

A Hidden Markov Model of Customer Relationship Dynamics

Oded Netzer

Graduate School of Business, Columbia University, New York, New York 10027, on2110@columbia.edu

James M. Lattin, V. Srinivasan

Graduate School of Business, Stanford University, Stanford, California 94305
{jlattin@stanford.edu, srinivasan_seenu@gsb.stanford.edu}

This research models the dynamics of customer relationships using typical transaction data. Our proposed model permits not only capturing the dynamics of customer relationships, but also incorporating the effect of the sequence of customer-firm encounters on the dynamics of customer relationships and the subsequent buying behavior. Our approach to modeling relationship dynamics is structurally different from existing approaches. Specifically, we construct and estimate a nonhomogeneous hidden Markov model to model the transitions among latent relationship states and effects on buying behavior. In the proposed model, the transitions between the states are a function of time-varying covariates such as customer-firm encounters that could have an enduring impact by shifting the customer to a different (unobservable) relationship state. The proposed model enables marketers to dynamically segment their customer base and to examine methods by which the firm can alter long-term buying behavior. We use a hierarchical Bayes approach to capture the unobserved heterogeneity across customers. We calibrate the model in the context of alumni relations using a longitudinal gift-giving data set. Using the proposed model, we probabilistically classify the alumni base into three relationship states and estimate the effect of alumni-university interactions, such as reunions, on the movement of alumni between these states. Additionally, we demonstrate improved prediction ability on a hold-out sample.

Key words: customer relationship management; hidden Markov models; dynamic choice models; segmentation; Bayesian analysis

History: This paper was received July 13, 2005, and was with the authors 18 months for 2 revisions; processed by Prasad Naik.

1. Introduction

"In order to implement CRM, a company must have an integrated database available at every customer 'touch point' and analyze that data well. . . . (CRM) allows companies to automate the way they interact with their customers, and to communicate with relevant, timely messages." (Source: Peter Heffring—president of Teradata's CRM division, 2002).

Customer relationship management (CRM) has been a prominent aspect of business marketing for the past decade. Given the wide adoption of CRM in the business world, we aim to develop a model that could help businesses analyze transaction data to assess customer relationships and put forward a support system for marketing decisions. Recently, marketing scientists have started to develop models that relate customer relationships and database marketing through measures like customer duration and customer lifetime value (e.g., Reinartz and Kumar 2003). However, far less attention has been given to modeling the *dynamics* of customer relationships and the effect of encounters

between the customer and the firm on customer-firm relationships and the customer's choice behavior.

Marketers often engage in activities that are aimed at creating an enduring impact on the relationship between the customers and the firm,¹ such as loyalty programs and university reunions. These interactions between the customer and the firm are designed to move the customer into a different state with different behavioral propensities (e.g., where the customer is less likely to switch to a competitor or to exhibit price sensitivity). Once the customer is engaged in a certain behavior, this behavior is likely to affect subsequent relationship with the firm.

The objective of this research is to capture the dynamics of customer relationships. We suggest a modeling framework for estimating and understanding the relationship dynamics which is formed by a series

¹ Because the model proposed in this paper applies to the customer's relationship with firms, brands, services, or nonprofit and for-profit organizations, we use the term "firm" to represent the business partner for the relationship with the customer.

of customer-firm interactions. The proposed model allows one to probabilistically identify the customer's state of relationship at any given time and enables comparing the impact of alternative customer-firm encounters on moving the customer to a higher state of relationship.

We propose a hidden Markov model (HMM) in which the states are a finite set of relationship states. The transitions between the states are determined by a set of time-varying covariates such as customer-firm interactions, leading to a nonhomogenous HMM. The *relationship-state dependence* is defined by the dependency between the relationship state and the likelihood of the customer's purchase behavior. The number of states is determined by the complexity of the relationship and its dynamics over time. To distinguish between relationship-state dependence and zero-order heterogeneity (Fader and Lattin 1993), unobserved heterogeneity is captured through a set of random-effect coefficients. The HMM is estimated using a Markov chain Monte Carlo (MCMC) hierarchical Bayes procedure.

We apply the model to a university-alumni customer relationship data set. This empirical application stresses the value of the model for CRM marketers. We identify three states, which correspond to dormant, occasional, and active (very frequent) donors. The states are relatively "sticky" (large diagonal elements in the transition matrix). Attending a reunion seems to have a strong impact on moving alumni from the dormant to the occasional donation state and from the occasional to the active state. In contrast to the commonly used highest customer lifetime value approach, using the HMM we find only a small effect of reunion attendance on alumni in the frequent donation state. Volunteering to university roles, on the other hand, seems to have its primary impact on alumni in the dormant and active states, but not on alumni in the occasional state. In our empirical application, we also find superior predictive validity of the HMM relative to a heterogeneous, yet static, latent class model and a dynamic and heterogeneous recency-frequency model.

The remainder of this paper is organized as follows. Section 2 relates the current work to the relationship marketing and dynamic choice modeling literature. Section 3 develops the HMM for capturing the dynamics of customer relationships and describes the hierarchical Bayes estimation procedure. In §4 we describe the application of the proposed model in the context of alumni relations using longitudinal gift-giving data from the alumni association of a major private university. Section 5 concludes this paper with a discussion of the theoretical and practical contributions of this research, as well as an outline of directions for future research.

2. Relationship Marketing and Dynamics in Buying Behavior

2.1. Relationship Marketing Dynamics

Research in the area of relationship marketing has been emerging in the past decade both from the consumer behavior perspective (e.g., Fournier 1998) and from the empirical modeling perspective (e.g., Bolton 1998, Thomas 2001).

Theoretical models (e.g., Dwyer et al. 1987) suggest that relationships evolve (not always monotonically) through several discrete levels. In particular, it is suggested that relationships develop as a consequence of changes in the relationship's environment and interactions between the relationship's partners (Aaker et al. 2004, Fournier 1998, Hinde 1979). Furthermore, Oliver (1997) suggests that a discrete shift in the relationship occurs if the aggregate satisfaction from a sequence of critical incidents is strong enough to move the customer to a different "conceptual plane" of loyalty. Thus, transitions between relationship stages might be triggered by discrete encounters between relationship parties. For example, offering an airline traveler an upgrade to business class could serve as a critical incident (Flanagan 1954). If the act of upgrade and the experiences of the traveler during the business class flight pass the customer's satisfaction threshold, this critical incident could have a long-term impact on the traveler's relationship with the airline and the traveler's subsequent choice of flights. A sequence of discrete encounters between the customer and the firm constructs a *relationship*. Such encounters include transactions, service encounters, customer initiated interactions, or exposure and response to marketing actions initiated by the firm. We use the notion that relationships are built from a series of customer-firm encounters as the building block of our model.

Recently, with the increase in popularity of CRM software applications in the business world, more academic research has been focused on building relationship models using marketing databases. This includes models of customer lifetime duration (e.g., Allenby et al. 1999, Bolton 1998, Reinartz and Kumar 2003, Schmittlein and Peterson 1994) and customer lifetime value (e.g., Libai et al. 2002, Rust et al. 2004). With the exception of Reinartz and Kumar (2003), these models do not take into consideration the dynamics in the relationship that result from changes in the customer's environment, which is the main focus of the current study. Indeed, in a review of service and relationship marketing models, Rust and Chung (2006) suggest that future research should model the dynamics in customers' preferences as a function of the dynamic interactions between the customers and

the firm. In a review of customer lifetime value models, Jain and Singh (2002) suggest incorporating factors that drive consumer purchase over time and the stochasticity in the buying behavior in the customer relationship model.

2.2. Dynamics in Buying Behavior and Hidden Markov Models

Methodologically, the model we developed is more similar to the literature on marketing dynamics. Many marketing settings involve dynamics in consumer behavior. These situations include both individual-level and market aggregate dynamics. The difficulty with capturing such dynamics is that in most marketing data sets the number of observations or time periods observed is relatively small, and the nature and structure of dynamics is often latent. To capture the latent structure of dynamics in a relatively parsimonious way, researchers developed various approaches that could be generally divided into discrete or continuous state space structures.

If the dynamics are assumed gradual or smooth, one could use a continuous state structure to capture the dynamics. For example, time series autoregressive error models are used to capture the dynamics in sales and the long-term effect of marketing activities such as advertising (see Dekimpe and Hanssens 2000 and Pauwels et al. 2004 for a review). Naik et al. (1998) and Xie et al. (1997) use Kalman filtering to capture the dynamics in advertising scheduling and new product introduction, respectively. In the choice modeling literature, a smooth dynamic effect is often captured by a state-dependent term in the utility function using an exponentially smoothed sum (Guadagni and Little 1983, Srinivasan and Kesavan 1976) or a simple running average (Bucklin and Lattin 1991) of past purchases.

However, the continuous state space is inadequate to capture dynamics that are postulated to develop in a discrete manner such as an instantaneous regime shift in the market conditions or consumer preferences (e.g., due to an inclusion or a drop of a brand from the consumer's consideration set). One could model such dynamics, by allowing consumers (or markets) to transition over time between a set of discrete states. Probably the simplest demonstration of such discrete states in the choice modeling literature is the state-dependent model (Heckman 1981). In this model, the observed previous choice of the customer (captured by a lagged dependent variable) constitutes the customer state in the current choice occasion. Choice modelers include state dependence in their econometric models to capture heterogeneity across individuals as well as the serial correlation in purchases over time (McAlister et al. 1991). Using scanner panel data, Keane (1997) and Erdem and Sun (2001)

find positive and significant state dependence effects across product categories even after controlling for heterogeneity; other studies find mixed results (e.g., Jeuland et al. 1980). In the context of CRM, Pfeifer and Carraway (2000) use a Markov model between the observed purchase recency states to capture dynamics in customer lifetime value. Morrison et al. (1982) modified the brand switching Markov model to classify Merrill Lynch's customers into "prime" and "not prime" states using managerial judgment. In the context of alumni donations, Soukup (1983) used an ad-hoc dichotomization of past donations (donor and noncontributor) to define the customer's state of donation behavior.

A limitation of the observed states models is their restrictive account for buyer behavior dynamics, whereby an ad-hoc specification of state dependence is added to an otherwise static model. A second shortcoming of these models is that they often ignore other important sources of dynamics in buying behavior, such as the enduring effects of marketing stimuli. Indeed, for exogenous variables that are correlated over time, and are not controlled for, previous behavior might be a determinant of current behavior simply because it captures the effect of the omitted variables (Erdem and Sun 2001). This problem is likely to be more severe in the context of relationship marketing because marketing actions such as loyalty programs (Lewis 2004) and customer initiated interactions such as service encounters (Bolton 1998) might alter the customer's relationship with the firm, and therefore might have an enduring effect on the customer's buying behavior. Finally, often the researcher or marketer does not observe the consumer or market states that govern the dynamics.

To overcome the problem of unobserved states one could describe a set of latent states and transitions between these states and translate these latent states to the observed behavior through a stochastic model. This process can be described as an HMM. MacDonald and Zucchini (1997, Chapter 4) describe several applications of HMMs in areas ranging from biology, geology, and climatology to finance and criminology. The most common application of HMMs is in the area of speech recognition (Rabiner 1989, Rabiner and Juang 1993). In econometrics, Hamilton (1989) proposed an HMM to estimate the impact of discrete regime shifts on the growth rates of the real gross national product.

Within the marketing literature, HMMs are closely related to the family of latent class models (Kamakura and Russell 1989). Like most latent class models, HMMs classify individuals into a set of states or segments based on their buying behavior. However, unlike the latent class models, in HMMs the membership in the latent states is dynamic and follows

a Markov process. A handful of attempts have been made to model dynamic change in the latent segment membership in marketing applications (e.g., Poulsen 1990, Ramaswamy 1997). Wedel and Kamakura (2000, chapter 10) and Dillon et al. (1994) survey these studies as well as alternative forms of dynamics in segmentations (e.g., Böckenholt and Dillon 2000, Böckenholt and Langeheine 1996). Wedel and Kamakura conclude that the issue of nonstationarity in marketing segmentation should be further investigated. More recently, Smith et al. (2006) develop a Markov switching criterion for HMMs and empirically tested it in the context of dynamic effectiveness of advertising on brand sales. Montgomery et al. (2004) used a time-continuous HMM, which combines discrete states and continuous transition times to study web-path analysis. Fader et al. (2004) proposed a changepoint model to predict new product sales. Liechty et al. (2003) applies an HMM to identify visual attention mode in advertising viewing. Du and Kamakura (2006) use an HMM to identify latent states in American families' life cycles. Moon et al. (2007) use a random-effect HMM to augment unobserved competitors' promotions in a pharmaceutical context.

Our HMM of customer relationships pushes forward the marketing literature related to dynamic latent class models in several aspects. First, relationships are constructed from a series of interactions between the customer and firm. Because we are interested in understanding the effect of these interactions on dynamics, we relax the assumption made in all the marketing HMM applications mentioned above of stationary transitions between the latent states. We use a nonhomogeneous HMM (Hughes and Guttorp 1994) in which the Markovian transitions are a function of time-varying covariates. To our knowledge, this is the first paper to do so in the marketing literature. Allowing for time-varying covariates in the transitions is important if one wishes to understand the drivers of the dynamics rather than merely build a model that fits the dynamics in the data. Second, CRM data sets are often collected at the individual level. When modeling dynamics using individual-level data it is crucial to account for heterogeneity in order to distinguish cross-individual heterogeneity from dynamics. Most HMM applications estimate the model at the aggregate level or using aggregate data (see Liechty et al. 2003, Montgomery et al. 2004, and Moon et al. 2007 for exceptions). Finally, from a substantive point of view, our application of the HMM to the area of customer relationships brings an advanced methodological modeling approach to help address an emerging managerial need to manage customer relationships over time.

3. Model Development

3.1. The Hidden Markov Model²

The model described in this section is an individual-level model of buying behavior. We consider a set of customers, each of whom is involved in repeated interactions with a brand, firm, service provider, or institution. The marketer observes the choice history for each individual and the marketing environment at every time period. These data are similar to typical transaction data commonly used in choice models.

We define a *relationship encounter* as an interaction between the customer and the firm. Such interactions might include purchase transactions, exposure to relationship marketing activities, or other nonpurchase related exposure to firm. Relationships are made of a longitudinal sequence of relationship encounters. We further define a set of hidden (latent or unobserved) relationship states, which differ with respect to the strength of the relationship between the customer and the firm and the conditional likelihood of choice given the relationship state. The transitions between the states are probabilistically determined and are affected by relationship encounters. This structure of latent states and observed behavior can be modeled by an HMM.

An HMM is a model of stochastic process that is not directly observable but can be observed only through another set of stochastic processes that produces a set of observations. In the proposed HMM, the transition between the relationship states is characterized by a Markov process. This stochastic process is then transformed into the observed buyer behavior through the stochastic process of choice. Specifically, we develop an HMM of repeated binary choices that relates the transitions between the latent relationship states to the observed buying behavior (see Figure 1 for a graphical representation of the proposed HMM).

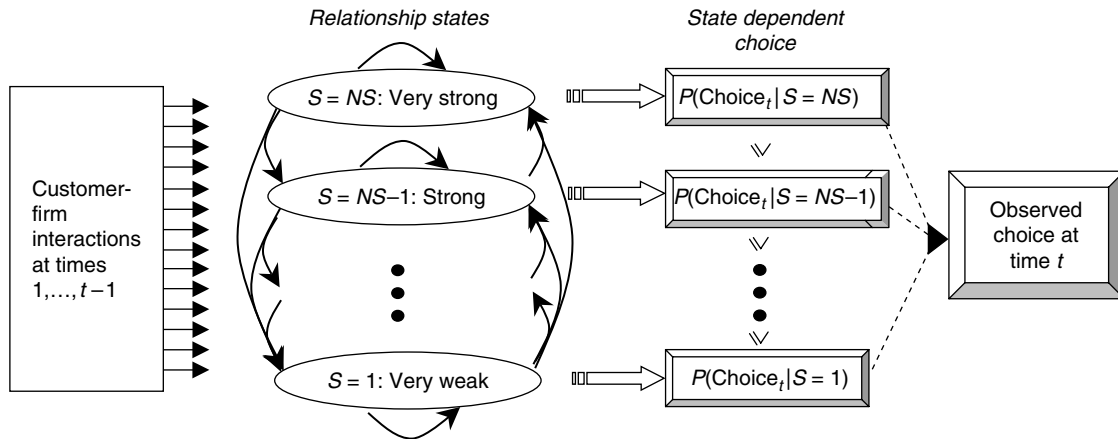
The proposed HMM consists of three main components:

(1) *The initial state distribution*—the probability that customer i is in state s at time 1 is $P(S_{i1} = s) = \pi_{is}$.

(2) *The transitions*—a sequence of Markovian *transitions* ($\mathbf{Q}_{i,t-1 \rightarrow t}$) that express, in a probabilistic manner, the likelihood that the series of customer-firm interactions in the previous time period were strong enough to transition the customer to another state. The probability that a customer transitions from state s_{t-1} at time $t - 1$ to state s_t at time t is $P(S_{it} = s' | S_{it-1} = s) = q_{its'}$.

² To keep this manuscript in a manageable length, we provide only a brief description of HMMs. The interested reader is referred to Rabiner (1989) and MacDonald and Zucchini (1997) for a detailed treatment of the topic.

Figure 1 A Hidden Markov Model of Customer Relationships



(3) *The state dependent choice*—the probability that the customer will choose the product at time t conditioned on her state is $P(Y_{it} = 1 | S_{it} = s) = m_{it|s}$,³ where S_{it} is the state of customer i at time t in a Markov process with NS states, and Y_{it} is the choice made by customer i at time t .

3.2. The Model's Components

3.2.1. The Markov Chain Transition Matrix. We model the transitions between states as a Markov process. The transition matrix is defined as

$$\mathbf{Q}_{i,t-1 \rightarrow t} = \begin{matrix} & \text{State at } t \\ \text{State at } t-1 & \begin{matrix} 1 & 2 & 3 & \cdots & NS-1 & NS \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ NS \end{matrix} & \begin{bmatrix} q_{i11} & q_{i12} & q_{i13} & \cdots & q_{i1NS-1} & q_{i1NS} \\ q_{i21} & q_{i22} & q_{i23} & \cdots & q_{i2NS-1} & q_{i2NS} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{iNS1} & q_{iNS2} & q_{iNS3} & \cdots & q_{iNSNS-1} & q_{iNSNS} \end{bmatrix} \end{matrix}, \quad (1)$$

where $q_{itss'} = P(S_{it} = s' | S_{it-1} = s)$ is the conditional probability that individual i moves from state s at time $t-1$ to state s' at time t , and where $0 \leq q_{itss'} \leq 1 \forall s, s'$, and $\sum_{s'} q_{itss'} = 1$. In applying our model in the context of alumni-university relationships (see §4), we put a structure on the general transition matrix in Equation (1). Specifically, we define the transition matrix as a random walk with a “sudden death,” whereas from each state the customer/alumni could move to an adjacent state or drop immediately to dormancy. This assumption was made primarily for model parsimony. Nevertheless, in the context of

alumni donations used in our empirical application, we found this assumption to be both behaviorally and empirically grounded.

Each one of the matrix elements in Equation (1) represents a probability of transition. We assume that the customer's propensity for transition is affected by his/her relationship encounters with the firm. We model the transitions between the states as a threshold model, where a discrete transition occurs if the propensity for transition passes a threshold level. As mentioned previously, the idea that a movement to a discrete level of relationship occurs when the aggregate measure of satisfaction or dissatisfaction from relationship encounters passes a threshold has roots in the relationship and service marketing literature (Oliver 1997).

The norm theory of Kahneman and Miller (1986) postulates that past experiences create a norm, against which current experiences are judged. Thus, it might be reasonable to expect that current relationship encounters be judged relative to the status quo. If the cumulative experience from the encounters between the customer and the firm is highly negative (e.g., service failure), it is likely to shift the propensity for transition below the threshold needed for a transition to a lower state. On the other hand, if the encounter is highly positive (e.g., an important product benefit learned from an advertisement campaign), it is likely to shift the propensity for transition above the threshold needed for a transition to a higher state. If the relationship encounters in the previous period did not have a strong impact on the customer, the customer is likely to stay in her current state.

Putting this process in mathematical terms, and assuming that the unobserved part of the propensity for transition is independently and identically distributed (IID) of the extreme value type, we can model the nonhomogeneous transition probabilities following the ordered logit model (Greene 1997).

³In what follows, we assume no information is available about the competition. We take this approach because this is generally the case in CRM transaction data sets, where the firm does not observe transactions with the competition. Nevertheless, if information about the competition exists, one could extend the model to incorporate competitive information.

Specifically, the terms $q_{itss'}$ in the transition matrix in Equation (1) could be written as

$$q_{its1} = \Pr(\text{transition from } s \text{ to state 1}) \\ = \frac{\exp(\mu(1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}{1 + \exp(\mu(1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}, \quad (2)$$

$$q_{itss'} = \Pr(\text{transition from } s \text{ to } s') \\ = \frac{\exp(\mu(s')_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}{1 + \exp(\mu(s')_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})} \\ - \frac{\exp(\mu(s' - 1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}{1 + \exp(\mu(s' - 1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}, \quad (3)$$

$$q_{itsNS} = \Pr(\text{transition from } s \text{ to } NS) \\ = 1 - \frac{\exp(\mu(NS - 1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}{1 + \exp(\mu(NS - 1)_{is} - \mathbf{a}'_{it}\boldsymbol{\rho}_{is})}, \quad (4)$$

for $s \in \{1, \dots, NS\}$ and $s' \in \{2, \dots, NS - 1\}$, where

$\boldsymbol{\rho}_{is}$ is a vector of parameters capturing the effect of relationship encounters of individual i on the propensity for transition from state s ,

\mathbf{a}_{it} is a vector of time-varying covariates for individual i between time $t - 1$ and time t , and

$(\mu(s')_{is})$ is the s' ordered logit threshold for individual i in state s , where $s' \in \{1, \dots, NS - 1\}$.

Note that the marginal effects of the time-varying covariates in the transition matrix are state specific, thus allowing for different impacts of the time-varying covariates depending on the customer's state.

3.2.2. The Initial State Distribution. For an HMM with time homogeneous transition matrix, the initial state distribution is commonly defined as the stationary distribution of the transition matrix (MacDonald and Zucchini 1997). However, because our transition matrix is a function of time-varying covariates, we calculate the stationary distribution of the transition matrix by solving the equation $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i \bar{\mathbf{Q}}_i$, under the constraint $\sum_{s=1}^{NS} \pi_{is} = 1$, where $\bar{\mathbf{Q}}_i$ is the transition matrix with the parameter estimates following §3.2.1 and all covariates are set to their mean value across individuals and time periods.⁴ Generally, the stationarity conditions above do not guarantee

⁴ An alternative approach would be to take the stationary distribution of the transition matrix with the covariates set to zero. For the empirical application in §4, the two approaches yielded very similar results. In general, one should use the stationary distribution of the transition matrix with the covariates set to zero if the data set is not left truncated (i.e., we observe the initial interaction between the customer and the firm), and the stationary distribution of the transition matrix at the mean of the covariates otherwise.

existence or uniqueness of a stationary distribution. In our empirical application, all the estimated individual transition probabilities were strictly positive confirming that the transition matrices are aperiodic and irreducible, thus guaranteeing an existence and uniqueness of the stationary distribution.

3.2.3. The State-Dependent Choice. Given the customer's state, the customer choices are assumed to be conditionally independent. Thus, given relationship state s , we model the probability of a dichotomous choice following the well-known binary logit model,⁵

$$m_{it|s} = \frac{\exp(\tilde{\beta}_{0s} + \mathbf{x}'_{it}\boldsymbol{\beta}_s)}{1 + \exp(\tilde{\beta}_{0s} + \mathbf{x}'_{it}\boldsymbol{\beta}_s)}; \quad s = 1, \dots, NS, \quad (5)$$

where

$\tilde{\beta}_{0s}$ is the state-specific coefficient for state s ,

\mathbf{x}_{it} is a vector of time-varying covariates associated with the choice of individual i at time t , and

$\boldsymbol{\beta}_s$ is a vector of state-specific response coefficients.

The full vector of conditional choice probabilities is $\mathbf{m}'_{it} = [m_{it|1}, m_{it|2}, \dots, m_{it|NS}]$.

The difference between the vectors of covariates (\mathbf{a}_{it}) to be included in the transition matrix and in the vector of covariates in state-dependent choice (\mathbf{x}_{it}) is noteworthy. The conceptual distinction is between those variables that have an enduring impact on the attitude of the customer toward the product or service and those that affect only the short-term choice behavior. For example, advertising is often assumed to have a long-term impact on attitude while a price promotion affects only the short-term choice behavior. In the transition matrix, one should include covariates that are hypothesized to have an enduring impact on the customer's buying behavior (e.g., advertisement, service encounters, or relationship-based marketing activities). On the other hand, the covariates in the state-dependent choice vector are assumed to primarily have an immediate effect on the customer (e.g., price and display promotions).⁶ The researcher could also test empirically (e.g., using fit measures) the appropriate location for each covariate in the model.

To ensure identification of the states, we restrict the choice probabilities to be nondecreasing in the relationship states. Because both the intercepts and the

⁵ In the context of alumni university donations, we model the state-dependent and dichotomous choices rather than donation amounts because we believe that the act of choice is a stronger determinant of relationship strength than quantity measures. Nevertheless, the model could be extended to capture quantity or amount outcomes using Poisson or Tobit models.

⁶ The covariate in the state-dependent choice vector could still have an indirect long-term effect through the longitudinal effect of the current choice on future buying behavior.

response parameters are state-specific, we impose this restriction at the mean of the vector of covariates, \mathbf{x}_{it} . Thus, the vector \mathbf{x}_{it} is mean-centered and the restriction $\tilde{\beta}_{01} \leq \tilde{\beta}_{02} \leq \dots \leq \tilde{\beta}_{0NS}$ is imposed by

$$\tilde{\beta}_{0s} = \beta_{01} + \sum_{s'=2}^s \exp(\beta_{0s'}); \quad s = 2, \dots, NS. \quad (6)$$

3.2.4. Accounting for Heterogeneity. Because our model is dynamic, one must ensure that the zero-order heterogeneity is fully accounted for to distinguish it from time dynamics. Heckman (1981) suggests that ignoring heterogeneity might lead to a strong spurious state dependence, even when the actual choices are not correlated over time. Similarly, a model that accounts for heterogeneity but ignores state dependence may overestimate the degree of heterogeneity (Keane 1997). Our proposed HMM addresses the second problem by offering a flexible specification of state dependence. To distinguish between heterogeneity and dynamics, we define random-effect parameters in the transition matrix $\mathbf{Q}_{i,t-1 \rightarrow t}$. We incorporate heterogeneity by allowing the threshold parameters ($\mu(s')_{is}$) in Equations (2)–(4) to vary across individuals.⁷ This specification allows for heterogeneity in the “stickiness” to different states, since the distance between the low and high thresholds could vary across individuals. Heterogeneity in the transition matrix also implies heterogeneity in the initial state distribution. An alternative heterogeneity specification would be to allow the state-dependent choice to vary across individuals. However, from a managerial point of view, such heterogeneity specification implies individual-specific state interpretation, thus losing the ability to classify the customers into a common set of states.

3.3. The Likelihood of an Observed Sequence of Choices

Due to the Markovian structure of the model, the individual choice probabilities are correlated through the common underlying path of the hidden states. Accordingly, the joint likelihood of a sequence of choices is given by the sum over all possible routes the individual could take over time between the underlying states:

$$P_i(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \sum_{s_1=1}^{NS} \sum_{s_2=1}^{NS} \dots \sum_{s_T=1}^{NS} \left[P(S_{i1} = s_1) \prod_{\tau=2}^T P(S_{i\tau} = s_\tau | S_{i\tau-1} = s_{\tau-1}) \cdot \prod_{\tau=1}^T P(Y_{i\tau} = y_{i\tau} | S_{i\tau} = s_\tau) \right]. \quad (7)$$

⁷ To keep the model parsimonious we avoid a full random-effect specification of the transition matrix $\mathbf{Q}_{i,t-1 \rightarrow t}$.

Using our notations for the three components of the HMM in Equations (1)–(5), we can rewrite Equation (7) as

$$P_i(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \sum_{s_1=1}^{NS} \sum_{s_2=1}^{NS} \dots \sum_{s_T=1}^{NS} \left[\pi_{is_1} \prod_{\tau=2}^T q_{is_\tau s_{\tau-1}} \cdot \prod_{\tau=1}^T m_{i\tau|s_\tau}^{y_{i\tau}} (1 - m_{i\tau|s_\tau})^{(1-y_{i\tau})} \right]. \quad (7a)$$

A problem with Equation (7a) is that it has NS^T elements and is therefore computationally intractable for even modest values of T . Following MacDonald and Zucchini (1997), we can rewrite Equation (7a) in a matrix products form that simplifies computation:

$$L_{iT} = P(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \boldsymbol{\pi}_i \tilde{\mathbf{m}}_{i1} \mathbf{Q}_{i,1 \rightarrow 2} \tilde{\mathbf{m}}_{i2} \dots \mathbf{Q}_{i,T-1 \rightarrow T} \tilde{\mathbf{m}}_{iT} \mathbf{1}', \quad (8)$$

where

$\tilde{m}_{it|s} = m_{it|s}^{y_{it}} (1 - m_{it|s})^{(1-y_{it})}$ and $\tilde{\mathbf{m}}_{it}$ is a $NS \times NS$ diagonal matrix with the elements of $\tilde{m}_{it|s}$ on the diagonal, and $\mathbf{1}'$ is a $NS \times 1$ vector of ones. To avoid underflow of the likelihood function in Equation (8) we divide the joint state likelihood after every time period by L_{it}/NS , accumulate the logarithms of these scale factors, and add it to the logarithm of the likelihood function (for details see MacDonald and Zucchini 1997, p. 79). The scaled log-likelihood function across individuals is simply the sum of the individual scaled log-likelihood over $i \in \{1, \dots, N\}$.

3.4. Estimation Procedure

In this section, we describe the procedure used to estimate our model. In choosing the estimation procedure, we focus on properly accounting for observed and unobserved heterogeneity. We estimate the HMM parameters: the transition matrix parameters and the state-dependent choice parameters described in Equations (2)–(6) using the joint likelihood function in Equations (7)–(8).

We estimate our HMM using a standard hierarchical Bayes estimation procedure (Rossi and Allenby 2003) using two sets of parameters: random-effect parameters ($\boldsymbol{\theta}_i$) and parameters that are common across individuals ($\boldsymbol{\Psi}$). We define $\boldsymbol{\theta}_i = \{\mu(s')_{is1}, \dots, \mu(s')_{iNS-1}, s' \in \{1, \dots, NS\}\}$ and $\boldsymbol{\Psi} = \{\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots, \boldsymbol{\rho}_{NS}, \beta_{01}, \beta_{02}, \dots, \beta_{0NS}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{NS}\}$. Heterogeneity is introduced into the model for the random-effect parameters by defining uninformative priors $\boldsymbol{\theta}_i \sim MVN(\bar{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}})$. We also estimate a model in which the observed individual characteristics, such as demographics, are introduced into the model in a hierarchical manner (e.g., Allenby and Ginter 1995). We complete the specification by assuming appropriate

and diffuse priors on Σ_0 and Ψ . We sequentially draw from the set of conditional posterior distributions.⁸ The conditional posterior distributions of θ_i and ψ do not have a closed form. Thus, we use the Metropolis-Hasting algorithm to draw from these posterior distributions. To reduce the degree of autocorrelation between draws of the Metropolis-Hasting algorithm and to improve the mixing of the MCMC we use an adaptive Metropolis adjusted Langevin algorithm (Atchadé 2006). We found this recent approach to be very useful in our HMM application.

It should be noted that unlike the HMM Bayesian estimation methods that augment the latent state memberships (Djuric and Chun 2002, Kim and Nelson 1999, Moon et al. 2007, Scott 2002), we use a Bayesian approach to estimate directly the likelihood function in Equation (8) with random-effect parameters.

3.5. Recovering the State Membership Distribution

An attractive feature of the HMM is the ability to use it to probabilistically recover the individual's state at any given time period. This measure could be directly derived from the likelihood function in Equation (8). Two approaches have been suggested for recovering the state membership distribution: "filtering" and "smoothing" (see Hamilton 1989 for a discussion). Filtering uses only the information known up to time t to recover the individual's state at time t , whereas smoothing uses the full information available in the data. The filtering approach is more appealing for marketing applications, where decisions are made based only on the history of the observed behavior. The filtering probability that individual i is in state s at time t conditioned on the individual's history of choices is given by

$$P(S_{it} = s | Y_{i1}, Y_{i2}, \dots, Y_{it}) \\ = \pi_i \tilde{m}_{i1} \mathbf{Q}_{i,1 \rightarrow 2} \tilde{m}_{i2} \dots \mathbf{Q}_{i,t-1 \rightarrow t,s} \tilde{m}_{it|s} / L_{it}, \quad (9)$$

where

$\mathbf{Q}_{i,t-1 \rightarrow t,s}$ is the s th column of the transition matrix $\mathbf{Q}_{i,t-1 \rightarrow t}$ and
 L_{it} is the likelihood of the observed sequence of choices up to time t from Equation (8).

4. Empirical Application

In this section, we describe the empirical application of the proposed HMM in the context of the relationship between alumni and their alma mater. We describe, in order, the data set, the alternative models estimated, the estimation results, the comparison

of the alternative models in terms of their prediction ability on a hold-out sample, and a survey-based analysis of the behavioral dimensions underlying the alumni-university relationship states.

4.1. Application to Alumni Relations

To empirically illustrate the ability of the proposed model to capture the dynamics in customer relationships and choice behavior, we apply the proposed HMM in the context of university-alumni relations and gift-giving (i.e., donation) behavior. The objective of the empirical application is to show how one can use observed alumni gift-giving data to (1) dynamically classify the alumni into relationship strength states, (2) understand the factors that influence the dynamics in gift-giving behavior, (i.e., assess what alumni-university interactions are likely to move alumni to a higher relationship state), and (3) predict future gift-giving behavior.

There are several reasons for choosing the alumni gift-giving data as an empirical application for our model. First, we believe that the dynamics in the gift-giving behavior, due to the strong relationship underlying its construct, is stronger than the dynamics found in typical scanner panel data. Second, this data set contains most of the components suggested for a good CRM data set. Specifically, it includes four out of the five elements suggested by Winer (2001): transaction data, customer contacts, descriptive information about the customers, and longitudinal data. As discussed further below, this data set is somewhat limited in terms of tracking exposure and response to marketing activities. As a result, we use other time-varying covariates that may affect the alumni state such as reunion attendance. Finally, in the sagging economy, private and public schools face severe financial problems. Charitable contributions to universities dropped in 2002 for the first time in 15 years (*New York Times* 2003). Therefore, addressing the problem of managing the \$24 billion market of U.S. alumni fundraising is of significant financial consequence.

Previous research on alumni gift-giving behavior is limited. The few articles that have been published in this area investigated the effect on alumni gift giving of (1) institutional characteristics (Baade and Sundberg 1996, Harrison et al. 1995), (2) reunions (Willemain et al. 1994), and (3) individual characteristics such as demographics, financial aid, and participation in college athletics (Okunade 1993, Okunade and Berl 1997, Taylor and Martin 1995). Recently, in the marketing literature, Arnett et al. (2003) related alumni-university relations to the behavioral identity salience model. To our knowledge, no study has previously investigated and modeled the dynamics of gift-giving behavior and the factors that can alter this dynamic behavior.

⁸ The complete set of conditional distributions and the iteration sequence of the MCMC simulation appear in Appendix A.

4.2. Data Description

The data used to calibrate and validate the model are sampled from the database provided by the alumni association of a large west coast university. Our data set consists of over 17,000 randomly sampled alumni. This represents 10% of the total university alumni base (see Appendix B in Netzer 2004 for a detailed description of the data set). From this data set, we use in our analysis 1,256 alumni who graduated with an undergraduate degree⁹ between 1966 and 1988 and donated at least once in the first 10 years following graduation (see Table 1 for descriptive statistics of the sample).

For each alumna/alumnus the data provide cumulative information on total gift giving since the time of graduation, as well as detailed disaggregate data about his/her gift giving since 1976 (or time of graduation, whichever is more recent). The data set also contains disaggregate information about different alumni-university interactions for the years 1976–2001, including participation in university events and volunteering for alumni roles. We use the observations in the years 1976–1998 to calibrate the model, and the last three years of possible gift giving for each alumna/alumnus (1999–2001) for validation.

A necessary condition for the identification of the proposed HMM is dynamics in donation behavior over time. To examine whether such dynamics exist in our data we used the Run Test (Frank 1962). The Run Test strongly supports the existence of individual-level dynamics in the donation behavior, in particular it suggest that alumni go through periods of donation and nondonation (for the full description of the Run Test results see Netzer 2004).

4.3. Variables Description

The variables of this data set can be divided into three categories:

1. *Alumni-university interactions*—This set of variables defines the interactions between the alumni and the university in the vector \mathbf{a}_{it} in Equations (2)–(4). These are recorded over time post graduation. In this study, we consider two types of interactions (besides donations): reunion attendance and volunteering for a university role. It should be noted that our modeling approach assumes that these alumni-university interactions occur, and are observed, prior to donation. This assumption seems to be valid for these two variables, because donations generally occur post the reunion attendance and volunteering decisions.

2. *Influence attempts*—These alumni-university interactions are postulated to have mainly short-term

Table 1 Descriptive Statistics of the Calibration and Hold-Out Sample

Key characteristics	Percentage/Average
Overall observations (gift opportunities)	27,526
Overall number of alumni	1,256
Mean observations per individual	21.9
Proportion of donation years	43%
Mean donation	\$478
Q25 yearly donation	\$30
Median yearly donation	\$100
Q75 yearly donation	\$200
Gender	
Female	40%
Male	60%
School	
Earth	3%
Humanities	79%
Engineering	15%
Other (e.g., Undeclared)	3%
Degree	
Undergraduate	73%
Undergraduate + Graduate	27%
Spouse is university alumna/alumnus	
Yes	56%
No	44%
Alumni association membership	
Member	69%
Not member	31%

effects. These covariates determine the vector \mathbf{x}_{it} in Equation (5). In the scanner data context, price promotion and display are examples of activities with short-term influence. In the current application, we use reunion year (any multiple of five years since graduation) as such a covariate. We consider a reunion year as a short-term influence to donate due to increased salience of the university during that year, which might raise the likelihood of giving in the specific year but not in subsequent years. However, it is actual participation in a reunion that might lead to a change in the strength of relationship between the alumna/alumnus and the university and to a long-term impact on subsequent donation behavior. Accordingly, reunion year is included in the state-dependent choice (the vector \mathbf{x}_{it}), and participation in a reunion in the transition matrix (the vector \mathbf{a}_{it}).

3. *Choice behavior*—Gift-giving behavior. This is the dependent variable, which is captured by the incidence of donation (0 or 1). If one is interested in merely predicting donations, this variable could be replaced by the actual amount donated using an ordered logit or a Tobit model. However, in the context of relationship dynamics model, the alumni association considers the actual act of giving as a stronger determinant of relationship than the amount given.

According to the alumni association, we aggregated gift-giving and events data by calendar year. To account for time dynamics that are not related to the

⁹ Following the alumni association’s recommendation, we sampled only alumni who received their undergraduate degree (possibly followed by higher degrees) from the university.

relational marketing encounters, a linear time trend, representing the number of years since graduation, is included in the state-dependent vector. Additionally, the data set contains alumni characteristics such as year and major of graduation, gender, marriage to a university alumna/alumnus, and membership in the alumni association.

4.4. Estimated Models

We estimated the HMM of customer relationship dynamics described in §3, using the MCMC hierarchical Bayes estimation described above. Additionally, we estimated a restricted version of this model, with no heterogeneity, as well as a nondynamic benchmark model and the state dependent recency-frequency model commonly used in relationship marketing.

Model 1—This is the full HMM.¹⁰

Model 2—HMM with no heterogeneity. This model is similar to Model 1, but with common parameter across individuals. A limitation of this model is that this model cannot distinguish between heterogeneity and dynamics in the gift-giving behavior.

Model 3—Nondynamic model. This is the latent class model (Kamakura and Russell 1989) with the same number of latent states as in Model 1. This model does not allow individuals to move between the states over time.

Model 4—Recency-frequency model. This model is frequently used in relationship management applications (e.g., Bult and Wansbeek 1995). In this model, the recency since the last donation and frequency of donations up to time t enter as covariates in the model.¹¹ In the recency-frequency model, the probability of choice follows the binary logit formulation:

$$P(Y_{it} = 1) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta}_i)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta}_i)}, \quad (10)$$

where

$\boldsymbol{\beta}_i$ is a vector of random-effect parameters, and \mathbf{x}_{it} is a vector of time-varying covariates.

The vector \mathbf{x}_{it} includes all the covariates used in the HMM (reunion years, reunion participation, volunteering to university roles and years since graduation) as well as recency since last donation and donation frequency. We incorporate heterogeneity in this model by estimating a random-effect intercept and the recency-frequency parameters using hierarchical Bayes estimation.

¹⁰ In addition to the full HMM, we also estimated a hierarchical HMM in which the random-effect coefficients are a function of individual characteristics (see Table 6).

¹¹ We do not model the monetary donation amounts to be consistent with the proposed HMM. Moreover, adding the running average of donation amounts as a predictor did not significantly improve the model's fit or prediction for donation incidents.

The distinction between the HMMs and the recency-frequency model is noteworthy. The recency-frequency model accounts for dynamics in an ad-hoc fashion in which the structure of the effect of past choices on the current choice is defined *a priori* through the recency and frequency terms. The HMM offers a more behaviorally structured approach for state dependence through a Markovian transition between a set of relationship states. Furthermore, although some structure is imposed on the transitions between the states, the choice of the number of states allows us to determine the structure of dynamics based on the complexity of the dynamics of the data at hand.

4.5. Estimation Results

Models 2 and 3 were estimated using the maximum likelihood procedure MAXLIK in the GAUSS statistical software. Models 1 and 4 were estimated using a MCMC hierarchical Bayes procedure, using the Gibbs sampler and the Metropolis-Hastings algorithm coded in GAUSS. In the hierarchical Bayes estimation, the first 90,000 iterations were used as a “burn-in” period, and the last 10,000 iterations were used to estimate the conditional posterior distributions and moments. To assess convergence of the MCMC we adopt the method proposed by Gelman and Rubin (1992), which compares the within to between variance for each parameter estimated across multiple chains. Across three parallel chains, the scale reduction estimate for all the parameters estimated is lower than 1.2, suggesting that convergence has been achieved.

4.5.1. Selecting the Number of States. The first stage in estimating the HMM is selecting the number of states. Model selection measures can be used to choose the number of states. Due to the sensitivity of some Bayesian model selection criteria, such as the marginal log-likelihood and Bayes factor, to the specified priors (Rossi and Allenby 2003), we compare alternative model selection criteria. Specifically, we contrast the Bayes factor with the log-marginal density,¹² the marginal validation log-likelihood measure (Andrews and Currim 2003), the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002), and the Markov switching criterion (MSC), recently developed for HMMs' states and variables selection by Smith et al. (2005).¹³ The validation log-likelihood is calculated following Equation (8), using

¹² The Schwarz Bayesian information criterion (BIC; Schwarz 1978) commonly used for model selection in classic statistical applications, asymptotically approximates the Bayesian posterior marginal density.

¹³ The MSC was originally developed for a stationary, aggregate data HMM. We adapted this criterion to our random-effect nonstationary HMM. The details of the modified MSC appear in Appendix B.

Table 2 Choosing the Number of States

Number of states	-2 Marginal log-density	Log Bayes factor	DIC	MSC	Validation -2 log-likelihood
1	26,708.5	—	26,493.0	50,476.5	4,768.5
2	25,557.1	575.7	25,378.8	48,806.6	3,689.1
3	25,397.1	80.0	25,176.3	48,332.8	3,583.6
4	25,689.1	-146.2	25,658.2	48,639.5	3,717.5

the hold-out sample (the years 1999–2001 in the data set). The log marginal density and Bayes Factor are calculated using a harmonic mean of the individual likelihoods across iterations (Newton and Raftery 1994) from the output of the Metropolis-Hasting sampler.

Based on all measures, the best-fitting model is the model with three states. This model minimizes the -2 log-marginal density, validation -2 log-likelihood, DIC and MSC, and shows a favorable Bayes factor¹⁴ (see Table 2).

4.5.2. HMM Estimates. Table 3 reports the posterior means and posterior standard deviations (parameter estimates and standard errors for Model 2) of the two variants of the HMM based on the calibration sample. In the heterogeneous HMM, the parameters that capture the effect of reunion attendance, reunion participation and volunteering are all positive as expected. With the exception of state 1, likelihood of donation is increasing with years since graduation. To get a better understanding of the magnitude of the parameter estimates in Table 3 we plug these parameters into Equations (2)–(6) to get the state dependent choice and transition probabilities.

The interpretation of the three states is primarily determined by the state-specific intrinsic propensity to donate (the parameters β_{01} , β_{02} , and β_{03}). At the mean of the covariates “years since graduation” and “reunion year,” the conditional probability of donation given state 1 is 6%, given state 2 it is 46%, and given state 3 it is 100%. Accordingly, we label these three states as “dormant,” “occasional,” and “active” states, respectively.

An interesting feature of our model is the ability to investigate the effects of time-varying covariates on the transitions between the states. Specifically, we compare the effect of reunion attendance and volunteering to university roles. The middle and right matrices in Table 4 demonstrate the average effect (across alumni) of reunion attendance and volunteering on the state transitions. Attending a reunion increases the likelihood that an alumna/alumnus in the dormant state moves to the occasional state from

Table 3 Estimation Results for the Hidden Markov Models

Parameter	Model 1 HMM with heterogeneity*	Model 2 HMM no heterogeneity
β_{01} state dependent intercept (State 1)	-2.778 (0.297)	-2.659 (0.092)
β_{02} state dependent intercept (State 2)	0.956 (0.244)	1.076 (0.038)
β_{03} state dependent intercept (State 3)	2.388 (0.432)	2.578 (0.225)
Reunion year (State 1)	0.646 (0.375)	0.592 (0.128)
Reunion year (State 2)	0.017 (0.300)	0.048 (0.070)
Reunion year (State 3)	0.359 (1.397)	-0.950 (0.128)
Years since graduation (State 1)	-0.071 (0.099)	-0.056 (0.009)
Years since graduation (State 2)	0.053 (0.099)	0.062 (0.008)
Years since graduation (State 3)	0.992 (0.458)	1.247 (0.287)
$\mu(hi)_1$ high threshold (State 1)	2.159 (0.132)	2.507 (0.078)
$\mu(lo)_2$ low threshold (State 2)	-1.807 (0.177)	-2.141 (0.078)
$\mu(hi)_2$ high threshold (State 2)	1.005 (0.140)	1.500 (0.037)
$\mu(lo)_3$ low threshold (State 3)	-3.461 (0.129)	-9.490 (8.214)
$\mu(hi)_3$ high threshold (State 3)	1.001 (0.137)	2.085 (1.021)
$V(\mu(hi)_1)^{**}$	0.200 (0.080)	—
$V(\mu(lo)_2)$	0.672 (0.227)	—
$V(\mu(hi)_2)$	0.232 (0.076)	—
$V(\mu(lo)_3)$	0.168 (0.103)	—
$V(\mu(hi)_3)$	0.212 (0.054)	—
Volunteering (State 1)	1.604 (0.523)	1.479 (0.230)
Volunteering (State 2)	0.092 (0.534)	-0.049 (0.190)
Volunteering (State 3)	0.969 (0.517)	1.080 (0.260)
Reunion participation (State 1)	2.895 (0.621)	2.526 (0.263)
Reunion participation (State 2)	1.044 (0.734)	0.959 (0.959)
Reunion participation (State 3)	0.779 (0.861)	-0.181 (0.677)
-2 log-marginal density/-2 log likelihood	25,397.1	25,565.9

*Numbers in parentheses are posterior standard deviations for Model 1 and standard errors for Model 2.

** $V(\cdot)$ refers to the posterior standard deviation across individuals.

10% to 68%. Reunion attendance also has a strong impact on moving alumni from the occasional to the active state; it increases this likelihood from 28% to 53%. More importantly, reunion attendance decreases the likelihood of dropping from the occasional state to the “sticky” dormant state from 14% to only 5%. In contrast, the effect of reunion participation on keeping alumni in the active state is moderate (i.e., from 68% to 82%). Volunteering to university roles, on the other hand, has its primary impact on alumni in the dormant and active states, whereas the effect on alumni in the occasional state is minimal. While targeting the most active customers is consistent with the customer lifetime value approach and with the practice of rewarding “loyal” customers, the result that reunion attendance has the highest impact on alumni in the dormant and occasional state is consistent with the theory of intermittent reinforcement. Due to the probabilistic nature of the transitions, active alumni are likely to transition into the occasional state once in

¹⁴ The log Bayes factor compares the model with NS states to the model with $NS - 1$ states. A log Bayes factor larger than five suggests strong evidence in favor of the model with NS states.

Table 4 The Mean Posterior Transition Matrices*

$t - 1$	No interactions			Reunion attendance			Volunteering		
	t			t			t		
	Dormant	Occasional	Active	Dormant	Occasional	Active	Dormant	Occasional	Active
Dormant	90% [89%–90%]	10% [10%–11%]	0% [— — —]	32% [17%–48%]	68% [52%–82%]	0% [— — —]	64% [49%–74%]	36% [25%–51%]	0% [— — —]
Occasional	14% [14%–15%]	58% [56%–59%]	28% [27%–30%]	5% [1%–12%]	42% [17%–56%]	53% [31%–80%]	13% [7%–21%]	57% [47%–60%]	30% [20%–45%]
Active	3% [3%–3%]	29% [28%–31%]	68% [66%–69%]	1% [0%–4%]	17% [2%–35%]	82% [61%–97%]	1% [0%–2%]	14% [9%–21%]	85% [77%–90%]

*95% confidence interval in parenthesis.

several time periods and therefore should be affected by reunion attendance.

It is important to note that the results presented in Table 4 are the mean of the posterior distribution across individuals. Inference from our estimation could and should be derived at the individual level in a similar manner.

The time-varying covariates reunion attendance and volunteering are endogenous to the alumna/alumnus, thus one should be careful about treating these as decision variables in the context of our data set. Nevertheless, given the observed behavior, one could still use these alumni-university interactions to dynamically segment alumni to the relationship states. Table 4 demonstrates the value of adding time-varying covariates such as customer initiated interaction or marketing interventions. Future research could explore the application of our model to data sets in which the time-varying covariates are also decision variables.

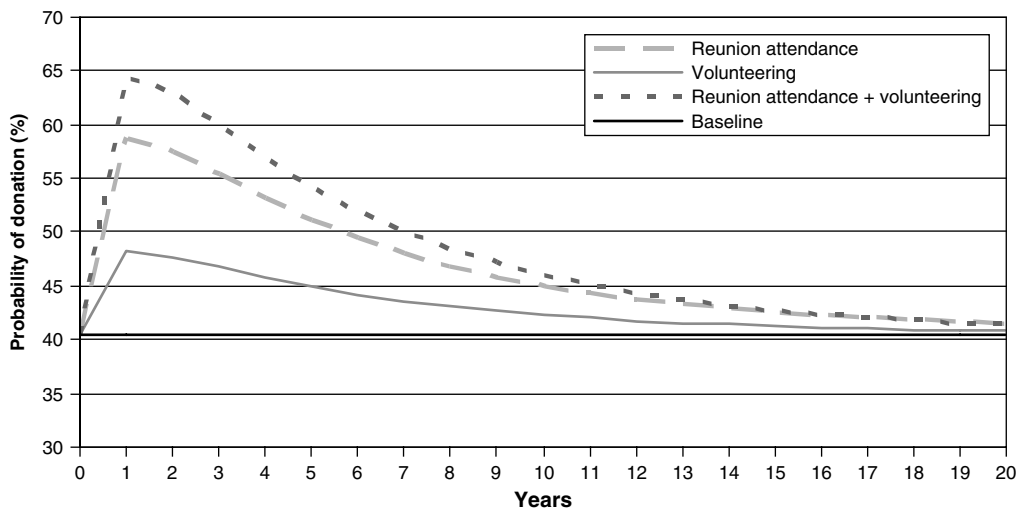
The transition probabilities in Table 4 suggest that reunion attendance might have an enduring impact on the donation behavior of alumni in the occasional

state, by moving them away from the “sticky” dormant state and by increasing the likelihood of a transition to the relatively “sticky” active state. We use the posterior mean of the transition matrix parameters to gain better understanding of the long-term effect of the time-varying covariates in the transition matrix. Specifically, we calculate the effect of reunion attendance and volunteering over the course of 20 years for an alumna/alumnus with an average initial state distribution at time 0 (average donation rate of 41%) and attended a reunion and/or volunteered to a university role at year 1.

As is evident from Figure 2, both time-varying covariates have a long-term impact on the propensity to donate. The immediate as well as long-term effects of reunion attendance are stronger than the effects of volunteering to a university role. Five years after the reunion attendance or volunteering year, approximately 50% of the effect of these events still carries over.

One interesting product of the transition matrices in Table 4 is the stationary distribution of the transition matrix at the mean of the covariates, which is

Figure 2 The Long-Term Effects of the Time-Varying Covariates



also our initial state distribution. The stationary distribution of the alumni in the three states is 46%, 29%, and 25% in the dormant, occasional, and active states, respectively.

In the HMM, following each choice, the model updates the state membership distribution. This allows estimating the effect of a donation at time t on the probability of being in state s at time t and therefore on the probability of donation at time $t + 1$ (see Table 5). Using Equation (11) we calculated the state membership probability at time t following a donation at time t .

$$P(S_{it} = s | Y_{it} = 1, \mathbf{S}_{it-1}) = (\Pr(\mathbf{S}_{it-1}) \mathbf{Q}_{i,t-1 \rightarrow t,s} m_{it|s}) / (\Pr(\mathbf{S}_{it-1}) \mathbf{Q}_{i,t-1 \rightarrow t} \mathbf{m}_{it} \mathbf{1}'), \quad (11)$$

where

$\Pr(\mathbf{S}_{it-1})$ is individual i 's state membership distribution at time $t - 1$,

$\mathbf{Q}_{i,t-1 \rightarrow t,s}$ is the s th column of the transition matrix $\mathbf{Q}_{i,t-1 \rightarrow t}$ and

\mathbf{m}_{it} is a $NS \times NS$ diagonal matrix with the elements of $m_{it|s}$ on the diagonal.

As expected, donations have a very strong impact on the transition probabilities.

The transition matrices in Tables 4 and 5 present the mean of the posterior distribution across alumni. Next, we describe the heterogeneity in these matrices across alumni.

4.5.3. Posterior Distributions and Observed Heterogeneity. In our HMM, alumni can differ with respect to their propensity to switch between states due to the random-effect transition threshold parameters. Figure 3 depicts the distribution of the alumni's posterior propensity to stay in each one of the states given in that state in the previous period (the diagonal elements of the first transition matrix in Table 4).

Table 5 Mean Posterior Transition Following a Donation at Time t

$t - 1$	Donation at t		
	t		
	Dormant	Occasional	Active
Dormant	53% [47%–58%]	47% [42%–53%]	0% [—]
Occasional	1% [1%–2%]	48% [43%–51%]	51% [47%–55%]
Active	0% [0%–0%]	16% [14%–18%]	84% [71%–86%]

Not only is the dormant state the most “sticky” on average, as suggested by Table 4, but also alumni in this state are most homogeneous in terms of their likelihood of staying in this state. On the other hand, alumni in the occasional and active states are relatively heterogeneous in terms of their propensity to stay in these states.

We relate the random-effect parameters to observed heterogeneity using a hierarchical Bayes structure (see Appendix A for details). Table 6 presents the parameter estimates of the observed heterogeneity covariates.

Several individual characteristics are significantly related the propensity for transition between the states. Specifically, membership in the alumni association increases the likelihood of transitioning up from the dormant and occasional states. Females and alumni families in which both household members are alumni tend to avoid falling into dormancy from the occasional state, but given in an active state tend either to stay active or drop to dormancy. Furthermore, graduates of the earth sciences, humanities and engineering schools are less likely to transition to the active state relative to the other majors.

Figure 3 Posterior Distribution of the Propensity to Stay in the Three States

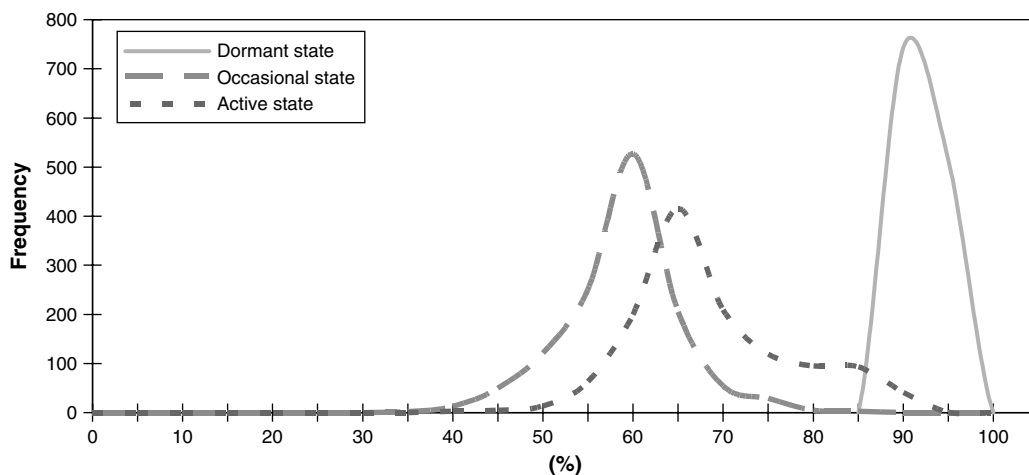


Table 6 Mean Posteriors for Observed Heterogeneity Parameters

Variable	$\mu(hi)_1$ high threshold dormant	$\mu(lo)_2$ low threshold occasional	$\mu(hi)_2$ high threshold occasional	$\mu(0)_3$ drop to dormant threshold active	$\mu(lo)_3$ low threshold active
Intercept	3.320**	-3.622*	-1.261	-1.720**	0.603
Alumni association member	-0.466**	-0.177	-0.260**	-0.277	0.243
Spouse is university alumna/alumnus	-0.687**	-0.362**	0.109	0.413*	-0.278**
Years since graduation in 1976	-0.085**	-0.047	-0.007	-0.081**	0.018
Female	0.180	-0.523**	0.204	0.722**	-0.401**
Earth sciences major	0.030	-8.641**	3.351**	0.150	-0.102
Humanities major	-0.023	2.660	2.429**	-1.678**	0.052
Engineering major	-0.076	1.587	2.609**	-0.494	-0.567
Only undergraduate degree from the university	-0.084	-0.300	0.113	0.193	0.335**

*The 90% confidence interval does not include zero.
 **The 95% confidence interval does not include zero.

4.6. Predictive Ability

We use the hold-out data to assess the prediction ability of the HMM and compare it to the four benchmark models.¹⁵ The parameters estimated based on the calibration period are used to predict the 1,256 (alumni) × 3 (years) = 3,768 observations of possible gift giving in the validation period. We compare the prediction ability of the alternative models using the hit rate measures¹⁶ (overall hit rates as well as hits and misses of donation and nondonation periods), the root-mean-square prediction error (RMSPE) between the predicted choice probabilities and the actual choices across alumni and time periods, and the validation log-likelihood (see Table 7).

The RMSPE of each model is compared to the RMSPE of a random choice rule. The random choice rule’s RMSPE was calculated based on the aggregate donation probability in the calibration period (43.0%). The random choice rule’s RMSPE is 0.495. The prediction ability of all models is significantly better than that of a random choice rule, based on the RMSPE.

The HMM predicts the hold-out choices significantly better than the nondynamic latent class model ($z = 9.5$; p -value < 0.001) and the dynamic, observed states, recency-frequency model ($z = 5.1$; p -value < 0.001). The improvement in prediction ability of the HMM relative to the alternative models is consistent across all prediction measures used. The recency-frequency model seriously under-predicts donation years, which are arguably more important to predict than nondonation years.

While the fit of the HMM with heterogeneity (Model 1) is better than that of the HMM with no heterogeneity (Model 2), the prediction ability of the parsimonious HMM with no heterogeneity is similar

to that of the full HMM. This is not surprising because in Model 2 heterogeneity in the hold-out period is elicited from the history of choices in the calibration period even when the model’s parameters do not vary across individuals. Consequently, in Model 2 one cannot distinguish between the effect of time dynamics and cross-individual heterogeneity.

One of the major advantages of the HMM in the context of customer relationships is the ability to dynamically segment the firm’s customer base. An alternative, static segmentation approach is the latent class model (Model 3). In Table 7, the HMM segmentation provides better donation predictive validity and fit relative to the latent class model. To investigate further the difference between the dynamic and the static segmentation models, we analyzed the predictive ability of the two models for alumni for whom the two segmentation methods coincide and alumni for whom they diverge. Specifically, for the HMM we classified each alumnus to the state with the highest state membership probability in the last year of

Table 7 Fit and Predictive Ability Measures

Measure	Model			
	Model 1 HMM with heterogeneity	Model 2 HMM no heterogeneity	Model 3 nondynamic latent class	Model 4 recency- frequency model
Calibration: -2 log marginal density/ -2 log likelihood	25,397.1	25,565.8	27,399.8	26,045.1
Overall hit rate (%)	77.6	77.3	67.9	72.5
Donation hit rate (%)	78.8	78.0	67.2	67.7
Non donation hit rate (%)	76.5	76.6	68.5	76.9
RMSPE	0.392	0.393	0.448	0.426
Improvement over the random RMSPE (%)	20.7	20.6	9.4	14.0
Validation -2 log likelihood	3,583.6	3,597.0	4,484.2	4,148.6

¹⁵ The parameter estimates of the recency-frequency model, and the latent class model can be obtained from the authors.

¹⁶ To estimate hit rates, we predicted a donation if the predicted probability of donation is larger than 50%.

Table 8 Predictive Ability by Segmentation Method

Segment allocation	<i>N</i>	Latent class hit rates (%)	HMM hit rates (%)
Dormant in HMM but not in latent class	142	42.0	77.2
Dormant in both	393	82.4	81.9
Occasional in HMM but not latent class	123	55.0	62.9
Occasional in both	154	52.6	68.2
Active in HMM but not in latent class	216	58.6	76.7
Active in both	228	85.2	85.4
Total	1,256	67.9	77.6

the calibration sample. The latent class solution suggested three segments that are similar to the three HMM states in terms of donation propensity (dormant 14%, occasional 45%, active 80%). We classified alumni into these three latent class states based on the highest probability rule. Table 8 compares the predictive ability of the HMM and the latent class models to predict the hold-out donation years of alumni that were classified to the same state/segment by the two methods and alumni for which the segmentation of the two methods departed.

The predictive ability of the latent class model is similar to that of the HMM for the alumni that were classified to the dormant or active segment/state by both models. However, for alumni that were classified by the latent class model into a different segment than the one suggested by the HMM, the predictive ability of the latent class model was poor. Additionally, the HMM outperformed the latent class model for alumni in the more transient occasional state. This analysis might suggest that the HMM segmentation matches the “true” alumni segmentation better than the static latent class model.

In summary, the previous sections demonstrated the insights one could gain from including time-varying covariates in the states’ transitions. The hold-out sample analysis suggests that an additional advantage of the relationship HMM is in improving the ability to predict future choices over the static and observed state models.

4.7. Behavioral and Attitudinal Dimensions

To broaden the applicability of the HMM to most common CRM data sets, the model uses only observed behavioral measures (such as choice) to elicit the relationship states. This could raise concerns that the relationship states in the HMM represent merely differences in the likelihood of choice rather than true differences in relationship reflected by attitude. It has been suggested that repeated transactions could be transformed into “true” relational

behavior through attitudinal bonds such as commitment, identity salience, self-connection, and satisfaction (e.g., Arnett et al. 2003, Fournier 1998, Morgan and Hunt 1994). To explore the underlying attitudinal dimensions of the three relationship states, we complement the observed behavior data with survey-based data. This approach is consistent with the call of Gupta and Zeithaml (2006) for incorporating attitudinal and perceptual constructs with behavioral outcome models.

In the years 1998, 2000, and 2002, the alumni association conducted a survey that measured alumni engagement and attitude towards the university. Over 1,600 randomly sampled alumni were surveyed. The vast majority of these alumni were surveyed only in one of the three years. The questionnaire included questions about the relationship between alumni and the university in dimensions such as satisfaction, emotional connection, pride and others. Of the total survey sample, 128 surveys matched our original data set of 17,000 alumni. For this subset of 128 alumni, we calibrated the HMM (Model 1) and calculated, using Equation (9), the alumni probability of membership in each of the three relationship states in the year of the survey.¹⁷ We then assigned each alumna/alumnus to the relationship state with the highest probability. Table 9 compares the mean responses to the relationship questions across alumni in the three relationship states.

On all relationship measures, except affinity to the graduating class, the average ratings toward the university are increasing in the states from dormant to occasional to active. The last column in Table 9 presents the ANCOVA of the different relationship measures on the states’ membership with lagged donation as a covariate. Controlling for the lagged choice provides a more conservative estimate of the difference in the relationship measures between the states over and beyond the impact of an immediate donation.

This analysis revealed that even after controlling for the effect of typical state dependence, alumni in the active and occasional states had stronger feelings toward the university than those in the dormant state. Specifically, the difference was significant in terms of the emotional connection to the university, affinity with the graduating class, the feeling of responsibility toward the university, the perception of being valued by the university, and the likelihood of recommending the university to others. Indeed, it has been suggested in the relationship marketing literature that these dimensions of commitment, responsibility, and self-connection (Fournier 1998, Morgan

¹⁷ Because the gift-giving data ended at the end of 2001, the relationship state in 2001 was used for the survey of 2002.

Table 9 Behavioral Dimensions of Relationship in the Three States

Question	Scale	Dormant	Occasional	Active	ANCOVA P-values
How <i>satisfied</i> are you with your university experience?	1–5	4.51	4.75	4.80	0.220
How strong is your <i>feeling</i> about the university?	1–5	4.31	4.50	4.60	0.098
Do you feel <i>proud</i> of your university degree?	1–4	3.47	3.62	3.69	0.833
Do you feel your university experience helped <i>shape your life</i> ?	1–4	2.89	3.24	3.43	0.069
Do you feel a strong <i>emotional connection</i> to the university?	1–4	2.72	3.14	3.22	0.022*
Do you feel a strong <i>responsibility</i> to help the university?	1–4	2.28	2.66	3.03	0.015
Do you feel strong <i>affinity</i> with your graduating class?	1–4	1.94	2.52	2.26	0.009
How strongly would you <i>recommend</i> the university to prospective students?	1–4	3.35	3.67	3.74	0.042
How good of a job is the university doing in <i>servicing</i> your needs as an alum?	1–4	2.74	2.93	2.96	0.223
To what extent do you feel the university <i>values</i> its alumni?	1–3	2.25	2.41	2.55	0.047
Do your <i>parents/grandparents</i> have a degree from this university?	Yes/No	19%	18%	12%	0.109
Did you receive <i>financial aid</i> from the university as a student?	Yes/No	40%	40%	39%	0.557
Median lifetime donation (from the actual donation data)		\$100	\$475	\$1,382	
Sample size (<i>N</i>)		64	29	35	128

*Bold numbers reflect significant differences (at the 5% level) between the three relationship states.

and Hunt 1994) are important factors of customer-firm relational behavior.

These attitudinal measures give behavioral support to the projection from observed donation behavior to latent states of relationship. Furthermore, the high ratings for positive word-of-mouth among alumni in the active and occasional states suggest that alumni in these states do not only have a higher propensity to donate, but are also more likely to be active on other dimensions. Indeed, Arnett et al. (2003) used actual donation and word-of-mouth to measure relationship marketing success in the context of alumni-university relationships.

5. General Discussion

In this paper, we use data on alumni gift-giving behavior to estimate a hidden Markov model of relationship dynamics. The HMM, which was estimated using a hierarchical Bayes MCMC procedure to account for observed and unobserved heterogeneity, offers several insights into the drivers of these dynamics.

The main contribution of this research is in suggesting a behaviorally grounded model that helps marketers to infer the underlying structure of relationship states. Using the model, the researcher can dynamically classify customers into the relationship states and assess the dynamic effect of alternative time-varying covariates on the transition between the relationship states and consequent buying behavior.

The empirical application to the problem of university-alumni relationships demonstrates the use of the model to a dynamic relationship problem. Examining the time-varying covariates in the transition matrix, we find that the impact of reunion attendance is the strongest for alumni in the dormant

and the transitory, occasional state, while the effect of volunteering is stronger on alumni in the “sticky” dormant and active states. Both of these covariates have a long-term impact on the donation behavior. In terms of prediction ability, we demonstrate that, for our empirical application, the proposed model predicts future donations significantly better than the static latent class model and the dynamic, observed states, recency and frequency model. Additionally, we use the hold-out sample to demonstrate that HMM dynamic segmentation provides superior segmentation to the static latent class segmentation. More generally, the empirical application demonstrates the value of the proposed model for CRM marketers. The HMM enables marketers to use buyer behavior data to dynamically segment their customers into relationship states and estimate the evolution of customer relationships over time.

The proposed model extends the marketing literature by suggesting a Markovian framework for estimating the dynamics in customer relationships. Using the nonhomogeneous HMM, one can estimate the long-term impact of customer-firm interactions on customer relationships, an issue that has been largely neglected in the customer relationship literature. Methodologically, the proposed model extends the observed states models, such as the state-dependence and recency-frequency models by offering a latent specification of state dependence, which incorporates the effect of time-varying covariates into the state-dependence structure. The ability to determine the number of states based on the data at hand relaxes the ad-hoc specification of dynamics in the observed state models. The proposed model also extends the family of HMMs. To our knowledge, this is the first

nonhomogeneous HMM in the marketing literature that investigates the impact of time-varying covariates, such as customer-firm interactions, on the transition between the latent states. Indeed, Wedel and Kamakura (2000, p. 176) point out that the issue of nonstationarity in marketing segmentation in general, and specifically nonstationarity that could be related to time-varying covariates, has received limited attention.

The ability to incorporate time-varying covariates in the transition matrix opens the opportunity to investigate which marketing activities are most effective in building customer-firm relationships and driving actionable behavior, and to determine the optimal targeting of such marketing activities. Future research could investigate this issue using a data set that incorporates both exposure to marketing activities and observed buying behavior.

It is worthwhile to distinguish between the proposed nonhomogeneous HMM, which incorporates time-varying covariates in the transition matrix, and a nonstationary HMM (Djurić and Chun 2002), which models the transition probabilities as a function of the state duration. Because our main interest is in investigating the effect of customer-firm interactions on HMM dynamics, we adopt the first approach. However, future research could investigate the value of the additional flexibility offered by nonstationary HMMs in marketing applications.

To broaden the applicability of the proposed model to most common CRM data sets, we used only observed buying behavior to elicit the relationship states. Using survey data, we show that the relationship states, which were estimated based on observed buying behavior, are also different in terms of behavioral dimensions of relationship, such as satisfaction, emotional connection, and responsibility. Future research could use longitudinal survey data to estimate the relationship states. Estimating the dynamics in relationship states directly from attitudinal variables would be helpful in providing insight into the mediating factors of the connection between observed buying behavior and the relationship states.

To increase the external validity of the model, it would be constructive to investigate the application of this model in relationship marketing contexts other than university-alumni relationships. Some possible applications are other institutional gift giving; continuous service provider relationships, which face high churn rates such as banks and telephone carriers (Bolton 1998); institutional memberships (Thomas 2001); direct selling efforts; and dynamic brand choice problems. The HMM framework could also be extended to study alternative measures of customer relationships. For example, one could explore the connection between alternative relationship measures

such as service encounters, satisfaction, and actionable behavior (Bolton 1998). Future research could also investigate an HMM with multivariate outcomes such as donation amounts, donation incidence and participation in university events, as multiple modes of behavior that determine the customer relationship state.

To summarize, we believe that from the modeling perspective, we have provided CRM practitioners with an implementable model for evaluating customer relationships, their evolution over time, and the effect of customer-firm interactions on altering this evolution. These factors are necessary to transform a CRM system from an information system into a decision support system. This research takes a fundamental step in that direction.

Acknowledgments

The authors thank Amy Paulsen and the alumni association members for help with the data, and Michaela Draganska, Ricardo Montoya, Olivier Toubia, and participants in seminars at Carnegie Mellon University, Chicago Business School, Columbia University, Hebrew University at Jerusalem, Hong-Kong University of Science and Technology, London Business School, New York University, Northwestern University, Stanford University, The Interdisciplinary Center at Hertzelia, The Israeli Institute of Technology, University of Maryland, University of Texas at Austin, University of Texas at Dallas, and University of Toronto for helpful comments and suggestions. The authors are also grateful for the support of the Marketing Science Institute through the Alden G. Clayton Dissertation Competition.

Appendix A. Hierarchical Bayes Estimation Algorithm

The parameters in our model could be divided into two groups: (1) parameters that vary across individuals (random-effect parameters) and (2) parameters that do not vary across individuals (fixed parameters). The set of parameters in each group is determined by the model's heterogeneity specification as described in §§3.2.4 and 3.4.

We denote by θ_i the set of random-effect parameters and by ψ the set of fixed parameters. For the HMM described in §3, the random-effect parameters includes: $\theta_i = \{\mu(s')_{i1}, \dots, \mu(s')_{iNS-1}\}$, $s' \in \{1, \dots, NS\}$. The vector ψ includes: $\psi = \{\rho_1, \rho_2, \dots, \rho_{NS}, \beta_{01}, \beta_{02}, \dots, \beta_{0NS}, \beta_1, \beta_2, \dots, \beta_{NS}\}$.

Observed individual characteristics, such as demographics, can be introduced into the model in a hierarchical manner (Allenby and Ginter 1995). Thus, the vector of heterogeneous parameters (θ_i) can be written as a function of observed and unobserved individual characteristics

$$\theta_i = \delta' \cdot Z_i + \varepsilon_{\theta_i}, \quad (A1)$$

where Z_i is a vector of individual characteristics for individual i , δ is a matrix of parameters, relating the individual characteristics to the random-effect parameters (θ_i), and $\varepsilon_{i\theta} \sim N(0, \Sigma_\theta)$.

The MCMC procedure recursively generates draws from the conditional distribution of the model's parameters:

$$\begin{aligned} & \{\boldsymbol{\theta}_i\} \mid \mathbf{Y}_i, \mathbf{X}_i, \mathbf{a}_i, \mathbf{Z}_i, \boldsymbol{\psi}, \boldsymbol{\delta}, \boldsymbol{\Sigma}_0 \\ & \boldsymbol{\delta} \mid \{\boldsymbol{\theta}_i\}, \mathbf{Z}, \boldsymbol{\Sigma}_0 \\ & \boldsymbol{\Sigma}_0 \mid \{\boldsymbol{\theta}_i\}, \mathbf{Z}, \boldsymbol{\delta} \\ & \boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{X}, \mathbf{a}, \{\boldsymbol{\theta}_i\}. \end{aligned}$$

where \mathbf{Y}_i , \mathbf{X}_i , and \mathbf{a}_i are the vectors of donations, covariates in the state-dependent choice, and covariates in the transition matrix, respectively, as described in §4.3.

(1) **Generate $\boldsymbol{\theta}_i$**

$$\begin{aligned} & f(\{\boldsymbol{\theta}_i\} \mid \mathbf{Y}_i, \mathbf{X}_i, \mathbf{a}_i, \mathbf{Z}_i, \boldsymbol{\psi}, \boldsymbol{\delta}, \boldsymbol{\Sigma}_0) \\ & \propto N(\{\boldsymbol{\theta}_i\} \mid \boldsymbol{\psi}, \mathbf{Z}_i, \boldsymbol{\delta}), \boldsymbol{\Sigma}_0 L(\mathbf{Y}_i) \\ & \propto |\boldsymbol{\Sigma}_0|^{-1/2} \exp[-1/2(\boldsymbol{\theta}_i - \boldsymbol{\delta}'\mathbf{Z}_i)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\delta}'\mathbf{Z}_i)] L(\mathbf{Y}_i), \quad (\text{A2}) \end{aligned}$$

where $L(\mathbf{Y}_i)$ is the likelihood function from Equation (8).

Because (A2) does not have a closed form, the Metropolis-Hastings algorithm is used to draw from the conditional distribution of $\boldsymbol{\theta}_i$. The Metropolis-Hastings algorithm proceeds as follows: Let's define $\boldsymbol{\theta}_i^{(k)}$ as the accepted draw of $\boldsymbol{\theta}_i$ in iteration k and $\boldsymbol{\theta}_i^{(k+1)}$ as the draw in iteration $k+1$. Then, the sequence of draws is given by $\boldsymbol{\theta}_i^{(k+1)} = \boldsymbol{\theta}_i^{(k)} + \Delta\boldsymbol{\theta}$, where $\Delta\boldsymbol{\theta}$ is a draw from $N(0, \sigma^2\Delta)$, and σ and Δ are chosen adaptively to reduce the autocorrelation among the MCMC draws, with an acceptance rate of approximately 20%, following Atchadé (2006).

The probability of accepting $\boldsymbol{\theta}_i^{(k+1)}$ is

$$\begin{aligned} & \text{Pr}(\text{acceptance}) \\ & = \min \left\{ \frac{[\exp(-1/2(\boldsymbol{\theta}_i^{(k+1)} - \boldsymbol{\delta}'\mathbf{Z}_i)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\theta}_i^{(k+1)} - \boldsymbol{\delta}'\mathbf{Z}_i))] L(\mathbf{Y}_i \mid \boldsymbol{\theta}_i^{(k+1)})}{[\exp(-1/2(\boldsymbol{\theta}_i^{(k)} - \boldsymbol{\delta}'\mathbf{Z}_i)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\theta}_i^{(k)} - \boldsymbol{\delta}'\mathbf{Z}_i))] L(\mathbf{Y}_i \mid \boldsymbol{\theta}_i^{(k)})}, 1 \right\}. \quad (\text{A3}) \end{aligned}$$

(2) **Generate $\boldsymbol{\delta}$**

Define $\mathbf{v}\boldsymbol{\delta} = \text{vec}(\boldsymbol{\delta}')$.

$$[\mathbf{v}\boldsymbol{\delta} \mid \{\boldsymbol{\theta}_i\}, \mathbf{Z}, \boldsymbol{\Sigma}_0] = MVN(\mathbf{u}_n, \mathbf{V}_n),$$

where

$$\begin{aligned} & \mathbf{V}_n = [(\mathbf{Z}'\mathbf{Z} \otimes \boldsymbol{\Sigma}_0^{-1}) + \mathbf{V}_0^{-1}]^{-1}, \\ & \mathbf{u}_n = \mathbf{V}_n[(\mathbf{Z}' \otimes \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\Theta}^* + \mathbf{V}_0^{-1}\mathbf{u}_0], \\ & \mathbf{Z} = (\mathbf{Z}'_1, \mathbf{Z}'_2, \dots, \mathbf{Z}'_N) \text{ is an } N \times nz \text{ matrix of covariates,} \\ & \boldsymbol{\Theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_N) \text{ is an } N \times n\theta \text{ matrix which stacks } \{\boldsymbol{\theta}_i\}, \\ & \boldsymbol{\Theta}^* = \text{vec}(\boldsymbol{\Theta}'), \\ & \mathbf{V}_0 \text{ and } \mathbf{u}_0 \text{ are prior hyperparameters,} \\ & n\theta = \text{dim}(\boldsymbol{\theta}_i), \text{ and} \\ & nz = \text{dim}(\mathbf{z}_i). \end{aligned}$$

We define diffuse priors by setting \mathbf{u}_0 to a $n\theta \cdot nz \times 1$ vector of zeros, and $\mathbf{V}_0 = 100\mathbf{I}_{n\theta \cdot nz}$.

(3) **Generate $\boldsymbol{\Sigma}_0$**

$$\begin{aligned} & [\boldsymbol{\Sigma}_0 \mid \{\boldsymbol{\theta}_i\}, \mathbf{Z}, \boldsymbol{\delta}] \\ & \sim IW_{n\theta} \left(f_0 + N, \mathbf{G}_0^{-1} + \sum_{i=1}^N (\boldsymbol{\theta}_i - \boldsymbol{\delta}'\mathbf{Z}_i)' (\boldsymbol{\theta}_i - \boldsymbol{\delta}'\mathbf{Z}_i) \right), \quad (\text{A4}) \end{aligned}$$

where f_0 and \mathbf{G}_0 are prior hyperparameters, f_0 is the degrees of freedom, and \mathbf{G}_0 is the scale matrix of the inverse Wishart distribution. We define diffuse priors by setting $f_0 = n\theta + 5$, and $\mathbf{G}_0 = \mathbf{I}_{n\theta}$.

(4) **Generate $\boldsymbol{\psi}$**

Similar to (A2) the conditional distribution of $\boldsymbol{\psi}$ can be defined by

$$\begin{aligned} & [\boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{X}, \mathbf{a}, \{\boldsymbol{\theta}_i\}] \\ & \propto L(\mathbf{Y}) N(\boldsymbol{\psi}_0, \mathbf{V}_{\boldsymbol{\psi}_0}) \\ & \propto |\mathbf{V}_{\boldsymbol{\psi}_0}|^{-1/2} \exp[-1/2(\boldsymbol{\psi} - \boldsymbol{\psi}_0)' \mathbf{V}_{\boldsymbol{\psi}_0}^{-1} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)] L(\mathbf{Y}), \quad (\text{A5}) \end{aligned}$$

where $\boldsymbol{\psi}_0$ and $\mathbf{V}_{\boldsymbol{\psi}_0}$ are diffused priors, and $L(\mathbf{Y})$ is the likelihood function from Equation (8). Because (A5) does not have a closed form, the M-H algorithm is used to draw from the conditional distribution of $\boldsymbol{\psi}$. The acceptance probability at step $k+1$ is defined by

$$\begin{aligned} & \text{Pr}(\text{acceptance}) \\ & = \min \left\{ \frac{[\exp(-1/2(\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}_0)' \mathbf{V}_{\boldsymbol{\psi}_0}^{-1} (\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}_0))] L(\mathbf{Y} \mid \boldsymbol{\psi}^{(k+1)})}{[\exp(-1/2(\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0)' \mathbf{V}_{\boldsymbol{\psi}_0}^{-1} (\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}_0))] L(\mathbf{Y} \mid \boldsymbol{\psi}^{(k)})}, 1 \right\}. \quad (\text{A6}) \end{aligned}$$

We define diffuse priors for the conditional distribution of $\boldsymbol{\psi}$ by setting $\boldsymbol{\psi}_0$ to a $n\psi \times 1$ vector of zeros, and $\mathbf{v}_{\boldsymbol{\psi}_0} = 30\mathbf{I}_{n\psi}$, where $n\psi = \text{dim}(\boldsymbol{\psi})$.

Appendix B. Adapting the Markov Switching Criterion to Our Estimation Algorithm

Recently, Smith, Naik and Tsai (2006; hereafter SNT) developed the Markov switching criterion (MSC) for HMMs' states and variables selection. The MSC as described by SNT was developed for a HMM with a stationary transition matrix, estimated using aggregate data and common parameters across individuals. In contrast, our HMM is estimated using individual-level data, with some of the parameters varying across individuals. Additionally, we allow the transition matrix to be a function of time-varying covariates, thus relaxing the assumption of stationary transition matrix commonly assumed in HMMs. Therefore, to apply the MSC as a criterion for choosing the number of states, we need to adapt it to our specific application.

Following Equation (15) in SNT the MSC is given by

$$MSC = -2 \log(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi})) + \sum_{s=1}^{NS} \frac{\hat{T}_s (\hat{T}_s + \lambda_s K)}{\delta_s \hat{T}_s - \lambda_s K - 2}, \quad (\text{B1})$$

where

$-2 \log(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi}))$ is the maximized log-likelihood value,
 $\hat{T}_s = \text{trace}(\hat{W}_s)$, $\hat{W}_s = \text{diag}(\hat{P}_s)$, and $\hat{P}_s = (\text{prob}(S_1 = s), \dots, \text{prob}(S_T = s))$,

$\lambda_s = NS$; following SNT recommendation, where NS is the number of states,

$\delta_s = 1$; following SNT recommendation, and

K is the number of covariates in the state-dependent vector.

To adapt the MSC to our application we need to modify the different components in (B1).

The first term in Equation (B1), $-2 \log(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi}))$, captures the model's fit. Because our model is estimated using an MCMC hierarchical Bayes estimation procedure, one needs to define the procedure used to calculate the model's fit measure using the simulation output. We adopt the deviance measure suggested by Spiegelhalter et al. (2002) as

the model fit component of the deviance information criterion (DIC). We calculate the deviance as the expectation over the estimates at each iteration of the MCMC simulation. Thus,

$$-2\log(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi})) = \sum_{j=1}^J \sum_{i=1}^N -2\log(f(\mathbf{Y}_i, \hat{\boldsymbol{\theta}}_i^j, \boldsymbol{\Psi}^j)), \quad (\text{B2})$$

where

J is the number of MCMC iterations used to obtain the posterior distributions after the burn-in period,

N is the number of individuals,

$\hat{\boldsymbol{\theta}}_i^j$ is the vector of parameter estimates for individual i in iteration j , and

$f(\mathbf{Y}_i, \hat{\boldsymbol{\theta}}_i^j, \boldsymbol{\Psi}^j)$ is the likelihood of observing a sequence of choices following Equation (8).

The term \hat{T}_s in (B1) captures the effective sample size at state s . Accordingly, we define

$$\hat{T}_s = \sum_{i=1}^N \sum_{t=1}^{T_i} P(S_{it} = s),$$

where $P(S_{it} = s)$ is the probability that individual i is in state s at time T . Thus, the only difference between our formulation of \hat{T}_s and that of SNT is the summation over individuals in our model.

Finally, we need to count the number of additional parameters in each state K . Unlike SNT, we include covariates both in the transition matrix and in the state-dependent vector. Accordingly, in our application, K is the number of covariates for each state in both the transition matrix and in the state-dependent vector. Note that because the parameters that capture the effect of the covariates are common across individuals we can simply “count” the number of parameters. If one estimates a full random-effect model using hierarchical Bayesian approach, we recommend using the effective number of parameters, pD of Spiegelhalter et al. (2002), divided by the number of states as a measure for K .

References

- Aaker, J., S. Fournier, A. S. Brasel. 2004. When good brands do bad. *J. Consumer Res.* **31**(June) 1–18.
- Allenby, G. M., J. L. Ginter. 1995. Using extremes to design products and segment markets. *J. Marketing Res.* **32**(November) 392–403.
- Allenby, G. M., R. P. Leone, L. Jen. 1999. A dynamic model of purchase timing with application to direct marketing. *J. Amer. Statist. Assoc.* **93**(446) 365–374.
- Andrews, R. L., I. S. Currim. 2003. A comparison of segment retention criteria for finite mixture models. *J. Marketing Res.* **39**(May) 235–243.
- Arnett, D. B., S. German, S. D. Hunt. 2003. The identity salience model of relationship marketing success: The case of nonprofit marketing. *J. Marketing* **67**(2) 89–105.
- Atchadé, Y. F. 2006. An adaptive version for the metropolis adjusted Langevin algorithm with a truncated drift. *Methodology and Comput. Appl. Probab.* **8**(2) 235–254.
- Baade, R. A., J. O. Sundberg. 1996. What determines alumni generosity? *Econom. Ed. Rev.* **15**(1) 75–81.
- Böckenholt, U., W. R. Dillon. 2000. Inferring latent brand dependencies. *J. Marketing Res.* **37**(February) 72–87.
- Böckenholt, U., R. Langeheine. 1996. Latent change in recurrent choice data. *Psychometrika* **61**(2) 285–302.
- Bolton, R. N. 1998. A dynamic model of the duration of the customer’s relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* **17**(1) 45–65.
- Bucklin, R. E., J. M. Lattin. 1991. A two-state model of purchase incidence and brand choice. *Marketing Sci.* **10**(1) 24–39.
- Bult, J. R., T. Wansbeek. 1995. Optimal selection for direct mail. *Marketing Sci.* **14** (4) 378–394.
- DeKimppe, M. G., D. M. Hanssens. 2000. Time-series models in marketing: Past, present and future. *Internat. J. Res. Marketing* **17**(2–3) 183–193.
- Dillon, W. R., U. Böckenholt, M. S. de Borrero, H. Bozdogan, W. DeSarbo, S. Gupta, W. Kamakura, A. Kumar, V. Ramaswamy, M. Zenor. 1994. Issues in the estimation and application of latent structure models of choice. *Marketing Lett.* **5**(4) 323–334.
- Djurić, P. M., J.-H. Chun. 2002. An MCMC sampling approach to estimation of nonstationary hidden Markov models. *IEEE Trans. Signal Processing* **50**(5) 1113–1123.
- Du, R., W. A. Kamakura. 2006. Household lifecycles and lifestyles in america. *J. Marketing Res.* **43**(February) 121–132.
- Dwyer, R. F., P. H. Schurr, S. Oh. 1987. Developing buyer-seller relationships. *J. Marketing* **51**(April) 11–27.
- Erdem, T., B. Sun. 2001. Testing for choice dynamics in panel data. *J. Bus. Econom. Statist.* **19**(2) 142–152.
- Fader, P. S., J. M. Lattin. 1993. Accounting for heterogeneity and nonstationarity in a cross-sectional model of consumer purchase behavior. *Marketing Sci.* **12**(3) 304–317.
- Fader, P. S., B. G. S. Hardie, C.-Y. Huang. 2004. A dynamic change-point model for new product sales forecasting. *Marketing Sci.* **23**(1) 50–65.
- Flanagan, J. C. 1954. The critical incident technique. *Psych. Bull.* **51**(4) 327–359.
- Fournier, S. 1998. Consumers and their brands: Developing relationship theory in consumer research. *J. Consumer Res.* **24**(March) 343–373.
- Frank, R. E. 1962. Brand choice as a probability process. *J. Bus.* **35**(January) 43–56.
- Gelman, A., D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–511.
- Greene, W. H. 1997. *Econometric Analysis*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ.
- Guadagni, P. M., J. D. C. Little. 1983. A logit model of brand choice Calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.
- Gupta, S., V. A. Zeithaml. 2006. Customer metrics and their impact on financial performance. *Marketing Sci.* **25**(6) 718–739.
- Hamilton, J. D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**(2) 357–384.
- Harrison, W. B., S. K. Mitchell, S. P. Peterson. 1995. Alumni donations and colleges’ development expenditures: Does spending matter? *Amer. J. Econom. Social.* **54**(October) 397–413.
- Heckman, J. J. 1981. Heterogeneity and state dependence. S. Rosen, ed. *Studies in Labor Markets*. University of Chicago Press, Chicago, 91–139.
- Hinde, R. A. 1979. *Towards Understanding Relationships*. Academic, London.
- Hughes, J. P., P. Guttorp. 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Res.* **30**(5) 1535–1546.
- Jain, D., S. S. Singh. 2002. Customer lifetime value research in marketing: A review and future direction. *J. Interactive Marketing* **16**(2) 34–46.
- Jeuland, A. P., F. M. Bass, G. P. Wright. 1980. A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Oper. Res.* **28**(2) 255–277.

- Kahneman, D., D. T. Miller. 1986. Norm theory: Comparing reality to its alternatives. *Psych. Rev.* **93**(2) 136–153.
- Kamakura, W. A., G. J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(November) 379–390.
- Keane, M. P. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15**(July) 310–327.
- Kim, C.-J., C. R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. M.I.T. Press, Cambridge, MA.
- Lewis, M. 2004. The influence of loyalty programs and short-term promotions on customer retention. *J. Marketing Res.* **41**(August) 281–292.
- Libai, B., D. Narayandas, C. Humby. 2002. Toward an individual customer profitability model: A segment-based approach. *J. Service Res.* **5**(1) 69–76.
- Liechty, J. C., M. Wedel, R. Pieters. 2003. The representation of local and global scanpaths in eye movements through Bayesian hidden Markov models. *Psychometrika* **68**(4) 519–541.
- MacDonald, I. L., W. Zucchini. 1997. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall, London.
- McAlister, L., R. Srivastava, J. Horowitz, M. Jones, W. Kamakura, J. Kulchitsky, B. Ratchford, G. Russell, F. Sultan, T. Yai, D. Weiss, R. Winer. 1991. Incorporating choice dynamics in models of consumer behavior. *Marketing Lett.* **2**(3) 241–252.
- Montgomery, A. L., S. Li, K. Srinivasan, J. C. Liechty. 2004. Predicting online purchase conversion using web path analysis. *Marketing Sci.* **23**(4) 579–595.
- Moon, S., W. A. Kamakura, J. Ledolter. 2007. Estimating promotion response when competitive promotions are unobservable. *J. Marketing Res.* Forthcoming.
- Morgan, R. M., S. D. Hunt. 1994. The commitment-trust theory of relationship marketing. *J. Marketing* **58**(3) 20–38.
- Morrison, D. G., R. D. H. Chen, S. L. Karpis, K. E. A. Britney. 1982. Modeling retail customer behavior at merrill lynch. *Marketing Sci.* **1**(2) 123–141.
- Naik, P. A., M. K. Mantrala, A. G. Sawyer. 1998. Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* **17**(3) 214–235.
- Netzer, O. 2004. A hidden Markov model of customer relationship dynamics. Ph.D. dissertation, Stanford University, Stanford, CA.
- Newton, M. A., A. E. Raftery. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. *J. Roy. Statist. Soc. B* **56**(1) 3–48.
- Okunade, A. A. 1993. Logistic regression and probability of business school alumni donations: Micro-data evidence. *Ed. Econom.* **1** 243–258.
- Okunade, A. A., R. L. Berl. 1997. Determinants of charitable giving of business school alumni. *Res. Higher Ed.* **38**(April) 201–214.
- Oliver, R. L. 1997. *Satisfaction: A Behavioral Perspective on the Consumer*. McGraw-Hill, New York.
- Pauwels, K., I. Currim, M. G. Dekimpe, E. Ghysels, D. M. Hanssens, N. Mizik, P. Naik. 2004. Modeling marketing dynamics by time series econometrics. *Marketing Lett.* **15**(4) 167–183.
- Pfeifer, P. E., R. L. Carraway. 2000. Modeling customer relationships using Markov chains. *J. Interactive Marketing* **14**(2) 43–55.
- Poulsen, C. S. 1990. Mixed Markov and latent Markov modeling applied to brand choice behavior. *Internat. J. Res. Marketing* **7**(1) 5–19.
- Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**(2) 257–286.
- Rabiner, L. R., B.-H. Juang. 1993. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ.
- Ramaswamy, V. 1997. Evolutionary preference segmentation with panel survey data: An application to new products. *Internat. J. Res. Marketing* **14**(1) 57–80.
- Reinartz, W., V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* **67**(1) 77–99.
- Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Sci.* **22**(3) 304–328.
- Rust, R. T., T. S. Chung. 2006. Marketing models of service and relationships. *Marketing Sci.* **25**(6) 560–580.
- Rust, R. T., K. N. Lemon, V. A. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *J. Marketing* **68**(1) 109–127.
- Schmittlein, D. C., R. A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* **13**(1) 41–67.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* **6**(2) 461–464.
- Scott, L. S. 2002. Bayesian methods for hidden Markov models, recursive computing in the 21st century. *J. Amer. Statist. Assoc.* **97**(475) 337–351.
- Smith, A., P. A. Naik, C.-L. Tsai. 2006. Markov-switching model selection using Kullback-Leibler divergence. *J. Econometrics* **134**(2) 553–577.
- Soukup, J. D. 1983. A Markov analysis of fund-raising alternatives. *J. Marketing Res.* **20**(August) 314–319.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc. Ser. B* **64**(3) 583–639.
- Srinivasan, V., R. Kesavan. 1976. An alternative interpretation of the linear learning model of brand choice. *J. Consumer Res.* **3**(September) 76–83.
- Storm, S. 2003. After years of cash flow, universities hit an ebb. *New York Times* **13**(March), 25.
- Taylor, A. L., J. C. Martin, Jr. 1995. Characteristics of alumni donors at a research I public university. *Res. Higher Ed.* **36**(3) 283–302.
- Thomas, J. S. 2001. The importance of linking customer acquisition to customer retention. *J. Marketing Res.* **39**(May) 262–268.
- Wedel, M., W. A. Kamakura. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer, Dordrecht, The Netherlands.
- Willemain, T. R., A. Goyal, M. Van Deven, I. S. Thukral. 1994. Alumni giving: The influences of reunion, class and year. *Res. Higher Ed.* **35**(2) 201–214.
- Winer, R. S. 2001. A framework for customer relationship management. *California Management Rev.* **43**(Summer) 89–105.
- Xie, J. X., M. Song, M. Sirbu, Q. Wang. 1997. Kalman filter estimation of new product diffusion models. *J. Marketing Res.* **34**(August) 378–393.