MANAGEMENT SCIENCE
Vol. 52, No. 5, May 2006, pp. 683–696
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5205 | 0683

DOI 10.1287/mnsc.1050.0504

Dynamic Catalog Mailing Policies

Duncan I. Simester
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, simester@mit.edu

Peng Sun
Fuqua School of Business, Duke University, Durham, North Carolina 27708, psun@duke.edu

John N. Tsitsiklis
Laboratory for Information and Decision Systems and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jnt@mit.edu

Deciding who should receive a mail-order catalog is among the most important decisions that mail-order-catalog firms must address. In practice, the current approach to the problem is invariably myopic: firms send catalogs to customers who they think are most likely to order from that catalog. In doing so, the firms overlook the long-run implications of these decisions. For example, it may be profitable to mail to customers who are unlikely to order immediately if sending the current catalog increases the probability of a future order. We propose a model that allows firms to optimize mailing decisions by addressing the dynamic implications of their decisions. The model is conceptually simple and straightforward to implement. We apply the model to a large sample of historical data provided by a catalog firm and then evaluate its performance in a large-scale field test. The findings offer support for the proposed model but also identify opportunities for further improvement.

Key words: dynamic optimization; catalog mailing; field test; Markov decision process
History: Accepted by Jagmohan S. Raju, marketing; received April 7, 2003. This paper was with the authors $12\frac{1}{2}$ months for 3 revisions.

1. Introduction

Catalog firms mailed almost 17 billion catalogs in 2000 (Direct Marketing Association 2001). To determine who should receive these catalogs, firms typically estimate the probability that a customer will purchase from historical data. They then mail catalogs to all customers for whom this probability exceeds the breakeven level, at which mailing costs equal expected profits. In doing so, firms focus solely on the response to the next catalog, overlooking any long-term effects on demand. Yet, there is considerable evidence that receiving a catalog has an enduring impact on customer purchasing behavior beyond the current period. We propose a model that allows firms to address the dynamic implications of mailing decisions. In developing this model, we have several goals. First, the model is intended to be managerially relevant—it is conceptually simple and straightforward to implement. Second, we seek a model that is modular in the components that firms may choose to implement. As we will discuss, the model has two components: (1) the design of a discrete state space and (2) the optimization of the mailing policy on that state space. There are alternative procedures that can be used to perform each component, and these alternatives are substitutable. For example, firms may choose an alternative method to design the state space while using the procedures that we propose for optimizing the mailing policy (and vice versa). Third, we would like the model to be modular in the segments of customers on which firms choose to implement it. A firm may choose to implement the proposed model on customers with certain characteristics while retaining its current policy for other customers. Because the characteristics of the customers change over time, this modularity requires that the model explicitly take into account the possibility that the customers will shift between policies. Our final goal focuses on validation. To validate the proposed model, we use both historical data and a large-scale field test.

Because dynamic considerations have little influence on mailing policies for prospective customers, we restrict attention to past (house) customers. The proposed model requires data describing both the mailing history and the transaction history for each customer. Although maintaining a record of a customer's mailing history is no more difficult than maintaining a record of the customer's purchase history, many catalog retailers do not store complete mailing histories. This might be interpreted as an explanation for why the mailing history is typically not used to design the mailing policy. It is more likely that the causation operates in the reverse; many firms do not store the mailing history because they do not

use it. One explanation for this omission is that the mailing history is highly correlated with the purchase history, so that the purchase history provides a sufficient statistic. However, in practice, stochasticity in the mailing policy ensures that the purchase history is not a sufficient statistic.

The proposed model requires stochasticity in the historical mailing policy. For example, if the firm historically mailed only to customers who had recently purchased, then the model cannot predict how other customers would respond if they received a catalog. Fortunately, there is often considerable stochasticity in historical mailing policies. There are at least two primary sources for this stochasticity. First, stochasticity is introduced by regular randomized split-sample testing. The company that provided data for this study regularly conducts these types of tests, and discussions with other catalog companies confirm that the practice is widespread. The second source of variation in mailing policies reflects changes in the mailing policy over time resulting from changes in the models used to predict customer response rates. The company employs analysts who are continually searching for opportunities to improve the profitability of the firm's mailing policies. Other changes in management policies and personnel have led to ongoing changes in the mailing policy.

The requirement for stochasticity in the historical mailing policy explains in part the desire for a model that is modular in the segments of customers on which firms choose to implement it. The level of stochasticity in the historical policy often will vary across customers with different characteristics. For example, in the sample of historical data used in this study, the firm mailed to an average of 59% of its customers in each time period. However, for some of the most valuable customers, this percentage increased to 93%, whereas it was as low as 9% for some of the less valuable customers. Modularity allows for restricting application of the model to states in which there is sufficient stochasticity.

Literature

There is extensive literature investigating topics relevant to the catalog industry. This includes a series of studies that use catalog data to investigate pricing cues and the impact of price promotions (see, for example, Anderson and Simester 2004). Other topics range from customer merchandise returns (Hess and Mayhew 1997), to customer privacy (Schoenbachler and Gordon 2002), and catalog copy issues (Fiore and Yu 2001). In addition, several researchers have investigated optimal catalog mailing strategies. Bult and Wansbeek (1995) present a model for making mailing decisions that builds on work by Banslaben (1992). They develop a model to predict whether customers will respond to a catalog and link the model to the firm's profit function to derive a profit-maximizing decision rule. They evaluate their model using a sample of historical data provided by a direct marketing company that sells books, periodicals, and music in the Netherlands. They show that their methodology offers strong predictive accuracy and the potential to generate higher net returns than traditional approaches.

Bitran and Mondschien (1996) focus on the role of cash flow constraints when making catalog mailing decisions. The cash flow constraint introduces a trade-off between mailing to prospective customers and mailing to house customers. Mailing to prospective customers is an investment that yields negative cash flow in the short term but builds the company's house list, whereas mailing to the house list enables the firm to harvest value from its earlier investments. The model incorporates inventory decisions, so that the profitability of the mailing policy depends on the availability of inventory. The authors present heuristics that approximate a solution to their model and test the model using a series of Monte Carlo simulations.

As early as 1960, it was recognized that catalog companies may be able to profit by focusing on long-run rather than immediate profits when designing their mailing policies (Howard 2002). This recognition has led to several attempts to design dynamic catalog mailing policies. The most widely cited example was published by Gönül and Shi (1998). Drawing on the structural dynamic programming literature (see, for example, Rust 1994), Gönül and Shi propose a model in which customers optimize a stochastic and dynamic Markov game. In particular, the model assumes that customers understand both the firm's mailing strategy and the stochasticity in their own purchasing decisions. Among other factors, customer utility is an increasing function of whether they receive catalogs, and, therefore, it is assumed that customers contemplate how their purchasing decisions will affect the likelihood that they will receive catalogs in the future. For any value of the parameters of the customer utility function, a value function is defined, which corresponds to the solution of the postulated stochastic game. This value function results in a model of firm and customer behavior. The "true" parameters of the utility function are estimated using maximum likelihood by comparing the behavior predicted by the model with available data. The authors test their predictions using the purchase histories for 530 customers of a durable household goods retailer. The findings suggest that their proposed policy has the potential to increase the firm's profits by approximately 16%.

If their assumptions hold, Gönül and Shi's (1998) approach offers an important advantage over the model that we propose: It provides a means of predicting how customers will behave under mailing policies that do not arise in the historical data. As such, it does not require the same level of stochasticity in the historical policy as our proposed model. For example, even if the company mailed only to customers who had recently purchased, the response model provides a means of estimating how other customers would respond if they received a catalog. Of course, these predictions will be more accurate if there is stochasticity in the historical policy, so that there are past examples of mailing to all types of customers. The Gönül and Shi approach also may be able to better account for changes in customer behavior resulting from changes in the mailing policy (see limitations in §6).

These benefits come at some cost. First, as Gönül and Shi (1998) acknowledge, computation is very difficult when there are more than two state variables. In practice, firms often use a rich array of historical measures when designing their mailing policies. Second, the model depends on the specification of the utility function and an assumption that, in the available data, customers derive their own optimal dynamic policy based on their observation of the firm's policy. This assumption that customers derive their optimal policy appears strong in light of Gönül and Shi's conclusion that the firm's current policy is suboptimal. Empirially, we observe stochasticity in the firm's policy resulting from constant experimentations and tests, which further limits a customer's ability to "observe" the firm's true policy. Third, the response function simply predicts whether a customer will purchase and does not consider the magnitude of that purchase.[1] In practice, we also would like to consider the size of customers' orders. Finally, the model requires that the data identify which catalog a customer ordered from, which raises practical difficulties. Customers often do not have the catalog code when they are placing an order; therefore, it is not possible to link transactions with specific catalogs. Discussions with different catalog managers reveal that the Internet has greatly aggravated this problem, as customers can use the Internet as an ordering mechanism without reference to a catalog.

For these and other reasons, the Gönül and Shi (1998) model has attracted more attention from academics than practitioners, which reflects, in part, a difference in objectives. The structural dynamic programming literature has traditionally focused on understanding what factors affect customer or firm decision making. Agents are assumed to be optimizing dynamically, and the model then searches for parameters that yield the observed behavior as an optimal outcome. In this paper, we have a different objective. We propose a model with additional practical relevance that is conceptually simple and straightforward to implement. The proposal uses a different approach to overcome the practical limitations in the Gönül and Shi model. As we will discuss, we calculate transition probabilities and one-step rewards directly from the data. This direct (nonparametric) estimation of the customers' response function from the data does not impose functional form assumptions and allows us to greatly expand the dimensionality of the problem. The method has its own limitations, for which we propose solutions.

In a recent paper, Elsner et al. (2003) present a description of the success that Rhenania, a German catalog company, enjoyed when implementing a dynamic approach to optimizing catalog mailing policies. They used a lengthy series of split-sample mailing tests to estimate the response to different mailing frequencies together with a chi-square automatic interaction detection (CHAID) algorithm to segment customers. Rhenania's success confirmed that mailing to low-valued customers may be profitable even when these customers are unlikely to respond immediately. In a related proof of concept, Pednault et al. (2002) use a publicly available sample of direct-mail promotion data to compare a myopic policy with a dynamic policy estimated using reinforcement learning. Their results provide further evidence that a dynamic approach can significantly outperform purely static approaches to solving direct-mail circulation problems (see also Abe et al. 2002).

2. Overview of the Proposed Model

Before presenting the proposed model, it is helpful to begin with a brief overview and several definitions. We interpret the company's sequence of mailing decisions as an infinite horizon task (there is no end point) and seek to maximize the discounted stream of expected future profits. Time is measured in discrete periods defined by exogenously determined catalog mailing dates. The intervals between mailing dates typically vary; therefore, we will allow time periods to have different lengths. We use the term *reward* to describe the profit earned in any time period (r_t). This reward is defined as the net profit earned from a customer's order less mailing costs.

Customers' histories (and their current status) will be described at each time period by a set of

[1] In an appendix, Gönül and Shi briefly describe an extension to their model that considers how much customers spend on each purchase (where the amount spent is discretized into k brackets). However, computational limitations prevented estimation of this model.

n variables, so that a point in an n-dimensional space represents each customer at each time period. The n variables span a vector space $\mathbf{X}$. We will segment the space into mutually exclusive and collectively exhaustive discrete states (we use S to denote the set of discrete states). Intuitively, each state groups together neighboring observations (customers at each time period) that have comparable histories and are expected to respond in a similar way to future policies. Obviously, the design of the states is an important challenge, which we address in §3.

There are two possible actions at each time period: mail or not mail. We identify the action at time period t by $a_{ts} \in \{0, 1\}$, where $a_{ts} = 1$ denotes a decision to mail at time period t to every customer in state s. A policy (π) describes a mailing decision for each state. The firm's objective is to choose a policy that maximizes the following objective function:

$$V^{\pi}(s) = \sum_{t=1}^{\infty} \delta^{T_t} r_t^{\pi}(s) \quad \forall s. \tag{1}$$

Because the lengths of the time periods may differ, we define T_t as the number of days between the beginning of the initial time period and the end of the tth time period, and δ as a discount factor. Here, $r_t^{\pi}(s)$ is the immediate reward expected in time period t, under policy π, given the initial state at period zero was s.

We attribute purchases to the time periods in which they occurred, rather than the date of the catalog from which the customer ordered. This offers three advantages. First, it overcomes the practical problem described in the previous section in that it is often difficult to link a purchase to a specific catalog. This problem arises for approximately 15% of the purchases in our data set. This percentage is larger in the more recent data (and is expected to grow) because of an increase in the number of orders placed over the Internet. Second, attributing profits to the time period in which they occurred, rather than the date of the catalog, overcomes the need to explicitly consider cannibalization across catalogs. A customer who is ready to purchase will often purchase from a prior catalog if they do not receive the most recent catalog. As a result, customers may be *less* likely to purchase from a prior catalog if they are mailed another catalog two weeks later. If we attribute purchases to a specific catalog when evaluating the profitability of each mailing, we would need to account for the adverse impact of this mailing decision on the profitability of previous mailing decisions. This problem does not arise if we record purchases in the period they are earned irrespective of which catalog they are ordered from. Finally, treating purchases as a consequence of the stock of prior mailing decisions rather than a specific mailing decision is more consistent with our claim that the effect of mailing a catalog to a customer extends beyond the immediate purchase occasion. Customers' experiences with a catalog are not limited to the catalog that they ordered from; therefore, their purchasing decisions are not determined solely by that mailing decision.

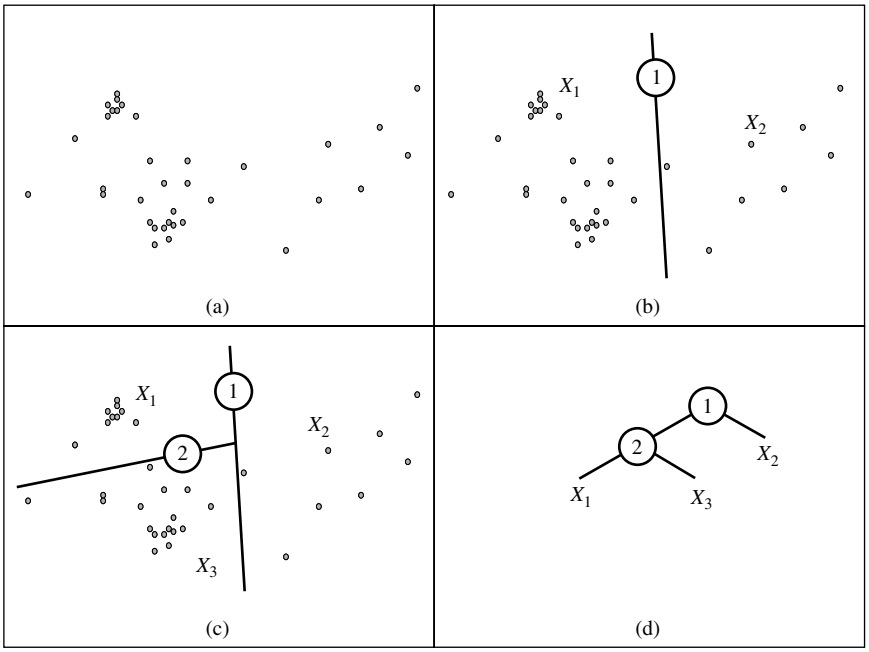
3. Constructing the State Space

The standard industry approach to designing a discrete state space is to tile the (continuous) state variables. There are several difficulties with this approach. Notably, it can yield a large number of states, and observations are often unevenly distributed across these states (many states are populated with few or no observations). An alternative approach is to develop a predictive model of how likely customers are to respond to a catalog and to discretize predictions from this model. The Direct Marketing Association (2001) reports that this approach, which will tend to yield fewer more evenly distributed segments, is used by approximately 28% of catalog firms. However, although this alternative is well suited to a myopic mailing policy, it is not well suited to a dynamic policy. There is no guarantee that grouping customers according to the predicted response to the next catalog will allow the model sufficient discrimination in a dynamic context. In particular, a new customer with few prior purchases may have the same purchase probability as an established customer who has extensive experience with the catalog. Yet, the long-term benefits of mailing the established customer may be different from the benefits of mailing the new customer.

In proposing a new algorithm for constructing a discrete state space, we adopt three objectives. First, the states should be "meaningful," so that each state $s \in S$ contains observations in the historical data. Second, the states should be "representative," so that data points in the same state are geometrically close to each other. Finally, the states should be "homogenous," so that the observations within a state share a similar profit stream given an identical mailing policy.

We begin by initially estimating a value function for each customer under the historical mailing policy. For a customer at point $x \in \mathbf{X}$, let $\tilde{V}^{\pi_H}(x)$ be an estimate of the present value of the expected discounted future profit stream under the historical mailing policy. Here, π_H indicates the historical mailing policy, and the tilde denotes the initial estimation. If the period of time covered by the historical data is sufficiently long, this estimate can be derived by fitting a function of the discounted aggregate profits earned for a representative sample of customers (see the implementation discussion in §5). Given $\tilde{V}^{\pi_H}(x)$, we use a series of separating hyperplanes to divide the state space into pieces organized by a binary tree structure.

Figure 1 State Space Design



We illustrate the intuition for the binary tree structure in Figure 1. Assume that we describe customers' history using just two variables ($n = 2$). A sample of data represented in this two-dimensional **X** space is portrayed in Figure 1(a). Line 1 represents a hyperplane in this **X** space that separates the sample in two subsegments (Figure 1(b)). The next iteration begins by selecting the segment with the highest variance for $\tilde{V}^{\pi_H}$ (not shown) and placing a second separating hyperplane (Line 2) through this segment. Following this second iteration, there are a total of three segments (Figure 1(c)). The process continues until a stopping rule is met, such as the desired number of segments or an upper bound on the largest variance in $\tilde{V}^{\pi_H}$ within any state.

The outcome is a tree structure (Figure 1(d)), in which the hyperplanes are branches on the tree and the segments are the leaves. A state space with N segments requires a tree with $N-1$ hyperplanes. Given the tree structure, the path from the root to each leaf node defines a set of inequalities identifying each state. Aggregation of states is also easily accomplished by pruning a large tree structure to a smaller one. This use of a binary tree structure is similar in spirit to the decision tree methods for classification (Duda et al. 2000) and the CHAID methods in customer segmentation (see, for example, Bult and Wansbeek 1995). The primary difference between the methods is the design of the hyperplanes determining the branches.

The algorithm that we use for identifying the hyperplanes proceeds iteratively, where each iteration has two steps. First, we select the segment for which the variance in $\tilde{V}^{\pi_H}(x)$ is the largest. Formally, we select the segment X_i for which $\sum_{x\in X_i}(\tilde{V}^{\pi_H}(x) - \bar{V}_{X_i})^2$ is largest, where $\bar{V}_{X_i}$ is the average of $\tilde{V}^{\pi_H}(x)$ calculated over all $x \in X_i$. This criterion favors the selection of segments that are least homogenous and/or have the most members. To prevent states with very few observations, we only select from among segments with at least 1,000 observations in them.

In the second step, we divide X_i into two segments X_i' and X_i''. To satisfy the homogeneity criterion, we would like the observations within each subsegment to have similar values of $\tilde{V}^{\pi_H}(x)$. To achieve this, we could fit a step function to the $\tilde{V}^{\pi_H}(x)$ values in X_i. However, computationally this is a difficult problem; therefore, we use a heuristic to approximate this step. The heuristic uses the following steps:

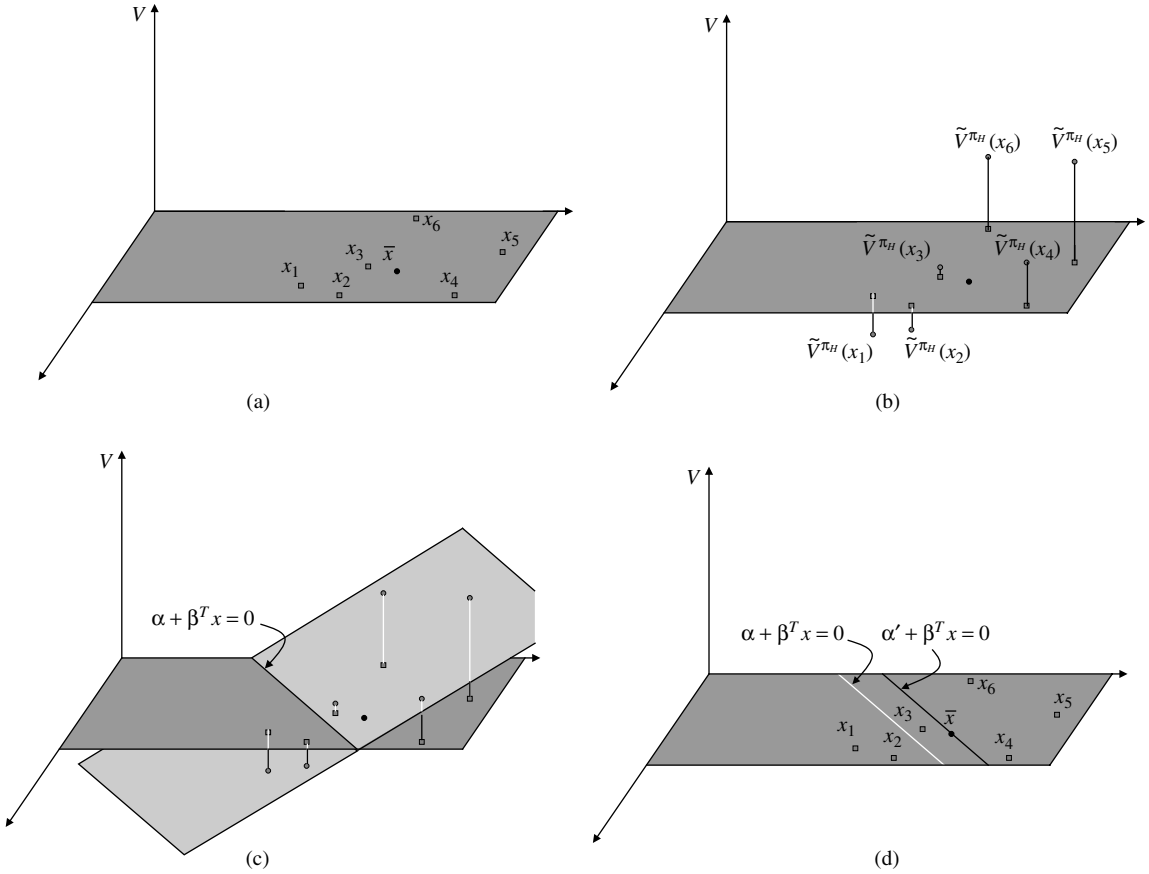
Step 1. Use ordinary least squares (OLS) to estimate α and β in $\tilde{V}^{\pi_H} = \alpha + \beta^T x + \varepsilon$ using all $x \in X_i$. That is, we find α and β that minimize $\sum_{x\in X_i}(\tilde{V}^{\pi_H}(x) - \alpha - \beta^T x)^2$.

Step 2. Find the center of the observations in the segment $\bar{x} = \sum_{x\in X_i} x$ by calculating the average of the observations on each of the n state variables.

Step 3. Compute α' such that $\alpha' + \beta^T \bar{x} = 0$ and divide segment X_i into two segments X_i' and X_i'' along the hyperplane defined by $\alpha' + \beta^T x = 0$.

We can again illustrate this process using a two-dimensional **X** space (see Figure 2). In Figure 2(a), we depict the observations in a selected segment. The center of these observations is defined by $\bar{x}$, and each observation has an estimated $\tilde{V}^{\pi_H}$ (Figure 2(b)). We use OLS to regress $\tilde{V}^{\pi_H}$ on x, which we illustrate in Figure 2(c) as a plane intersecting with the **X** space.

Figure 2 Dividing Segments



The intersection of the regression function and the **X** space defines a separating hyperplane ($\alpha + \beta^T x = 0$) that separates the chosen segment into two subsegments. The slope of the hyperplane is given by β, and its location is determined by α. To satisfy the meaningfulness objective, we locate the hyperplane so that it passes through the center of the observations in the segment (Figure 2(d)). We accomplish this by dividing along $\alpha' + \beta^T x = 0$.

The primary difference between this approach and other binary tree methods (such as CHAID) is that the hyperplanes need not be perpendicular to the axes of the **X** space. The use of a response measure ($\tilde{V}^{\pi_H}$) and the continuous nature of this response variable also distinguish this approach from both clustering and classification methods. Clustering methods generally do not include a response variable. They focus on the representative objective without regard to the homogeneity criterion. Classification methods do use a response measure, but they require that the response measure be binary or discrete.

4. Dynamic Optimization

Recall that the firm's objective is to maximize its discounted aggregate profits. Having designed a discrete state space, two tasks are required to identify the optimal policy: (1) For each state, we need to estimate the (one-period) rewards and transition probabilities for both the "mail" and "not mail" actions, and (2) using these estimated rewards and transition probabilities, we can use standard techniques to calculate the value function for a given policy and then iterate to improve on that policy.

Estimating the Rewards and Transition Probabilities

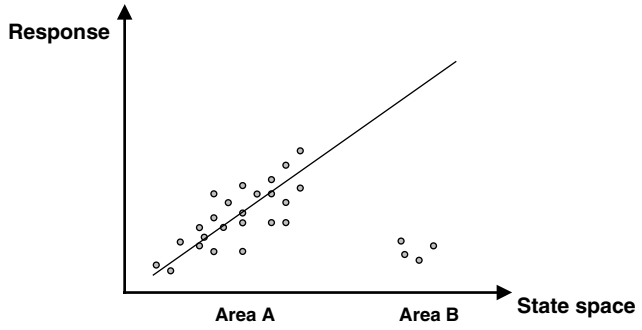
In the original applications for which dynamic programming was first proposed, the rewards and transition probabilities were known. However, in this application, and indeed in almost any social science application, these model parameters are not known and, instead, must be estimated from historical data. The traditional approach to estimating the rewards and transition probabilities is to estimate an underlying response process as a continuous function of the state variables according to an assumed functional form. This *parametric* approach is used by Gönül and Shi (1998), who estimate the probability that a customer will purchase as the underlying response process. We propose a different approach for estimating the rewards and transition probabilities. For each

state and mailing decision, we simply observe from the historical data the average (one-period) reward and the proportion of times customers transitioned to each of the other states. We claim that this *nonparametric* approach offers four advantages.

First, the next state, after a transition, depends not only on whether a customer purchased but also on how much they spent. Explicitly estimating a customer response function under the parametric approach, therefore, requires a model of purchase probabilities, together with a second jointly estimated model describing the size of the purchase (conditional on purchase). The process would be both extremely complex and sensitive to errors; therefore, it is unlikely that such a model would have practical relevance. Gönül and Shi (1998) abstract away from this problem in their model by both focusing on a very simple state space and only considering whether customers purchase (ignoring the variation in the size of those purchases).

Second, the functional form assumptions under the parametric approach can cause problems if the steady-state probabilities change. We use an example with a one-dimensional state space to illustrate this in Figure 3. Most of the historical observations are clustered in one portion of the state space (Area A). Because we can only estimate the response function using historical data, imposing a functional form favors accuracy in states with a lot of historical data (Area A) at the expense of states with few historical data (Area B). In Figure 3, we illustrate this trade-off by imposing a linear functional form. This may not be a problem if the steady-state probabilities do not change under different policies. But, if they do change, so that under the optimal policy customers transition to Area B, the errors introduced by the functional form assumption can be severe. One solution is to introduce additional degrees of freedom to the functional form, so that the response function is no longer linear. The nonparametric approach that we propose can be interpreted as an extreme interpretation of this suggestion. By estimating specific parameters for each state, we allow for any and all nonlinearities (and interactions) across states. Of course, the nonparametric estimates are less precise when there are fewer historical data, but the estimates are unhindered by the functional form restriction.

Figure 3 Functional Form Assumptions



Third, the optimization portion of the dynamic programming algorithm favors actions for which (1) the errors in the expected rewards are positive, and (2) errors in the transition probabilities favor transitions to more valuable states. This leads to upward bias in the value function estimates (and can be interpreted as an application of Jensen's inequality). We will later show that under the proposed nonparametric approach, we can overcome this bias through cross-validation, as drawing a new sample of data yields an independent set of errors. In contrast, this solution is not available under the parametric approach. The distribution of the data across the states will tend to be stable across draws of the data; therefore, the functional form assumptions ensure that the errors are not independent (the errors illustrated in Figure 3 will occur in each sample).

Finally, under the proposed nonparametric approach, the precision of the transition probabilities is known. In particular, under weak assumptions, the estimates of the transition probabilities follow a multinomial distribution. As a result, it is possible to approximate the bias and variance in the value function estimates (Mannor et al. 2005). Under the parametric approach, it is not clear from which distribution the estimates of the transition probabilities in each state are drawn. We could calculate the estimation errors for each state using the residuals; however, this is equivalent to reverting to the nonparametric approach.

As we acknowledged when distinguishing our approach from the Gönül and Shi (1998) model, the parametric approach does offer an advantage. If the assumptions hold, it provides a means of predicting how customers will behave under mailing policies that do not arise in the historical data. Of course, the Gönül and Shi approach also benefits from stochasticity in the historical data. Indeed, absence of stochasticity will make the continuous parametric approach particularly sensitive to the errors illustrated in Figure 3.

Policy Evaluation and Improvement

With the rewards and transition probabilities in hand, the value function can be calculated using Bellman's (1957) optimality equation:

$$V(s) = \max_{\pi} \mathrm{E}_{r,\, T(s),\, s'}[r_{s,\, \pi(s)} + \delta^{T(s)} V(s') \mid s,\, \pi(s)] \quad \forall s \in S. \quad (2)$$

Here, we use the notation $r_{s,\,a}$ for the random variable representing the immediate profit from the Markov chain after visiting state s and taking mailing action a,

δ for the discount factor per unit time, and T for the length of the intermailing time period after visiting state s. Because we anticipate that T would generally be included in the state variables used to define and construct the state space, we write T as $T(s)$, recognizing that T is a random variable whose distribution is determined by s.

For any fixed policy π, the following equation characterizes the expected discounted aggregate profits (value function) when starting at state s:

$$V^\pi(s) = \mathrm{E}_{r,T,s'}[r_{s,\pi(s)} + \delta^T V^\pi(s') \mid s, \pi(s)] \quad \forall s \in S. \quad (3)$$

If we use $\bar{r}_{s,a}$ to represent the expected rewards earned from a customer in state s when the firm chooses mailing action a, the above system of equations can be expressed as

$$\begin{aligned} V^\pi(s) &= \bar{r}_{s,\pi(s)} + \mathrm{E}_{T,s'}[\delta^T V^\pi(s') \mid s, \pi(s)] \qquad \forall s \in S \\ &= \bar{r}_{s,\pi(s)} + \sum_{s'} V^\pi(s') \sum_T \delta^T p_{s,\pi(s)\to T,s'} \quad \forall s \in S. \quad (4) \end{aligned}$$

Here, $p_{s,\pi(s)\to T,s'}$ represents the joint probability that a customer in state s after the mailing action a will transition to state s' and that the duration of the time period will be T. In the computation, we can directly estimate $p_{s,s',a} \equiv \sum_T \delta^T p_{s,a\to T,s'}$ from the data, which takes care of both the transition probability and the discounting. With a slight modification of notation, we can express Equation (4) in vector form. Let $\mathbf{P}^\pi$ denote a matrix for a given policy such that $\mathbf{P}^\pi_{i,j} = p_{i,j,\pi(i)}$, let $\bar{\mathbf{r}}^\pi$ denote the vector of expected rewards (with the ith element equal to $\bar{r}_{i,\pi(i)}$), and let $\mathbf{v}^\pi$ denote the vector with elements $V^\pi(i)$. Given this notation, we have $\mathbf{v}^\pi = \bar{\mathbf{r}}^\pi + \mathbf{P}^\pi \mathbf{v}^\pi$, which yields $\mathbf{v}^\pi = (\mathbf{I} - \mathbf{P}^\pi)^{-1}\bar{\mathbf{r}}^\pi$ as the value function under policy π.

Following the above notation, we can define a policy π_H for the historical mailing decisions. We assume that the historical mailing actions out of each state s follow the probability distribution observed in the data. The corresponding $\mathbf{P}^{\pi_H}$ and $\bar{\mathbf{r}}^{\pi_H}$ can be directly estimated from the data as well, which leads to the value function under this historical policy: $\mathbf{v}^{\pi_H} = (\mathbf{I} - \mathbf{P}^{\pi_H})^{-1}\bar{\mathbf{r}}^{\pi_H}$. This provides both a benchmark against which to evaluate the optimal policy and an obvious starting point for computing the optimal policy.

Having $\mathbf{v}^{\pi_H}$, we use the classical policy-iteration algorithm to compute the optimal mailing policy. The algorithm iterates between policy evaluation and policy improvement. In particular, the algorithm begins with a policy for which we calculate the value function. We then use this value function to improve the policy, which yields a new policy with which to begin the next iteration. The sequence of policies improves monotonically until the current policy is optimal. It is well known that the policy-iteration algorithm converges to a stationary policy that is optimal for the finite state infinite time horizon Markov decision process (Bertsekas 1995). In practice, the speed of convergence is surprisingly fast (Puterman 1994).

5. Implementation

We implemented the model on a data set provided by a women's clothing catalog that sells items in the moderate to high price range. We received data describing the purchasing and mailing history for approximately 1.73 million customers who had purchased at least one item of women's apparel from the company. The purchase history data included each customer's entire purchase history. The mailing history data was complete for the six-year period from 1996 through 2002 (the company did not maintain a record of the mailing history prior to 1996). In this six-year period, catalogs containing women's clothing were mailed on 133 occasions, so that on average a mailing decision in this category occurred every two to three weeks. The company also mails catalogs for other product categories, and the historical data received from the company contained a complete purchasing record for the other product categories.

State Variables

With the assistance of the firm, we identified a set of 13 explanatory variables to describe each customer's mailing and purchase histories. These variables can be grouped into three categories: purchase history, mailing history, and seasonality. We begin with a discussion of the purchase history variables.

Women's Clothing Purchase History

Purchase Recency$_{it}$: Number of days since customer i's most recent purchase prior to period t.

Purchase Frequency$_{it}$: Number of orders placed by customer i prior to period t.

Monetary Value$_{it}$: Average size in dollars of orders placed by customer i prior to period t.

Monetary Value Stock$_{it}$: Discounted stock of prior purchases.

Customer Age$_{it}$: Number of days between period t and customer i's first purchase.

Purchase History for Other Categories

Purchase Frequency$_{it}$: Number of orders placed by customer i prior to period t for items outside the women's clothing category.

The *Monetary Value Stock*$_{it}$ measure can be distinguished from the *Recency*, *Frequency*, and *Monetary Value* measures by the increased weight that it gives to more recent transactions. In particular, the measure is calculated as follows: $p_{it} = \sum_{j\in J_{it}} \eta^{T_j} x_j$, where J_{it} is the set of purchases by customer i prior to period t,

$\eta \in [0, 1]$ is a decay rate per unit of time, T_j denotes the number of units of time between period t and the jth purchase, and x_j describes the amount spent on the jth purchase. In preliminary analysis, we considered different values for these decay variables. This led to inclusion of two *Monetary Value Stock*$_{it}$ variables with decay rates 0.9 and 0.8 per month, respectively.

We used two mailing stock variables to describe the history of women's clothing catalogs mailed to each customer. These were defined analogously to the purchase stock measures: $m_{it} = \sum_{k \in K_{it}} \eta^{T_k}$, where K_{it} identifies the set of catalogs mailed to customer i prior to period t. The decay rates for these two mailing stock variables were set at 0.9 and 0.8 per week. These values were chosen because they yielded greater variance in the optimal mailing policies. The final estimates of the value function V were relatively stable to different values of these decay rates. We also considered a variety of variables describing customers' mailing and purchase histories from other product categories, but these variables had little effect on estimates of the optimal value function (V) or the optimal mailing policies.

Analysis of the raw data confirmed the presence of seasonality in both the purchasing and mailing histories. We used the following three variables to capture seasonality:

Purchase Seasonality$_t$: Average number of orders received in the corresponding week across all years in the data set.

Mailing Seasonality$_t$: Average number of catalogs mailed in the corresponding week across all years in the data set.

Individual Seasonality$_{it}$: Discounted sum of the number of purchases by customer i in the same quarter in prior years.

To smooth the purchase and mailing seasonality variables, we used a moving average for each measure. In the individual seasonality measure, we gave greater weight to more recent purchases by decaying prior purchases using an exponential weighting function (using a decay rate of 0.9 per year). We also considered including dummy variables identifying the four quarters in a year. However, these had little impact on the findings. In general, although it is obviously important to include variables describing seasonality, the findings were robust to modifications in these variables (such as the use of different decay rates in the *Individual Seasonality*$_{it}$ measure).

Finally, an additional variable was included to control for the variation in the length of each mailing period. This variable was labeled *Period Length*$_t$, and was defined as the number of weeks in the current mailing period (period t).

Design of the State Space

Having defined the vector space $\mathbf{X}$, we discretized it using the approach described in §4. This process is computationally intensive; therefore, we focused on a random sample of 100,000 of the 1.73 million customers for this step. Data for the first 25 mailing periods (1996 through July 1997) were used to initialize the mailing and purchase stock measures, and the period from August 1997 through July 2002 was used as the estimation period. This estimation period comprised 108 mailing periods.

To obtain initial estimates of the value function for the current policy ($\tilde{V}^{\pi_H}$), we randomly selected a mailing period in 1996 for each of the 100,000 customers and calculated the discounted profits earned from each customer in the subsequent periods. The randomization ensured that all values of the seasonality variables were represented. Using the total discounted profit as a dependent measure, we regressed $\tilde{V}^{\pi_H}$ as a quadratic function of the (n) explanatory variables describing the customers' mailing and purchase histories. To ensure that the estimates were robust, we repeated this process 100 times and averaged the resulting parameter estimates to derive final estimates for $\tilde{V}^{\pi_H}$.

The company supplements its purchase history data with additional information from other sources to make mailing decisions for inactive customers (defined as customers who have not purchased within the past three years). Because we do not have access to this additional data, this introduces the potential for bias in the calculated optimal value function. For this reason, we calculate only the optimal mailing policy for customers who purchased in the three years prior to the current time period. Specifically, we divided the vector space $\mathbf{X}$ into two half spaces $\mathbf{X}'$ and $\mathbf{X}''$, where observations in $\mathbf{X}'$ represent customers who purchased within three years of the current time period. The state space discretization procedure was then conducted separately to design 500 states in each of the $\mathbf{X}'$ and $\mathbf{X}''$ spaces.

Dynamic Optimization

Having discretized the state space, we calculated the value function estimates for both the current and optimal policies. The policy improvement procedure was conducted only on states in $\mathbf{X}'$. Before calculating the transition probabilities and expected rewards, we first randomly selected a validation sample of 100,000 customers (none of these customers were in the sample used to design the state space). For comparison purposes, we then separately estimated the transition probabilities and expected rewards for two different samples. Sample 1 represents the 1.63 million customers that remained after removing the validation sample; Sample 2 is a smaller sample of 100,000

customers randomly selected from this sample of 1.63 million customers.

As we discussed, the policy improvement algorithm focuses on the 500 states in which customers are active ($\mathbf{X}'$). We need separate transition probabilities and expected rewards for each of the two possible decisions ("mail" and "not mail"), yielding 1,000 "state–action pairs." Estimating the rewards simply requires calculating the mean reward for each of these state–action pairs. However, from each state, there are potentially 1,000 possible transitions (including the other 499 active states, the 500 inactive states, and back to the same state). Therefore, the transition matrix has one million elements (1,000 potential transitions from 1,000 state–action pairs). Fortunately, most of the transitions are infeasible. For example, a customer who has placed three orders cannot transition to states for customers who have purchased fewer orders. Indeed, if customers do not purchase in a mailing period, they frequently transition back to the same state. In Table 1, we summarize the amount of data available to estimate the expected rewards and transition probabilities under the 1.63 million and 100,000 customer samples.

Table 1 Sample Sizes Used to Calculate the Transition Probabilities and Expected Rewards

	Sample 1	Sample 2
Number of customers in the sample	1,639,363	100,000
Total number of observations across all mailing periods	82,404,362	4,702,845
Expected rewards		
Average sample size	82,404	4,703
Minimum sample size	301	71
Transition probabilities		
Percentage of transitions to the same state	42	38
Percentage of elements with zero transitions	80	90
Average sample size for each nonzero transition	405.0	45.5

Notes. An *observation* is defined as an active customer in a single mailing period. The missing data reflect the acquisition of some customers after the first mailing period. *Zero transitions* describe elements of the transition matrix that were never observed in the data.

Results

For ease of exposition, we refer to the improved policy as the "optimal" policy. However, we caution that the optimality of the policy is conditional on the design of the discrete state space and the accuracy of the transition probabilities and expected rewards. In Table 2, we report estimates of the current and optimal policy value functions for different discount rates. The discount rates are monthly interest rates, with a rate of 0.87% corresponding to an annual rate of 10%. The estimates for the current policy are derived using Sample 1. We restrict attention to active customers and weigh the estimates for each state by the number of visits to each state in the training sample. The table also reports the average percentage of (active) customers mailed a catalog in each mailing period.

Table 2 Value Function Estimates by Monthly Interest Rate

Monthly interest rate (%)	Current policy	Optimal policy	
		Sample 1	Sample 2
15	\$11.29	\$12.36	\$13.20
10	\$17.80	\$19.57	\$20.90
5	\$35.45	\$42.40	\$44.90
3	\$55.82	\$74.14	\$78.31
0.87	\$141.29	\$260.03	\$275.45
	Percentage of customers mailed (%)		
15	59	21	27
10	59	43	43
5	59	64	68
3	59	73	77
0.87	59	76	79

The value function for the current policy varies across interest rates. Although the policy does not vary, the rate at which future transactions are discounted affects the value function. The value function estimates for the optimal policy also vary with the interest rate. However, this variance reflects both the change in the rate at which future transactions are discounted and differences in the optimal policy. At lower interest rates, it is optimal to mail a higher proportion of customers because the model gives more weight to the favorable impact that mailing has on future purchasing.

The value function estimates for the optimal policy in Table 2 are consistently higher in Sample 2, which is the smaller of the two samples. Because the transition probabilities and expected rewards are calculated directly from the data, they inevitably contain error; the observed transition probabilities and expected rewards are only estimates of the true transition probabilities and expected rewards. The error in these estimates raises three important issues.

First, imprecision in the transition probabilities and expected rewards leads to variance in the value function estimates. Second, because the expression used to evaluate the value function $\mathbf{v}^\pi = (\mathbf{I}-\mathbf{P}^\pi)^{-1}\bar{\mathbf{r}}^\pi$ is nonlinear in the transition probabilities, errors in transition probabilities lead to bias in the value function estimates. Third, in choosing actions to maximize future discounted returns, the optimization algorithm favors actions for which the errors in the expected rewards are positive and errors in the transition probabilities favor transitions to more valuable states. This also leads to upward bias in the value function estimates and can be interpreted as an application of Jensen's inequality (recall the discussion in §4).

These issues have received little attention in the literature. Fortunately, the use of a nonparametric

Table 3 Corrected Value Function Estimates

Monthly interest rate (%)	Current policy ($)	Optimal policy	
		Sample 1 ($)	Sample 2 ($)
15	11.14	11.88	11.57
10	17.58	18.82	18.18
5	35.13	40.40	38.76
3	55.25	71.16	67.76
0.87	140.45	248.81	235.07
	Standard errors		
15	0.05	0.05	0.05
10	0.08	0.08	0.08
5	0.17	0.22	0.21
3	0.30	0.51	0.47
0.87	1.47	3.89	3.39

approach to estimate the transition probabilities and expected rewards yields a solution to all three issues. As we discussed, under weak assumptions, the estimates of the transition probabilities follow a multinomial distribution. The properties of this distribution can be used to derive expressions for the bias and variance in the value function estimates (Mannor et al. 2005). The nonparametric estimation of the model parameters also ensures that when redrawing a new sample of data, the errors are independent. As a result, we can test for the bias induced by the optimization by reestimating the value function for the optimal policy using the validation sample.[2] In Table 3, we summarize the corrected value function estimates and the standard errors of these estimates. Comparison of these findings with Table 2 highlights the impact that imprecision in the transition probabilities and expected rewards has on the value function estimates. After correcting the estimates, the value function estimates derived from the (larger) Sample 1 are now consistently higher than those derived from Sample 2.

In other findings of interest, we see that the benefits of adopting the optimal policy (compared with current policy) depend on the monthly interest rate. At monthly interest rates higher than 10%, the value function for the optimal policy is similar to that of the current policy. At these high interest rates, the objective function is relatively myopic, giving little weight to transactions that occur in later periods. The findings indicate that the improvement on the current policy is relatively small in these conditions. This is perhaps unsurprising given the myopic focus of the current policy and the extensive feedback that the firm receives about the immediate response to its mailing policies. However, as the interest rate decreases, so that more value is attributed to future earnings, the difference in the estimated value functions increases.

6. Field Test

Whereas comparisons of internal validity are common in the management science literature, tests of external validity at the individual level are rare. In this section, we describe the validation of the proposed model in a large-scale randomized field test conducted with the company that provided the historical data. The field test was conducted over a period of six months and included 12 mailing dates and a total of 60,000 customers. These customers were randomly selected by the firm from its database of 1.73 million customers, subject to the restriction that the customers had purchased within three years of the starting date of the field test. This restriction is consistent with the focus on active customers in our application and was designed to limit the number of inactive customers in the test.

Because the mailing strategies were varied for six months only, the predicted differences in the profits earned under the optimal and current policies are smaller than the differences presented in Table 3 (which evaluates permanent changes in the mailing policy). Moreover, the predicted improvements vary based on the initial states. In low-value states, the predicted profit improvements occur faster than in higher value states. For this reason, we broke the 60,000 customers into three approximately equally sized subsamples based on values of the states in which customers started the test. We label the subsamples as low-value, moderate-value, and high-value.[3]

The field test employed a 3×2 experimental design, reflecting the value of the states in which customers started the field test (the three subsamples) and the mailing policies used during the test period (two conditions). In particular, customers in each of the three subsamples were randomly assigned to either a treatment or a control group. In the treatment group, mailing decisions for all 12 catalogs mailed during the six-month test period used the proposed model, and the firm's current mailing policy was used for the customers in the control group.

Following the company's guidance, we adopted a 3% monthly interest rate when designing the mailing policy for the treatment group. Because of a time constraint, the estimated rewards and transition

[2] In contrast, for the reasons described earlier (see Figure 3), the errors are not independent across samples when using the parametric approach to estimate the transition probabilities and expected rewards. Under that approach, cross-validation is not available, nor is there an obvious way to derive expressions for the bias and variance in the value function estimates.

[3] We used (current policy) value function cutoffs of $24 and $71 to demarcate the three subsamples. There were 20,030, 20,061, and 19,909 customers in the low-, moderate-, and high-value subsamples, respectively.

probabilities were estimated using Sample 2 (100,000 customers). In this respect, the field study can be considered a conservative test of the model's potential.

Results

The results are summarized in Table 4, which shows both the profits earned during the six-month test period and the value function estimates for the customers at the end of the period.[4] We also report the sum of these two measures (which we label "total profit"). For each measure, we report the *observed* and *predicted* differences between the treatment and control groups. The findings for the low- and moderate-value customers are reassuring. The optimal policy transitioned customers to more valuable states by the end of the field test. For the moderate-value customers, this was done without incurring any lost profits during the six-month test period. For the low-value customers, transitioning customers to the more valuable states required increased mailing frequencies during the test period, and these additional mailing costs outweighed the additional revenue earned during this period. Although the observed reduction in profits during the test period was larger than we had predicted (−96% versus −56%), the total improvements were roughly consistent with the predicted improvements (7% versus 11%).

Unfortunately, the outcome of the field test for the high-value customers was less favorable. For these customers, the optimal policy mailed less frequently than the control policy and generated less revenue as a result, with the reduction in sales outweighing the savings in mailing costs. Although the findings for the high-value customers were disappointing, a more detailed examination of the data reveals some promising evidence even for these customers. In Figure 4(a), we report the percentage difference between conditions in the average weekly gross profit earned from these customers during the field test. The gross profit measure includes revenue less the cost of goods sold (it does not include mailing costs). In Figure 4(b), we report the percentage of customers mailed on each of the 12 mailing periods in the treatment and control conditions.

Throughout the test, the mailing rates in the treatment condition were considerably lower than in the control condition. Although there is an upward trend in the mailing rates under the treatment condition, at the end of the test, the number of catalogs mailed in the treatment condition was still almost 20% fewer than in the control condition. The gross profits (excluding mailing costs) are initially a lot lower in the treatment condition, but by the end of the test period they meet or exceed the profits in the control condition, despite the lower mailing rates. Indeed, in the last four weeks of the test, 20% fewer catalogs were mailed to customers in the treatment condition, yet the company earned over 2% more in gross profit (compared with the control condition).

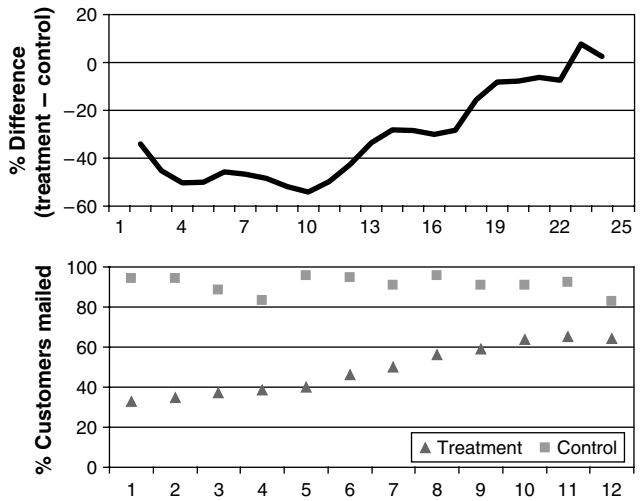
Further investigation revealed an explanation for the poor initial outcome with the high-value customers. In the historical data, the firm mailed over 85% of the time to the high-value customers, so that only 15% of the data in these states was available to evaluate what would happen if the firm did not mail to these customers (it was as low as 7% in one state). Moreover, almost all of these "not mail" data occurred on just nine of the 108 mailing dates in the historical data. It seems that there is simply insufficient data to reliably predict the impact of not mailing to the firm's most valuable customers. To identify an optimal pol-

Table 4 **Field Test Results: Percentage Difference (Treatment – Control)**

	Low-value	Moderate-value	High-value
Six-month profit			
Predicted	−56	−2	0
Observed	−96	0	−26
Final value function prediction			
Predicted	22	7	5
Observed	27	14	−15
Total profit			
Predicted	11	5	3
Observed	7	10	−16

Note. The findings reflect the percentage difference between the treatment and control groups, calculated as (treatment – control)/control.

Figure 4 **(a) High-Valued Customers: % Difference in Gross Weekly Profit (Three-Week Centered Moving Average); (b) High-Valued Customers: Percentage Mailed Each Mailing Period**



[4] The value function estimates at the end of the test period were calculated using the firm's current policy. Using the optimal policy instead of the current policy has little effect on the pattern of results.

icy for its most valuable customers, the firm would first need to introduce stochasticity in its mailing policy to better predict the outcome of not mailing to these customers.

Limitations

The results of the field test are subject to some important limitations. In Table 4, total profit is calculated as the sum of the profits earned during the six-month test period, together with the value function estimates for customers at the end of that six-month period. The results rely on the accuracy of the value function estimates. Ideally, we would measure the discounted future profit stream actually earned from these customers after the field test. However, these data are not available.

A related issue raised by one of the reviewers is that the changes to the mailing policy may have affected customers' expectations about the future mailing policy. As a result, the behavior of customers in a given state at the end of the field test may differ from the historical behavior of customers in that state. It was this concern that motivated Gönül and Shi (1998) to explicitly model customers' expectations regarding a firm's mailing policy. Our model assumes that the state space is rich enough to describe changes in customers' purchasing behavior resulting from any changes in the mailing policy because mailing variables are included in the state space. In particular, we assume that any change in the mailing policy that affects customers' purchasing behavior is accounted for by customers transitioning to different states. If customers stay in the same state, the change in customer behavior would require a change in the Markov chain parameters within that state, for which the model does not allow. In practice, the state space may not be rich enough to discriminate between different mailing treatments. In this respect, the reviewer is correct that the model will not be capable of fully capturing customers' behavior changes (within a state).

Another related issue concerns the content of the catalogs. If there were changes in the catalog content across periods, this would also need to be included in the state variables. In this application, we attempted to describe these changes through the seasonality variables and limited our attention to catalogs from a single product category (recall that the catalog company also sells products in several product categories). To the extent that these variables did not fully capture variation in the catalog content, then, we have introduced additional noise. In practice, it will generally be impossible to fully capture variation in catalog content. For example, the design of garments may simply be more attractive in some seasons than in other seasons. This is presumably an important source of the stochasticity in the model, helping to explain why there is variation in rewards and transitions across observations.

7. Conclusions

We have presented a model that seeks to improve catalog mailing decisions by explicitly considering the dynamic implications of those decisions. The proposed model is conceptually simple and straightforward to implement. Moreover, it is modular both in the components that firms choose to implement and the segments of customers on which they implement it. We have validated the model using both historical data and a large-scale field test. The findings show considerable promise and also highlight opportunities for further improvement.

A limitation of the proposed model is that it requires stochasticity in the historical policy. In particular, if the historical policy mailed only to customers who had recently purchased, then we cannot estimate the effects of mailing to customers who have not purchased recently. Fortunately, there is often considerable stochasticity in the historical policy because of both randomized testing of mailing policies and changes in mailing policies over time. This is apparently true of catalog companies in general, not just the company from which we received data.

The level of stochasticity in the historical mailing policy varies across states. In the data that we analyze, the firm almost always mailed to its most valuable customers, so that there were insufficient data available to evaluate what would happen if the firm did not mail to these customers. Fortunately, the modularity of the model offers the firm the flexibility to implement the model only on customers for which there is sufficient stochasticity in the historical mailing policy. In particular, on receiving the results of the field test, the firm was enthusiastic about implementing the proposed model with its less valuable customers but preferred to maintain its current policy with its more valuable customers.

The results are also subject to two other important limitations. First, the model assumes that any change in the mailing policy that affects customers' purchasing behavior is accounted for by customers transitioning to different states. In practice, the state space may not be rich enough to discriminate between different mailing treatments. Second, we estimate the future behavior of the customers that participated in the field test using value function estimates from the proposed model. Ideally, we would measure the discounted future profit stream actually earned from these customers after the field test. Because these data are not yet available, we leave this issue for future research.

There are other important issues worthy of future research. When designing the state space, we sought

to group customers with similar value function estimates. We did so using an initial estimate of the value function under the historical policy. It would be interesting to investigate the extent to which the design of the state space would change if we used final value function estimates under the optimal policy. Although this issue is not specific to our proposed model, the issue has not received much attention in the literature. Nor is it obvious how to address the problem as most of the potential solutions have difficulties. For example, the value function estimates for the optimal policy are the same for all of the customers in each state; therefore redesigning the state space with these values will yield essentially the same states. The field test provides only six months' worth of data; therefore, it is not possible to calculate the on-policy value function directly from these data. Even comparing the steady-state probabilities under the current and optimal policies is a challenge. In the field test, we can evaluate the distribution of customers' final states. However, the field test focuses on a fixed sample of customers, whereas the dynamics in the overall system also reflect new customers arriving.

We have not addressed the issue of how many states to use in the analysis. This issue introduces a trade-off. Classifying the observations more finely by using a larger number of states offers additional degrees of freedom with which to optimize. On the other hand, as we have shown, the accuracy of the value function estimates depends on the accuracy of the rewards and transition probabilities. Using a larger number of states results in fewer observations to estimate these model parameters. Validating the policies on a separate sample of validation data offers one option for resolving this trade-off. Because this issue is not specific to the proposed model, we leave further investigation to future research.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant 0322823 and was partially funded by the MIT Center for Innovation in Product Development (CIPD), the MIT eBusiness Center, and the Singapore-MIT Alliance. This paper has benefited from comments by Charles Elkan, John Hauser, Shie Mannor, Olivier Toubia, and workshop participants at the AGSM (Sydney), Duke University, Queens University, MIT, Texas A&M University, University of Auckland, University of Chicago, University of Florida, University of Iowa, University of Western Ontario, the 2002 Fall INFORMS Conference, the Winter 2003 National Center for Database Marketing (NCDM) Conference, and the Spring 2005 New England Mail Order Association Spring Conference. The authors gratefully acknowledge the contribution of the company that provided the data for this study, and research assistance from Stephen Windsor.

References

Abe, Naoki, Edwin Pednault, Haixun Wang, Bianca Zadrozny, Wei Fan, Chid Apte. 2002. Empirical comparison of various reinforcement learning strategies in sequential targeted marketing. *Proc. 2002 IEEE Internat. Conf. on Data Mining*. Japan IEEE Computer Society, Maebashi City, Japan, 3–10.

Anderson, Eric, Duncan I. Simester. 2004. Does promotion depth affect long-run demand? *Marketing Sci.* **23**(1) 4–20.

Banslaben, John. 1992. Predictive modeling. E. L. Nash, ed. *The Direct Marketing Handbook*. McGraw-Hill, New York, 626–636.

Bellman, Richard. 1957. *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Bertsekas, Dimitri P. 1995. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, MA.

Bitran, Gabriel R., Susana V. Mondschien. 1996. Mailing decisions in the catalog sales industry. *Management Sci.* **42**(9) 1364–1381.

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

Direct Marketing Association. 2001. *Statistical Fact Book*, 23rd ed. Direct Marketing Association, New York.

Duda, Richard, Peter Hart, David Stork. 2000. *Pattern Classification*. Wiley-Interscience, New York.

Elsner, Ralf, Manfred Krafft, Arnd Huchzermeier. 2003. Optimizing Rhenania's mail-order business through dynamic multilevel modeling (DMLM). *Interfaces* **33**(1) 50–66.

Fiore, Ann Marie, Hong Yu. 2001. Effects of imagery copy and product samples on responses toward the product. *J. Interactive Marketing* **15**(2) 36–46.

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

Hess, James D., Glenn E. Mayhew. 1997. Modeling merchandise returns in direct marketing. *J. Direct Marketing* **11**(2) 20–35.

Howard, Ronald. 2002. Comments on the origin and application of Markov decision processes. *Oper. Res.* **50**(1) 100–102.

Mannor, Shie, Duncan I. Simester, Peng Sun, John N. Tsitsiklis. 2005. Bias and variance approximation in value function estimates. *Management Sci.* Forthcoming.

Pednault, Edwin, Naoki Abe, Bianca Zadrozny. 2002. Sequential cost-sensitive decision making with reinforcement learning. *Proc. Eighth ACM SIGKDD Internat. Conf. on Knowledge Discovery and Data Mining* (*SIGKDD*). Edmonton, Alberta, Canada.

Puterman, M. L. 1994. *Markov Decision Problems*. Wiley, New York.

Rust, John. 1994. Structural estimation of Markov decision processes. Robert F. Engle, Daniel McFadden, eds. *Handbook of Econometrics*, vol. 4. Elsevier Sciences, The Netherlands, 3081–3143.

Schoenbachler, Denise D., Geoffrey L. Gordon. 2002. Trust and customer willingness to provide information in database-driven relationship marketing. *J. Interactive Marketing* **16**(3) 2–16.