \left[ 1 - \pi \sum_{i=1}^{k} P_i \right] + s_k \ln \pi + \sum_{i=1}^{k} y_i \ln P_i. \] (7)

If we take the first-order derivative with respect to \( \pi \) and set equal to 0, then we have

\[ \frac{d}{d\pi} \ln L_\gamma(\pi) = -\frac{(N - s_k) \sum_{i=1}^{k} P_i}{1 - \pi \sum_{i=1}^{k} P_i} + \frac{1}{\pi} s_k = 0. \] (8)

Solving this equation gives us the maximum likelihood estimator of the response rate \( \pi \) as follows:

\[ \hat{\pi} = \frac{s_k}{n} \left( \sum_{i=1}^{k} P_i \right)^{-1}. \] (9)

If we plug \( \hat{\pi} \) in (9) into the log-likelihood function in (7) and rearrange the expression, we have

\[ \ln L_\gamma(q, d) \propto \sum_{i=1}^{k} y_i \ln P_i - s_k \ln \sum_{i=1}^{k} P_i, \] (10)

where \( \propto \) denotes “is proportional to”.

The maximum likelihood estimates \( \hat{q} \) and \( \hat{d} \) are the ones that maximize this log-likelihood function in (10). Any optimization software, such as Microsoft Excel – Solver, can be used to find the maximum likelihood estimates of \( q \) and \( d \). With \( \hat{q} \) and \( \hat{d} \), we then find the maximum likelihood estimate of \( \pi \) from (9).

Note that \( P_i \) is a function of \( q \) and \( d \), where the resistance rate \( q \) is an unknown constant and the delay factor \( d \) is a random variable. If a specific distribution of the delay factor \( d \) is given, then we can specify the probability \( P_i \) in the log-likelihood function in (10). In the next section, we consider three different types of probability distribution function of the delay factor \( d \).

**VARIOUS DELAY FACTORS**

The reply of a respondent is delivered \( i \) days after the launch of a direct marketing campaign due to (i) his or her procrastination (i.e., resistance rate \( q \)) and (ii) the delivery time (i.e., delay factor \( d \)). Thus, in the geometric response model, the probability \( P_i \) that his or her reply will be received on day \( i \) is

\[ P_i = \sum_{j=1}^{i} q^{i-j} (1-q) P[d_{j-1}] = \sum_{j=1}^{i} q^{i-j} (1-q) P[d_{i-j}], \] (11)

where \( P[d_j] \) is the probability mass function of the delay factor. In a special case in which there is no delay, the probability distribution becomes
\[ P[d_j] = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \geq 1. \end{cases} \] (12)

In such a case, we simply have

\[ P_i = q^{i-1}(1 - q). \] (13)

Let us consider three different probability mass functions of \( d \) with a single parameter. First, suppose that the delay factor \( d \) has a discrete uniform distribution as in Basu, Basu, and Batra (1995):

\[ P[d | u] = \frac{1}{u+1} \quad d = 0, 1, 2, \ldots, u, \] (14)

where \( u \) is the upper limit of the uniform random variable. There is no delay if \( u=0 \). The expected value of the uniform delay is

\[ E[d | u] = \frac{u}{2}. \] (15)

It follows from (11) and (14) that

\[ P_i = \frac{1-q}{u+1} \sum_{j=\max(1,i-u)}^{i} q^{i-j}, \] (16)

which can be simplified further as

\[ P_i = \frac{1}{u+1} \min \{(1-q'), (q^{i-u-1} - q')\}. \] (17)

Second, suppose that the delay factor has a geometric distribution:

\[ P[d | r] = r^d (1 - r), \quad d = 0, 1, 2, \ldots, \infty, \] (18)

where \( r \) is a parameter, \( 0 < r < 1 \), to be estimated empirically. If \( r \) is close to zero, then the delay factor is negligible. The expected value of the geometric random variable is

\[ E[d | r] = \frac{r}{1-r}. \] (19)

It follows from (11) and (18) that

\[ P_i = (1-q)(1-r) \sum_{j=1}^{i} q^{j-1} r^{i-j}. \] (20)

Third, suppose that the delay factor has a Poisson distribution:
P[d | s] = \frac{s^d e^{-s}}{d!}, \hspace{1cm} d = 0, 1, 2, \ldots, \infty,

(21)

where s is a parameter, s>0, to be estimated empirically. The delay factor is ignored if s is close to 0. The average delay in (21) is

E[d | s] = s.

(22)

With the Poisson delay, it follows from (11) and (21) that

\[ P_i = (1 - q)e^{-r} \sum_{j=1}^{s-1} \frac{s^{j-1}}{(j-1)!}. \]

(23)

Figure 3 illustrates the three probability distributions in which the average delay is all \( E[d] = 2 \) days. Among the three distributions, the Poisson delay in Figure 3 (c) appears to be the most realistic in most practical situations.

![Figure 3](image)

(a) Uniform delay with \( u=4 \)  \hspace{1cm} (b) Geometric delay with \( r=2/3 \)  \hspace{1cm} (c) Poisson delay with \( s=2 \)

Figure 3. Various delay factors with \( E[d]=2 \) days.

Note that we may consider other discrete probability distributions with more than one parameter (Nadarajah and Kotz 2009). For example, the negative binomial distribution has been widely used in various consumer behavior models (Wagner and Taudes 1987) and product inspection models (Chun and Sumichrast 2007). However, we restrict our attention to the single-parameter delay factor to have a parsimonious response model. Thus, our geometric response model has only three parameters (i.e., response rate, resistance rate, and delay factor), all of which have meaningful interpretations. With a weekly response data, we compare the performance of the three delay factors in the next section and propose the best one.

**NUMERICAL EXAMPLE**

To illustrate our response model with a delay factor, we use the response data collected by Huxley (1980) as a part of his dissertation research. He mailed out questionnaires to \( N=4,314 \)
manufacturing firms and recorded the number of responses received by the end of each week during the 17-week period. His response data has been extensively used as a benchmark in subsequent studies by Hill (1981), Parasuram (1982), McGowan (1986), and Bauer (1991), among others.

Using the Huxley’s response data, we estimate the parameter values of our geometric response model with various delay factors. The results are given in Table 1. As a performance measure, we consider the sum of squared errors (SSE) of the cumulative number of responses $s_i$. The maximum value of the likelihood function in (10) is also considered as a performance measure.

<table>
<thead>
<tr>
<th>Delay Model</th>
<th>Response rate $\pi$</th>
<th>Resistance rate $q$</th>
<th>Delay factor $d$</th>
<th>SSE</th>
<th>Maximum likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>No delay factor</td>
<td>0.58972</td>
<td>0.88355</td>
<td>-</td>
<td>428,948</td>
<td>-5959.347</td>
</tr>
<tr>
<td>Uniform distribution</td>
<td>0.54976</td>
<td>0.83647</td>
<td>$u=2.0000$</td>
<td>129,894</td>
<td>-5628.563</td>
</tr>
<tr>
<td>Geometric distribution</td>
<td>0.53570</td>
<td>0.74138</td>
<td>$r=0.7415$</td>
<td>135,515</td>
<td>-5700.481</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td>0.53080</td>
<td>0.77457</td>
<td>$s=2.1629$</td>
<td>91,876</td>
<td>-5578.286</td>
</tr>
</tbody>
</table>

Table 1. Various delay models with estimates of $\pi$, $q$, and $d$.

Without the delay factor, the ultimate response rate is estimated as $\pi=0.58972$. The maximum likelihood estimate of the weekly resistance rate is $q=0.88355$. The SSE of our geometric response model without a delay factor is 428,948, which is much better than the SSE = 649,503 of the Huxley’s (1980) classical regression model in (2). If we include a delay factor, the

Figure 4. Actual and fitted values of the cumulative number of responses at time $k=17$. 

669472-9
geometric response model performs even better as shown in Table 1.

Among the three probability distributions of the delay factor, the Poisson distribution appears to be the best, followed by the uniform distribution. The Poisson delay has the smallest SSE and the largest value of the likelihood function. The outstanding performance of the Poisson delay distribution is anticipated from Figure 3, where the Poisson delay looks more realistic than the uniform or geometric delay distribution. In practice, we suggest using the geometric response model with the Poisson delay.

Figure 4 illustrates the Huxley’s (1980) original response data, along with the cumulative number of responses $s_i$ predicted by our geometric response model with the Poisson delay. The dotted curve in Figure 4 is the predictions of the Huxley’s (1980) classical response model. As contrasted in the figure, our S-shaped response curve with a delay factor is clearly a better choice than the Huxley’s banana-shaped concave curve for the 17-week mail survey data.

Figure 5 displays the cumulative number of responses $s_k$ up to $k=25$, predicted by the Huxley’s model and by our geometric response model with the Poisson delay. When the first $k=10$ week data is available, the Huxley’s growth curve has a negative value at $k=0$ as shown in Figure 5 (a), and overestimates the actual values from $k=11$ to 25 significantly. In fact, it approaches $N=4,314$ as $k$ approaches infinity. On the other hand, our geometric response model with a Poisson delay slightly underestimates the actual values from $k=11$ to 17, but it fits much better than the Huxley’s response model.

The predicted values based on the first $k=15$ week data are shown in Figure 5 (b). The S-shaped growth curve of our geometric response model predicts the cumulative number of responses by the end of the 25th week much better than the Huxley’s banana-shaped concave curve.

**CONCLUDING REMARKS**

In the paper, we propose a geometric response model with a Poisson delay that has many desirable properties. We also fitted our response model to the Huxleys’ (1980) empirical data to show its outstanding performance. However, Huxley’s response data has some anomalies; the first week is only two days, while other weeks have five days. In addition, follow-up mails were sent in weeks 4 and 7. To compare the performance of our proposed response model with those of conventional models, we may need more empirical data or extensive simulation studies. In any cases, we believe that our response model with the Poisson delay is clearly an improvement over traditional growth curve models.

Certainly it is possible to construct richer and more complex response models with more model parameters. For example, we assume that the resistance rate $q$ is constant throughout the entire process, but it could be a function of time or could be changed by some forms of follow-up or reminder mailings. Although we only considered a discrete-time case in the paper, our response model could be extended to a continuous-time case, in which each time period is not necessarily the same. This can be achieved by making appropriate modifications to our geometric response model with varying degrees of difficulty.
Figure 5. Predictions of the cumulative number of responses.
Another potentially fruitful area of research lies in a Bayesian response model that could incorporate our prior knowledge from similar direct marketing campaigns or expert opinions (Baesens et al. 2002; Rossi and Allenby 2003). Unlike other conventional response models that only give point estimates of unknown parameters, the Bayesian model can construct confidence intervals of parameters and test various hypotheses under different loss functions. The geometric response model in this paper has three unknown parameters, but the computational difficulties with three prior distributions could be overcome with an appropriate Monte Carlo Markov chain method or a Gibbs sampler (Hruschka 2006; Chun 2008).

With the increasing popularity of personal computers and Internet, many researchers have analyzed different shopping behavior of online customers (Van den Poel and Buckinx 2005). Thus, it would be interesting to compare the response rate, resistance rate, and delay factor between a traditional mail survey and a Web-based survey (Cobanoglu, Warde, and Moreo 2001; Kwak and Radler 2002). We can also analyze the effects on those parameter values of various response stimulants such as providing advance notice to respondents, utilizing different forms of postage, giving a variety of monetary and non-monetary premiums, and so on (Jobber and Saunders 1988).

Our response model can be applied to other areas as well. Meade and Islam (1998) reviewed various “diffusion models” for the spread of technological innovation or the penetration of a new product into the market. The response rate in a direct marketing campaign can be represented as a growth curve over time. Thus, it would be possible to use our geometric response model with a delay factor for diffusion models that describe the process of how new products get adopted over time (Tapiero 1983; Gottardi and Scarso 1994; Shore and Benson-Karhi 2007).

REFERENCES


