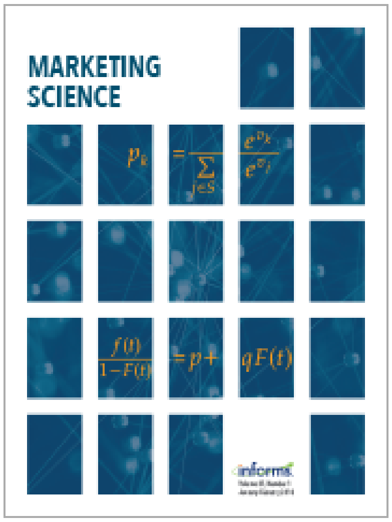

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

How to Compute Optimal Catalog Mailing Decisions

Füsun F. Gönülhttp://www.fusunconsulting.com, Frenkel Ter Hofstede,

MARKETING SCIENCE
Vol. 25, No. 1, January–February 2006, pp. 65–74
ISSN 0732-2399 | EISSN 1526-548X | 06 | 2501 | 0065

DOI 10.1287/mksc.1050.0136

How to Compute Optimal Catalog Mailing Decisions

Füsun F. Gönül
401 E. Waldheim Road, Pittsburgh, Pennsylvania 15215-1936
{fusungonul@hotmail.com, http://www.fusunconsulting.com}

Frenkel Ter Hofstede
Department of Marketing, McCombs School of Business, University of Texas at Austin, CBA 7.234/B6700,
Austin, Texas 78712, terhofstedef@mccombs.utexas.edu

We develop, estimate, and test a response model of order timing and order volume decisions of catalog customers and derive a Bayes rule for optimal mailing strategies. The model integrates the *when* and *how much* components of the response; incorporates the *mailing decision* of the firm; and uses a Bayesian framework to determine the optimal mailing rule for each catalog customer. The *Bayes rule* we propose for optimal mailing strategy allows for a broad set of objectives to be realized across the time horizon, such as profit maximization, customer retention, and utility maximization with or without risk aversion. We find that optimizing the objective function over multiple periods as opposed to a single period leads to higher expected profits and expected utility. Our results indicate that the cataloguer is well advised to send fewer catalogs than its current practice in order to maximize expected profits and utility.

Key words: catalog mailing; database marketing; econometric models; hierarchical Bayes
History: This paper was received March 1, 2002, and was with the authors 32 months for 2 revisions; processed by Michel Wedel.

1. Introduction

Customer response models provide feedback to database managers who need to formulate a mailing policy specific to a customer. In this paper we develop, estimate, and test a customer response model that controls for unobserved heterogeneity across customers, that simultaneously predicts the incidence (timing) and volume of orders from catalogs, that controls for response to the timing of mail, and allows for a broad set of objective functions to be optimized. Our model generates optimal mailing strategies for each customer using a Bayesian decision-theoretic framework.

In the extant literature simultaneous decisions of the customers, such as whether to buy, when to buy, and how much to buy, have been modeled before. See, for example, Gupta (1988), Chiang (1991), Bucklin and Gupta (1992), among others. Our work adds one more dimension by incorporating the firm's decision as well—that is, when the catalog is mailed. If the voluminous amounts of catalogs that clutter our mailboxes today are mailed efficiently and effectively—that is, at the right time to the right targets—distribution costs could be reduced, more happy customers can be won, and environmental waste could be reduced.

Database managers observe order timing and order volume decisions of their customers. The most common approach to derive optimal mailing strategies is with the recency, frequency, monetary-value (RFM) procedure (Roberts and Berger 1999). RFM uses information on previous orders to classify customers according to the recency (elapsed time since last order), frequency (number of orders), and monetary value (average dollar amount spent so far by the customer) of their purchases. Database managers segment customers into $R \times F \times M$ cells and rank order them according to their expected lifetime profits (customer lifetime values). The current trade practice is one step ahead of traditional marketing in that it segments customers by their purchase histories rather than the old-fashioned demographic profiling approach.

The importance of past behavior as a better future predictor than other characteristics such as demographics or lifestyles has been well documented in the marketing literature (see, for example, the seminal work of Guadagni and Little 1983 and, more recently, Rossi et al. 1996). Although the RFM procedure helps determine who on the house list should receive the next catalog, it is silent on the timing of mailings. Catalog managers tend to send a maximum of five catalogs per year on a more or less regular basis (four for each season and one for the winter holiday season). Lands' End and Quill regularly send out catalogs to their customers to assure the availability of a catalog (Schmid 1999). Because the mailing of catalogs comes

Table 1 A Contrast of the Literature with Our Work

	(1) Integrates "when" and "how much" to buy decisions	(2) Integrates the mailing decision	(3) Accounts for unobserved heterogeneity	(4) Considers whether the firm is risk neutral or risk averse	(5) Incorporates customer expectations of the future	(6) Chooses a theoretic basis for determining optimal mailing rules
Bult and Wansbeek (1995)	Y	Y	N	N	N	Y
Gupta (1988); Chiang (1991); Bucklin and Gupta (1992)	Y	N	N	N	N	N
Gönül and Shi (1998)	N	Y	N	N	Y	Y
Allenby et al. (1999)	N	N	Cross-sectional and temporal	N	N	Y
Elsner et al. (2004)	N	Y	N	N	Y	Y
Zhang and Krishnamurthi (2004)	Y	Y	Segment-level	N	N	Y
This study	Y	Y	Y	Y	N	Y

at a considerable cost, such approaches may lead to suboptimal mailing strategies and waste environmental resources.

Recent marketing research aims to provide guidelines for the optimal mailing strategy. A summary of the relevant literature and its contrast with our work is placed in Table 1.[1]

Bult and Wansbeek (1995) develop a method that selects targets from a mailing list for direct mail. The optimal selection strategy prescribed by the model maximizes profits by equating marginal returns to marginal costs, which results in a closed-form solution of the optimal selection strategy.

Gönül and Shi (1998) propose an estimable dynamic programming model that helps identify key determinants of direct-mail policies. The model is formulated under the assumption that customers maximize utility and the cataloguer maximizes profits. Allenby et al. (1999) develop a Hierarchical Bayes duration model of interorder times that captures cross-sectional and temporal heterogeneity. The model helps to predict at the individual level when and if a particular customer is likely to increase the time between orders that may indicate that a retention intervention is needed.

Elsner et al. (2004) investigate when to mail, who to mail, and the length of the mailing list decisions using a medium-term (not a myopic) approach for direct-mail campaigns. Zhang and Krishnamurthi (2004) estimate a model to simultaneously predict when to promote, how much, and to whom on the Internet. They develop an optimization approach to determine the optimal promotion for each household on each online shopping occasion.

We propose an empirical model that generates an optimal mailing strategy for each catalog customer. Our model endogenizes order volume decisions of customers in addition to order incidence decisions. Our model generates a customer-specific optimal mailing policy in accordance with the spirit of one-to-one (or direct) marketing, based on maximizing the manager's expected profits and/or utility over a finite horizon.[2] We also compare our approach with a simpler model where the manager aims to maximize retention only. This comparison is similar to sales maximization versus profit maximization in the economics literature. Last, but not least, we do not neglect to control for unobserved heterogeneity across customers in our specification.

2. A Hierarchical Bayes Model of Catalog Orders

In this section, we formalize the order decision process of customers, based on data from the house list of the cataloguer. The house list consists of the timing and volume of orders of customers across time and is typically used for predicting future purchases of customers. (The customers contained in a house list are those that placed at least one order from the catalog, as opposed to a rental list that consists of prospective buyers.) In addition, we have data on the mailing schedules of the firm that we utilize in the model formulation.

[1] In other direct-marketing-related works, Schmittlein et al. (1987) develop a method to identify the number of inactive customers in a database. This approach is empirically tested by Schmittlein and Peterson (1994) and is extended to predict the volume of purchases in future time periods, although volume and incidence are predicted in two independent models. Anderson and Simester (2004) report results from field studies to measure response from new and existing customers to deep discounts over two years.

[2] Of course, ours is not a fully dynamic approach where future expectations affect the current purchases, as in Gönül and Srinivasan (1996) or Gönül and Shi (1998), but rather a multiperiod approach that is essentially myopic in nature. In addition, we use the term expected profits to refer to net present value of future discounted profits and not to any future uncertainty.

2.1. Order Incidence

Customers' order incidence decisions are likely to depend on their previous purchases and the mailings they receive. We formalize the timing of orders through a discrete-time proportional hazard function. Assume $k = 1, \ldots, K_i$ orders are observed for customer i with length T_{ik}. The last spell is likely to be right-censored because the observation period ends before the customer has another chance to order.[3] In a particular spell, the discrete hazard function is the conditional probability of observing an order in time period t, given that no order has taken place so far (Kalbfleisch and Prentice 1980). That is, the hazard of customer i placing the kth order at time t is:

$$h_{ik}(t) = P[T_{ik} = t \mid T_{ik} > t-1]. \tag{1}$$

We specify individual-level (denoted by i) hazard functions because we are interested in predicting the responses of individual customers in the spirit of one-to-one or direct marketing. Under these assumptions, Equation (1) defines a semi-Markov process, where an order depends on the previous order through the duration between the two orders that is in line with the treatment of multiple spells in the hazard literature (Cox 1972, Kiefer 1988).

The hazard model specifications proposed in the marketing literature are either continuous or discrete. For example, Wedel et al. (1995) suggest a discrete piecewise hazard specification, based on a Poisson counting process. The underlying process we observe is that of durations between orders placed. Because the actual durations are continuous (customers can place orders at any point in time), but the observed data are discrete, we use a complementary log-log specification for the baseline hazard (please see Equation (2) following). The complementary log-log hazard specification follows from a process where the real underlying process is generated according to the continuous proportional hazard model, but the data are interval censored, i.e., only the interval, such as the week of the order, is known to the researcher. Results based on the complementary log-log link should be more robust to the choice of categories, which is a concern for discrete piecewise hazard models (see Ter Hofstede and Wedel 1998).

The order likelihood may vary with time elapsed since the mailing of the last catalog. A new catalog serves as a promotional offer and may tempt a customer to place an order. It may serve as a reminder ad, especially if the contents of catalogs remain largely the same from one issue to another. Customers may need some time to select products from the catalog, which will not result in a higher likelihood of an order in the first periods. However, as time passes the catalog may be dislocated, discarded, or considered outdated. This would negatively affect the order likelihood over time. A new catalog could also be considered a nuisance and may cause advertising wearout for some consumers (Little 1979, Simon 1982). We have no prior information to let us formulate hypotheses about customers' order behavior. Therefore, we formulate flexible functions accommodating these effects and let the data tell us the answers to these questions.

Hence, our hazard function specification also includes the time since the last mailing, $v_{ik}(t)$:

$$\begin{aligned}&\ln\bigl(-\ln(1-h_{ik}(t))\bigr)\\ &\quad = \phi_i + \rho_{i1} t + \rho_{i2}\ln(t) + \alpha_{i1} v_{ik}(t) + \alpha_{i2}\ln(v_{ik}(t)).\end{aligned} \tag{2}$$

Equation (2) links our model to the RFM trio in the following way. The intercept (ϕ_i) becomes a measure of customer-specific frequency (F), and ρ_{1i} and ρ_{2i} represent the (nonlinear) effect of recency (R) on order incidence.[4] Tendency to purchase at a given point in time may differ among customers. Therefore, we let intercept terms (ϕ_i) vary across customers. In addition, budget and storage constraints may curb the buying tendencies of buyers once the ordered goods arrive. The tendency to order may also fluctuate over time, thus yielding a variety of nonlinear curves depending on the signs of ρ_{i1} and ρ_{i2}. For example, if ρ_{i1} is positive and ρ_{i2} is negative, then the hazard function is U-shaped. When the signs are reversed the hazard function exhibits an inverse U. Similar arguments hold for the coefficients of the time since last mailing, α_{i1} and α_{i2}.

Our model is flexible. Please think of the right-hand side of Equation (2) as the right-hand side of a multiple regression equation. For example, we could have used quadratic terms instead of the log terms. However, the correlation between linear and quadratic terms exceeds 0.8 in our data and pose a danger of multicollinearity, whereas the correlation between linear and log term is about 0.6.

The first $K_i - 1$ spells are complete spells that end in an order so that the likelihood of observing duration T_{ik} equals the product of the hazard functions with the survivor function $S(t)$. The survivor function is the (unconditional) probability of not observing an order before t, and in a discrete-time framework it reduces to the product of hazard functions:

$$S_{ik}(t) = \prod_{u=1}^{t-1} (1 - h_{ik}(u)). \tag{3}$$

[3] In hazard function models a spell is defined as the elapsed time between two consecutive events; in this case, between two purchases.

[4] We center the right-hand-side variables on each customer-specific mean. That is, we use $(t_i - \bar{t})$ instead of t_i, etc.

The last spell T_{iK_i}, however, does not have to coincide with an order; it often remains incomplete because the observation period ends before the customer has a chance to make a purchase. We let d_i be an indicator such that $d_i = 0$ if the last spell of customer i is censored, and $d_i = 1$ otherwise. The likelihood function of the last spell equals the survivor function, $S_{iK_i}(T_{iK_i}) = S_{iK_i}(T_{iK_i} - 1)(1 - h_{iK_i}(T_{iK_i}))$, if the spell is censored ($d_i = 0$), and equals $S_{iK_i}(T_{iK_i} - 1)h_{iK_i}(T_{iK_i})$ if it is completed ($d_i = 1$). Because of the conditional independence of orders given the duration of the previous order, the likelihood of observing all order durations $T_{i\cdot} = (T_{i1}, \ldots, T_{iK})'$ of customer i is the product of likelihood functions over all spells:[5]

$$\begin{aligned} p(T_{i\cdot} \mid \phi_i, \rho_{1i}, \rho_{2i}, \alpha_{1i}, \alpha_{2i}) &= \left[\prod_{k=1}^{K_i-1} h_{ik}(T_{ik}) S_{ik}(T_{ik} - 1)\right] \\ &\quad \times \left[S_{iK_i}(T_{iK_i} - 1) h_{iK_i}(T_{iK_i})^{d_i} (1 - h_{iK_i}(T_{iK_i}))^{(1-d_i)}\right]. \end{aligned} \tag{4}$$

Even though orders seem to be independent of each other because they are all made by the same customer i, they are connected with the common thread of the individuality of the customer. We capture this connection by building a hierarchical structure on Equation (4), and hence accommodate heterogeneity of the hazard function across customers, as we explain at the end of §2.2.

2.2. Order Volume

We expect the volume of orders to be dependent on the timing of orders and timing of mailings. For example, if a customer recently ordered, she or he might buy again, in case she or he recognizes the need for a component of an augmented product originally ordered. Thus, a reorder, albeit in smaller amounts, is likely to occur immediately following a purchase. A customer may also wait and consolidate orders to save on shipment and handling costs. Such behavior gives rise to orders with higher volumes as time passes. In the case in which a customer decides to order, order volume may be higher over time because the customer has had more time to select products from the catalog and because the inventory at home has depleted. We have no a priori hypotheses about customers' spending behavior. Therefore, we refrain from speculating further and let the issue be empirically resolved.

Let $Z_{ik}(t)$ represent the volume of an order placed by customer i in the tth time period in order spell k. Conditional on an order ($T_{ik} = t$), we specify a lognormal distribution for $Z_{ik}(t)$ with mean $\mu_{ik}(t)$ and variance σ_Z^2. If no order is placed ($T_{ik} = t$), then $Z_{ik}(t)$ equals zero with perfect certainty. This leads to the following lognormal distribution for $Z_{ik}(t)$ with a spike at zero for $T_{ik} = t$:

$$\left[Z_{ik}(t) \mid T_{ik}, \mu_{it}(t), \sigma_Z^2\right] \sim \begin{cases} \operatorname{Log} N(\mu_{it}(t), \sigma_Z^2) & \text{if } T_{ik} = t, \\ I(Z_{ik}(t) = 0) & \text{otherwise.} \end{cases} \tag{5}$$

The lognormal distribution ensures that the predicted order volume is always positive. The lognormal distribution is known to perform well in other applications of estimating inherently positive variables, such as quantity demanded, wage income, and others.

We parameterize $\mu_{ik}(t)$ as:

$$\mu_{ik}(t) = \mu_{i0} + \varphi_{i1} t + \varphi_{i2} \ln t + \beta_{i1} v_{ik}(t) + \beta_{i2} \ln v_{ik}(t), \tag{6}$$

where the linear and log-linear terms accommodate monotonic and nonmonotonic effects. Note that the intercept (μ_{i0}) relates to the monetary value component (M) of the RFM trio.

The response parameters in Equations (4) and (6) are customer specific, allowing for different response curves among customers. Unobserved characteristics of customers may affect the shape of these curves. For example, larger households may order more on average, which leads to higher values of the monetary value parameter μ_{i0}. At the same time, the response parameters may be correlated. For example, households ordering less frequently could order larger amounts once they place an order, which will result in a negative correlation between ϕ_i and μ_{i0}. Therefore we let the customer-specific parameters ϕ_i, ρ_{i1}, ρ_{i2}, α_{i1}, α_{i2} follow normal distributions as follows:

$$\begin{aligned} &[\phi_i, \rho_{i1}, \rho_{i2}, \alpha_{i1}, \alpha_{i2}, \mu_{i0}, \varphi_{i1}, \varphi_{i2}, \beta_{i1}, \beta_{i2}]' \\ &\quad \sim N\big([\bar{\phi}, \bar{\rho}_1, \bar{\rho}_2, \bar{\alpha}_1, \bar{\alpha}_2, \bar{\mu}_0, \bar{\varphi}_1, \bar{\varphi}_2, \bar{\beta}_1, \bar{\beta}_2]', \Lambda\big), \end{aligned} \tag{7}$$

where Λ is a 10×10 covariance matrix. Thus, with this formulation, unobserved customer characteristics (e.g., household size, household member tastes, etc.) are captured by allowing a distribution of the parameters across customers.

2.3. Estimation

To estimate the customer response model, we use Markov Chain Monte Carlo (MCMC) methods (see, for example, Gilks et al. 1996). Our objective is to derive the posterior distribution of the parameters, given the data. We develop an estimation procedure that approximates the desired posterior distribution of the parameters by sampling parameters from their full conditional distributions. The prior distributions

[5] The equation assumes the process is stationary and is typical of hazard function models.

of the population parameters $\{\bar{\phi}, \bar{\rho}, \bar{\alpha}, \bar{\mu}, \bar{\varphi}, \bar{\beta}, \Lambda\}$ are taken from standard conjugate hyperdistributions:

$$(\bar{\phi}, \bar{\rho}, \bar{\alpha}, \bar{\mu}, \bar{\varphi}, \bar{\beta})' \sim N((\phi_0, \rho_0, \alpha_0, \mu_0, \varphi_0, \beta_0)', \Sigma_0)$$

and

$$\Lambda \sim \mathit{Wishart}(\lambda_0, \Lambda_0).$$

Hence, the full conditional posterior distributions of these parameters reduce to normal, gamma, and multinomial distributions. Values from full conditional distributions are sampled using Gibbs samples (see, for example, Gelfand and Smith 1990, Hoberts and Casella 1996).

The full conditionals of the customer-specific parameters $\{\phi_i, \rho_i, \alpha_i, \mu_i, \varphi_i, \beta_i\}$ and the variance parameter σ_Z^2 cannot be obtained in closed form. We use the Metropolis-Hastings algorithm to sample ϕ_i, ρ_i, α_i, μ_i, β_i from their full conditional distributions (Hastings 1970, Metropolis et al. 1953). We use slice sampling to obtain samples of σ_z^2 (Neal 2003). We use a normal proposal density in the Metropolis-Hastings steps and constrain the interval of σ_z^2 to the positive domain in the slice sampler. Taking random starting values, and iteratively applying the Gibbs, Metropolis-Hastings, and slice samplers, the sampled values converge to the posterior distribution of the parameters (Robert and Casella 1999).

3. A Bayes Rule for Optimal Mailing Strategy

We offer a decision-theoretic framework to derive the optimal mailing strategy by taking into account the expected profit and expected utility maximization motives of a database manager. We derive an optimal strategy for mailing catalogs to customers by building on the response model developed in the previous section.

Database managers need to decide on a regular basis (e.g., each month) whether or not to send a catalog to a specific customer. This does not mean, however, that expected profits in the future can be ignored. Our model instructs that customers' order and volume decisions depend on the durations since a previous order and since a previous promotion (catalog mailing). For example, if a customer's order volume strongly increases with the time since the last catalog is mailed, the mailing of a catalog to that customer may be postponed to generate profits in the long run, in spite of the short-run profits that an immediate mailing may generate.

Let $U(\Theta, D)$ be the utility function of the database manager, where Θ is the parameter set and D is a particular mailing schedule. If a particular mailing strategy D_1 is implemented and Θ_1 turns out to be the true characterization of the ordering process of customers, then a utility of $U(\Theta_1, D_1)$ is attained. Because Θ is not obtained with complete certainty, the uncertainty is integrated out of the utility by taking the expectation across the posterior distribution of the parameters given observed data, i.e., $E[U(\Theta, D) \mid T(D), Z(D)]$. This expression is the *expected utility* and is maximized with respect to mailing strategy, D. The value of D that maximizes the expected utility is an optimal Bayes rule with respect to the specified utility function.

First, we consider pure (risk-neutral) profit maximization. Let π_i be the profits over the next P months and Θ_i be the vector of customer-specific parameters. Let λ be the profit margin of orders and c the cost of sending a catalog.[6] Then the profit-maximizing utility function across P future months is:

$$U^{\pi}(\Theta_i, D_i) = \sum_{p=1}^{P} \kappa^p [\tilde{Z}_{ip}(\Theta_i, D_i)\lambda - D_{ip}c], \tag{8}$$

where κ is a discount factor $(0 < \kappa < 1)$ to discount future profits, and $\tilde{Z}_{ip}(\Theta_i, D_i)$ and $D_i = (D_{i1}, \ldots, D_{iP})$ are order volume and mailing decisions in the pth future decision period, respectively.

The optimal mailing strategy is likely to change if the decision maker is risk averse. The future volume of orders $(\tilde{Z}_i(\Theta_i, D_i))$ is unknown and introduces uncertainty into the profit equation (8). We incorporate risk aversion as follows:

$$\begin{aligned} U^{\tau}(\Theta_i, D_i) &= -\exp(\tau \pi_i) \\ &= -\exp\left(\tau \sum_{p=1}^{P_i} \kappa^p [\tilde{Z}_{ip}(\Theta_i, D_{ip})\lambda - D_{ip}c]\right) \end{aligned} \tag{9}$$

where τ accounts for the degree of risk aversion of the database manager. From the first-order condition, $dU^{\tau}(\Theta_i, D_i)/d\tau = \pi * U^{\tau}$, we observe that if profits are positive (negative), then as the degree of risk aversion increases, utility increases (decreases). Thus, the formulation suggests that when the business is flourishing, a risk-averse manager aims to preserve the status quo and is more cautious and conservative. In contrast, if the manager is losing money, then a higher degree of risk aversion reduces her/his utility. Therefore, the manager is better off taking some risks.[7] For a good review of risk aversion, see, for example, Keeney and Raiffa (1993).

Consider the case of deciding on a mailing strategy for one month ahead. Then, each customer can

[6] We do not model whether profit margins are affected by mailings and assume a constant profit margin (λ). In addition, we assume P is the same for all customers on the house list. In a more comprehensive model these issues can be addressed. We thank an anonymous reviewer for this point.

[7] We thank an anonymous reviewer for this point.

be mailed a catalog or not. Each of these decisions has a different (posterior) distribution of the profits (or losses) realized over the next month. For example, not sending a catalog may lead to a relatively limited expected profit with a 1% chance of monetary loss, whereas sending a catalog may lead to higher expected profits, but with a 40% chance of monetary loss. A risk-averse manager may choose not to mail to that customer even if it leads to lower expected profits. Accordingly, a higher value of τ indicates a higher degree of risk aversion.

The expected utility is obtained by taking the expectation of Equations (8) or (9) conditional on observed data:

$$E\big[U(\Theta_i, \widehat{D}_i) \mid T_i, Z_i, D_i\big] = \int_{\Theta_i} U(\Theta_i, \widehat{D}_i) p(\Theta \mid T_i, Z_i, D_i)\, d\Theta_i. \qquad (10)$$

To obtain the optimal Bayes rule $D_i^* = D_i$, the expected utility in Equation (10) is maximized. The integral in Equation (10) cannot be derived in closed form. However, the MCMC sampling scheme developed in the previous section allows us to generate a sample predictive distribution of order volume by imputing draws from the posterior distribution $p(\Theta_i \mid T_i, Z_i, D_i)$. We evaluate the expected utility by substituting the sampled values of $\widetilde{Z}_i(\Theta_i, D_i)$ into Equation (10). The mean value of the utility functions across the sampled values of $\widetilde{Z}_i(\Theta_i, D_i)$ provides an estimate of the expected utility.

The optimal mailing strategy is obtained by comparing the expected utility for each of the 2^P combinations of monthly mailing decisions (mail or do not mail in each of the P months). The mailing decision that corresponds to the highest expected utility is the optimal Bayes rule for mailing catalogs to that customer (D_i^*).

4. Empirical Application

4.1. Data

We apply our methodology to a sample of 424 customers from the house list of a major U.S. catalog company.[8] We are grateful to the catalog executives who shared with us their mailing schedule as well as order information. While order data are often easier to obtain, mailing information is more difficult because of the sensitivity of companies to sharing private information.

The items sold in the catalog are household products with an approximate lifespan of one to two years. The products are consumable all year round, and the items displayed in the catalog do not have large seasonal fluctuations in sales. The content of the catalog does not change substantially from one issue to the next. Only the cover of the catalog changes to reflect the season and to emit a sense of newness.

The customer database contains order histories of the company's customers, starting from the time their first order is placed. The observation period spans nine years, from January 1986 to December 1994. For predictive validity assessment, we use the observations in the last six months (July 1994 to December 1994) to construct a holdout sample. This leaves us with an average order history of about 44 months in the estimation sample, with a minimum of 6 and a maximum of 98 monthly observations. On average, customers place slightly less than one order per year (0.8 orders per year), spending an average of \$81 per order (see Table 2).

In addition to purchase histories in the customer database, we use the company's mailing schedule to derive for each customer when s/he received a catalog. We use this information to compute the recency of mailing variables, $v_{ik}(t)$. Similar to the order histories, we exclude the last six months for predictive validity assessment.

The company periodically employs the RFM procedure to select customers to the current list. A maximum of five catalogs are sent to a customer in a given year. Customers generally receive only some of the catalogs. In other words, not every customer gets each of the five catalogs because of cutoffs based on customer lifetime value considerations. On average, customers receive one catalog nearly every three months.

The average cost of sending a catalog is \$2.53. The average net revenue is \$28, which is the percentage profit margin (35% for this company) multiplied by the size of the average order in the data (\$81). Other than purchase and promotion (mailing) history there is limited customer-specific information in the data. They are: gender (92% female) and sensitivity to privacy (99%). Because there is little variation in these variables, they are omitted from our analysis.

4.2. Results

In estimating the model, we generate random starting values and sample iteratively using the MCMC

Table 2 Descriptive Statistics

Estimation period characteristics	Mean	Std. dev.	Min.	Max.
Observation period (months)	44.15	21.73	6.00	98.00
Order frequency (per month)	0.07		0	1
Order volume (\$)	80.91	82.31	10.99	599.98
Months since last order	10.20	9.79	1.00	60.00
Months since last mailing	2.98	2.71	1.00	24.00

[8] The information on the identity of the catalog and the product category listed in the catalog cannot be disclosed because of a confidentiality agreement with the catalog company.

sampling scheme.[9] We run the Gibbs sampler for 60,000 iterations after inspecting the sequence of samples of the posterior distributions across iterations. The MCMC chains appear to converge after a modest number of samples. However, to assure the accuracy of our results, we delete the first 30,000 samples and use the remaining 30,000 samples for further analysis.

We estimate the full model that permits correlations between each pair of the 10 parameters in our model and a simpler model without correlations across parameters. We find that all of the correlation credible intervals contain 0 (corresponding to insignificance at the 95% level in a frequentist terminology). We also find that allowing for correlations does not impact the other parameter estimates. Thus, for space and scope considerations we report results from the more parsimonious model in Table 3. All other results are available upon request.

In Table 3, we report the population parameters of interest, i.e., the median effects of purchase recency and mailing on incidence and volume of orders as well as the heterogeneity of these effects across customers. Our discussion in this section is general (for the median customer only). Individual responses vary due to heterogeneity.

The duration effects of orders, $\bar{\rho}_1$ and $\bar{\rho}_2$, as well as the mailing effects, $\bar{\alpha}_1$ and $\bar{\alpha}_2$, are substantial, as demonstrated by the relatively small credible intervals around their posterior means.

The heterogeneity of parameters across customers appears to be small. The standard deviations range from 0.3% to 10% of their corresponding means. Immediately following a purchase there is a 12% chance that a customer will reorder, possibly to complete the order with complementary products or with any other forgotten item in the previous order. For example, if a customer has recently ordered a major item, s/he may order a minor matching item immediately afterwards to complete the set. The median curve reaches its minimum after 5 months and remains relatively flat until about 15 months. Afterwards, the replacement effect comes into play and the likelihood of placing another order rises steadily.

The conditional order probability strongly increases in the first two months after mailing a catalog. Apparently, it takes some time before customers are exposed to the content of a catalog and decide upon their order. On average, the order probability slowly decreases after it reaches its maximum at four months. This points to a wearout effect of catalogs after it is being mailed. A catalog may be lost or misplaced, or perceived as old and outdated after six months.

Table 3 Posterior Median and 95% Credible Sets of Population Parameters

Parameter		Posterior median (95% credible set)	Parameter	Posterior standard deviation across customers (95% credible set)
Order incidence				
Intercept	$\bar{\phi}$	−2.716 (−2.792, −2.639)	$\sqrt{\lambda_{11}}$	0.1247 (0.0318, 0.3599)
t	$\bar{\rho}_1$	0.057 (0.050, 0.064)	$\sqrt{\lambda_{22}}$	0.0058 (0.0031, 0.0121)
$\ln t$	$\bar{\rho}_2$	−0.277 (−0.320, −0.240)	$\sqrt{\lambda_{33}}$	0.0067 (0.0031, 0.0244)
v	$\bar{\alpha}_1$	−0.078 (−0.101, −0.043)	$\sqrt{\lambda_{44}}$	0.0054 (0.0032, 0.0125)
$\ln v$	$\bar{\alpha}_2$	0.253 (0.204, 0.314)	$\sqrt{\lambda_{55}}$	0.0050 (0.0030, 0.0128)
Order volume				
Intercept	$\bar{\mu}_0$	4.052 (3.980, 4.127)	$\sqrt{\omega_{11}}$	0.3280 (0.2577, 0.4061)
t	$\bar{\varphi}_1$	−0.007 (−0.027, 0.010)	$\sqrt{\omega_{22}}$	0.1005 (0.0900, 0.1136)
$\ln t$	$\bar{\varphi}_2$	0.105 (0.001, 0.208)	$\sqrt{\omega_{33}}$	0.2784 (0.2119, 0.3744)
v	$\bar{\beta}_1$	0.030 (−0.047, 0.100)	$\sqrt{\omega_{44}}$	0.2030 (0.1523, 0.2358)
$\ln v$	$\bar{\beta}_2$	−0.064 (−0.221, 0.112)	$\sqrt{\omega_{55}}$	0.3397 (0.2364, 0.4819)

Note. The interval spanned by the 2.5 and 97.5 percentiles is referred to as the *credible interval* that is the Bayesian equivalent of a 95% confidence interval in a frequentist setting. If zero is not in included in the credible interval of a parameter, the probability that this parameter was less than or equal to zero is less than 0.05.

Thus, a cataloger is well advised to resend catalogs to increase response rates. The precise timing of the catalog mailing to generate peak response varies from customer to customer.

Recency of orders or recency of catalogs does not seem to have a substantial impact on order volume for the *median* customer. (The 95% credible intervals contain zero, except for $\bar{\varphi}_2$ being just contained in the interval.) Thus, at the aggregate level, there appears to be a small duration dependence effect of recency of orders on order volume. Order volume slightly increases when time since a previous purchase increases. The heterogeneity of the effects across customers, however, is high. Thus the relevance of elapsed time since the last order and elapsed time since the last catalog cannot be dismissed for a bulk of the customers, because the coefficients display relatively large variations across customers.

Why does there appear to be more heterogeneity in order volume than in order timing? We learn from company reports that the product is usable by all members in the household. It seems that the order timing response is more or less homogeneous across

[9] In order for the priors not to influence the posterior estimates, we use diffuse, weakly informative prior distributions, i.e., $\phi_0 = \rho_0 = \alpha_0 = \mu_0 = \beta_0 = 0$; $\Sigma_0 = 0.001 \cdot I_{10}$; $\lambda_0 = 6$; $\Lambda_0 = 10 \cdot I_5$.

customers and that the product is replaceable after a few months since the last purchase. Thus, we may conjecture that the wear and tear and therefore the replacement cycle of the product is about the same across households. However, some households, perhaps those with more members, tend to spend more than others do, conditional on an order. Therefore it may be that while the order timing follows a more or less consistent pattern across households no discernible pattern exists for order volume.

In summary, our model helps us understand customers' ordering behavior from catalogs. We find that the elapsed time since the last order and elapsed time since the last catalog is mailed, appear to be highly relevant in deciding when to order. We find that order volume displays considerable heterogeneity across customers in terms of its sensitivity to the recency of orders and to the recency of catalogs.

4.3. In-Sample and Out-of-Sample Fit Comparisons

We use the log-marginal density (LMD) function:

$$\mathrm{LMD}=\ln\left[\left(\frac{1}{Q}\sum_{q}\prod_{i,k}p\big(T_{ik},Z_{ik}\mid\phi_i^q,\rho_{i1}^q,\rho_{i2}^q,\alpha_{i1}^q,\alpha_{i2}^q,\mu_{i0}^q,\mu_{i1}^q,\mu_{i2}^q,\beta_{i1}^q,\beta_{i2}^q\big)^{-1}\right)^{-1}\right] \quad (11)$$

to compare our model with others. Further, we show the out-of-sample predictive power of our model vis-à-vis other models using minimum absolute deviations (MAD) of order incidence and order volume. The holdout sample consists of observations in the last six months of the purchase histories. The results are in Table 4. For ease of comparison, we rank the models from best to worst starting with **[1]** = Best.[10]

In the related literature that we discuss in the introduction, Gupta (1988) can be thought of as a version of our model without heterogeneity, and Bucklin and Gupta (1992) can be viewed as one with a discrete heterogeneity structure.

In-Sample Comparisons. The full model (Model 1 in the table) outperforms the simpler models (Models 2–5) because it has the highest LMD value. Changing the specification of heterogeneity does not affect the performance (Model 3, Model 4) as much as changing the functional form (Model 2, Model 5). Thus, it appears that functional form specification is more important here than heterogeneity, although the improvement due to heterogeneity cannot be ignored in achieving in-sample fit.

[10] A model without heterogeneity and without any effects would be even more simplistic than the RFM. We do not present it, as it would be easily beaten even by the RFM framework.

Table 4 Predictive Performance and Model Comparisons

Model	In-sample fit Log-marginal density	Out-of-sample fit: Minimum absolute deviation (order incidence)	Out-of-sample fit: Minimum absolute deviation (order volume)
1. Full model	41,386.215 **[1]**	0.1041 **[1]** (0.1021, 0.1064)	3.964 **[1]** (3.813, 8.358)
2. Mailing excluded	40,882.873 **[4]**	0.1086 **[2]** (0.1076, 0.1113)	3.993 **[2]** (3.890, 9.232)
3. Discrete heterogeneity	41,071.470 **[2]**	0.1124 **[4]** (0.1092, 0.1278)	4.269 **[4]** (3.814, 6.919)
4. No heterogeneity	40,996.333 **[3]**	0.1283 **[5]** (0.1153, 0.1317)	5.351 **[5]** (3.904, 6.831)
5. Log-terms dropped	38,409.423 **[5]**	0.1120 **[3]** (0.1091, 0.1155)	3.983 **[3]** (3.882, 7.581)
6. RFM	—	0.133 **[6]**	11.707 **[6]**

Note. In parentheses 95% credible intervals are shown. Models are ranked from **[1]** = best to **[6]** = worst in terms of fit (set in boldface).

Out-of-Sample Comparisons. Our model clearly predicts better than other models and the RFM procedure in terms of minimum absolute deviations of both order incidence and order volume. The worst out-of-sample fit is given by the RFM procedure (Model 6 in the table), followed by the model with no heterogeneity (Model 4). For out-of-sample fit, heterogeneity specification seems to matter (Models 3 and 4) more than functional form (Models 2 and 5).

4.4. Implementation of the Bayes Rule

We apply the optimal mailing rule described in §3 to the holdout sample to assess the profit implications of our approach. We derive optimal mailing strategies using several criteria (profit maximization, which is equivalent to risk neutrality, utility maximization with risk aversion) and two time horizons (one period versus multiple periods) for each individual customer. The risk-neutral and risk-averse utility functions are $U^{\pi}(\Theta, D(Y,Z))$ and $U^{\tau}(\Theta, D(Y,Z))$, given in Equations (8) and (9), respectively. The utility function for retention is:

$$U_{\pi}(\Theta,\widehat{D}_i)=I\left\{\sum_{p=1}^{P}\widetilde{Z}_{ip}(\Theta,D_{ip})>0\right\},$$

where $I(.)$ is an indicator function and the manager's optimization criterion is to maximize the number of retained customers (and not expected profits or utility). Table 5 summarizes the results.

The highest expected profits are achieved when the manager is risk neutral and when the optimization horizon is longer (23.82, entry 4 in Table 5). The highest expected utility is achieved when the manager is risk averse and when the horizon is longer (−5.874, entry 5).

Table 5 Optimal Strategies Under Different Criteria

Manager's optimization criterion	Time horizon	Expected profits (risk neutral)	Expected utility (risk averse, $\tau = 0.01$)	Average number of mailings per month (std. dev.)
1. Risk neutral	$P = 1$	21.04	−5.880	0.0991 (0.1213)
2. Risk averse ($\tau = 0.001$)	$P = 1$	21.35	−5.878	0.0295 (0.0937)
3. Risk averse ($\tau = 0.01$)	$P = 1$	21.54	−5.876	0.0295 (0.1114)
4. Risk neutral	$P = 6$	23.82 [*]	−5.879	0.2770 (0.1453)
5. Risk averse ($\tau = 0.001$)	$P = 6$	22.81	−5.876	0.1258 (0.0522)
6. Risk averse ($\tau = 0.01$)	$P = 6$	21.67	−5.874 [*]	0.1248 (0.0615)
7. Retention	$P = 6$	14.07	−5.937	0.5004 (0.2230)
8. RFM		12.80		0.4172

Note. [*] indicates the highest value in a column.

The results indicate that optimization across multiple time periods offers a substantial improvement over single-period optimization. The longer decision horizon proves more profitable because it takes into account the interaction of the customer with the company through the duration effects of orders and catalog mailings, as our empirical model shows.

According to our findings the cataloguer is well advised to send fewer catalogs than its current practice. Our model that yields the highest expected profits in a longer time horizon suggests mailing 0.277 catalogs per month (slightly over three per year). (See criterion 4 in Table 5.) In the data the average number of catalogs mailed per year is four to five per year, one for each season, and one for the winter holiday season. We also verify this in the last row of Table 5, where we show that the RFM procedure suggests slightly over five catalogs per year. The customer retention model, which is a simpler version of our full model, also suggests a high number of catalogs to be mailed in a year (about six). Catalog managers may save considerably on costs, and increase expected profits or utility by mailing less frequently.

5. Conclusion

In this study, we develop, estimate, and test a customer response model to predict order incidence and order volume decisions of catalog customers. We use a hierarchical Bayes approach to allow for individual-level predictions of order incidence and volume to be made. Our model integrates the mailing decision component of the problem as well. We develop an optimal Bayes rule for mailing strategies, based on a formal decision theory that allows a (risk-averse or risk-neutral) manager to optimize expected utility over a finite time horizon, where utility is defined as profits tempered by any risk aversion.

The response model we estimate shows that the effects of duration since the last order and duration since the last catalog is mailed, on decisions of order incidence and order volume, cannot be ignored. Of course, the degree of influence of each covariate varies across customers, which we capture by individual heterogeneity. While the house list appears to display a more or less consistent pattern in order timing decisions, we find considerable heterogeneity in order volume decisions.

Our model achieves a better in-sample fit and has a higher predictive accuracy, both in order incidence and order volume out of sample, than all the models it is compared to and the current company practice.

The Bayes rule we develop for optimal mailing strategy also indicates that expected profits and utility are higher when a longer time horizon is employed than a single-period horizon. This implies that current mailing decisions affect future decisions and that mailing strategies should optimize over a longer time horizon, updating strategies periodically. Cataloguers of household items typically follow the tradition of sending seasonal catalogs and one for the holiday season. However, our results show that the traditional RFM procedure will lead cataloguers to send out far too many catalogs than optimal.

No study is without limitations. In our case, lack of data prevents us from making general conclusions. For example, we observe when a catalogue was sent to a customer, but do not have data on when customers actually receive the catalog. Also, we do not observe competitive activities the cataloguer faces from stores, catalogs, or websites. We leave the research on competition (or synergy) between Web shopping and catalog shopping to future research, provided data sets with the desired level of richness are available. Our model focuses on existing customers only, and not on acquisition of new customers, which is of course an important part of any business. We do not have data on acquired customers.

The technique we present is sufficiently flexible to accommodate more complex objective functions. In our study, we specify cost functions that are independent across customers and could be optimized for each customer separately. However, economies of scale or budget constraints will render the cost functions dependent among customers so that the integer programming techniques are required to obtain the optimal strategy. We leave this issue to future research. Time aggregation may affect model performance; however, simulation studies have shown that difference between weekly and monthly calibration are small, as addressed in Ter Hofstede and Wedel (1998). However, please note that our method does not need any closed-form solutions for the Bayesian utility function; we can generate samples of the posterior distribution of any objective function with or without constraints.

Acknowledgments

The authors are grateful for the valuable comments of three anonymous reviewers, the AE, and the Summer Camp participants at Carnegie Mellon University, May 2000; the Marketing Workshop participants at New York University, June 2000; the Marketing Science Conference participants, Wiesbaden, Germany, July 2001; and the Strategic Leadership Forum in Pittsburgh, Pennsylvania, February 2002. The usual disclaimer applies.

References

Allenby, Greg M., Robert P. Leone, Lichung Jen. 1999. A dynamic model of purchase timing with application to direct marketing. *J. Amer. Statist. Association* **94**(446) 365–374.

Anderson, E., D. Simester. 2004. Long-run effects of promotion depth on new versus established customers: Three field studies. *Marketing Sci.* **23**(1) 4–20.

Bucklin, Randall E., Sunil Gupta. 1992. Brand choice, purchase incidence, and segmentation: An integrated approach. *J. Marketing Res.* **May** 201–215.

Bult, Jan R., Tom Wansbeek. 1995. Optimal selection for direct mail. *Marketing Sci.* **14**(4) 378–394.

Chiang, Jeongwen. 1991. A simultaneous approach to whether, what, and how much to buy questions. *Marketing Sci.* **10**(4) 297–315.

Cox, David R. 1972. Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Bull.* **34** 187–220.

Elsner, R., M. Krafft, A. Huchzermeier. 2004. Optimizing Rhenania's direct marketing business through dynamic multilevel modeling (DMLM) in a multicatalog-brand environment. *Marketing Sci.* **23**(2) 192–206.

Gelfand, A. E., A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Association* **85** 398–409.

Gilks, W. R., S. Richardson, D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Boca Raton, FL.

Gönül, F. F., Mengze Shi. 1998. Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Sci.* **44**(9) 1249–1262.

Gönül, F. F., K. Srinivasan. 1996. Estimation of the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Sci.* **15**(3) 262–279.

Guadagni, Peter M., John D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.

Gupta, Sunil. 1988. Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* **November** 342–355.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

Hoberts, J. P., G. Casella. 1996. The effect of improper priors on Gibbs sampling in hierarchical linear models. *J. Amer. Statist. Association* **91** 1461–1473.

Kalbfleisch, J. D., R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.

Keeney, Ralph L., Howard Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, UK.

Kiefer, Nicholas M. 1988. Economic duration data and hazard functions. *J. Econom. Literature* **26**(June) 646–679.

Little, John D. C. 1979. Aggregate advertising models: The state of the art. *Oper. Res.* **27** 629–667.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. 1953. Equations of state calculations by fast computating machines. *J. Chemical Phys.* **21** 1087–1092.

Neal, Radford M. 2003. Slice sampling. *Ann. Statist.* **31**(3) 705–767.

Robert, Christian P., George Casella. 1999. *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Roberts, Mary Lou, Paul D. Berger. 1999. *Direct Marketing Management*. Prentice-Hall Inc., Upper Saddle River, NJ.

Rossi, Peter E., Robert E. McCulloch, Greg M. Allenby. 1996. Value of household information in target marketing. *Marketing Sci.* **15** 321–340.

Schmid, Jack. 1999. When "W" factor. *Target Marketing* **22**(12) 43–44.

Schmittlein, D. C., Robert A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* **13**(1) 41–67.

Schmittlein, D. C., D. G. Morrison, R. A. Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Sci.* **33**(1) 1–24.

Simon, H. 1982. ADPLUS: An advertising model with wearout and pulsation. *J. Marketing Res.* **28**(February) 29–41.

Ter Hofstede, Frenkel, Michel Wedel. 1998. Time aggregation effects on the baseline of continuous time and discrete time parametric hazard models. *Econom. Lett.* **58** 149–156.

Wedel, Michel, Wagner A. Kamakura, Wayne S. DeSarbo, F. Ter Hofstede. 1995. Implications for asymmetry, nonproportionality, and heterogeneity in brand switching from piecewise exponential mixture hazard models. *J. Marketing Res.* **32**(November) 457–462.

Zhang, J., L. Krishnamurthi. 2004. Customizing promotions in online stores. *Marketing Sci.* **23**(4) 561–578.