

Research Paper No. 1904 (R)

**A Hidden Markov Model of Customer
Relationship Dynamics**

**Oded Netzer
James M. Lattin
V. Srinivasan**

May 2007

RESEARCH PAPER SERIES

STANFORD
GRADUATE SCHOOL OF BUSINESS



<http://ssrn.com/abstract=998603>

A HIDDEN MARKOV MODEL OF CUSTOMER RELATIONSHIP DYNAMICS

Oded Netzer*
Columbia University

James M. Lattin
Stanford University

V. Srinivasan
Stanford University

May 2007

* Oded Netzer is Assistant Professor of Marketing, Graduate School of Business, Columbia University (e-mail: on2110@columbia.edu). James M. Lattin is the Robert A. Magowan Professor of Marketing, Graduate School of Business, Stanford University (e-mail: jlattin@stanford.edu). V. "Seenu" Srinivasan is the Adams Distinguished Professor of Management, Graduate School of Business, Stanford University (e-mail: srinivasan_seenu@gsb.stanford.edu). The authors are indebted to Amy Paulsen and the alumni association members for their help with the data. The authors are also grateful for the helpful comments and suggestions received from Michaela Draganska, Ricardo Montoya, Olivier Toubia and the 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. We are grateful for the support of the MSI through the Alden G. Clayton Dissertation Competition.

A HIDDEN MARKOV MODEL OF CUSTOMER RELATIONSHIP DYNAMICS

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 may 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 the 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 dataset. Using the proposed model, we are able to 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.

Keywords: Customer relationship management, hidden Markov models, dynamic choice models, segmentation, Bayesian analysis

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 models 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 firm¹, 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 that captures the impact of the series 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

¹ Since the model proposed in this paper applies to the customer’s relationship with firms, brands, services, or non-profit and for-profit organizations, we use the term “firm” to represent the business partner for the relationship with the customer.

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 dataset. 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 states and from the occasional to the active states. 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 hidden Markov model for capturing the dynamics of customer relationships and describes the hierarchical Bayes estimation procedure. In Section 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, Schurr and Oh 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, Fournier and Brasel 2004; Fournier 1998; Hinde 1979). Further, 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 may 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 may 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, Leone and Jen 1999; Bolton 1998; Reinartz and Kumar 2003; Schmittlein and Peterson 1994) and customer lifetime value (e.g., Rust, Lemon and Zeithaml 2004; Libai, Narayandas and Humby 2002). With the exception of Reinartz and Kumar (2003), these models do not take into consideration the dynamics in the relationship that result from the 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 situations exhibit dynamics in consumers' behavior. These situations include both individual consumer behavior (often represented by discrete observations such as brand choice or purchase incidence) as well as in the firm's or market performance (often represented by continuous measures such as sales or market share). The difficulty with capturing such dynamics is that in most marketing datasets 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, the 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 Pauwels et al. 2004 and Dekimpe and Hanssens 2000 for a review). Naik, Mantrala and Sawyer (1998) and Xie et al. (1997) used 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. 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) found positive and significant state dependence effects across product categories even after controlling for heterogeneity; other studies found mixed results (e.g., Jeuland, Bass and Wright 1980). In the context of customer relationship management, 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, non-contributor) 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, for example, the enduring effects of marketing stimuli. Indeed, for exogenous variables that are correlated over time, and are not controlled for, previous behavior may be a determinant of current behavior just 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 since marketing actions such as loyalty programs (Lewis 2004), and customer initiated interactions like service encounters (Bolton 1998) may alter the customer's relationship with the firm, and therefore may 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 a Hidden Markov Model (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 a HMM to estimate the impact of discrete regime shifts on the growth rates of the real GNP.

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) surveyed these studies, as well as alternative forms of dynamics in segmentations (e.g., Böckenholt and Dillon 2000; Böckenholt and Langeheine 1996). Wedel and Kamakura concluded that the issue of non-stationarity in marketing segmentation should be further investigated. More recently, Smith, Naik and Tsai (2006) developed 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, Hardie and Huang (2004), proposed a changepoint model to predict new product sales. Like HMMs, their changepoint model allows for a discrete shift in preferences, however, unlike HMMs, their model's state space is continuous. Liechty, Wedel and Pieters (2003) applied a HMM to identify visual attention mode in advertising viewing. Du and Kamakura (2006) used a hidden Markov model to identify latent states in American families' lifecycles. Moon, Kamakura and Ledolter (2007) used a random-effect HMM to augment unobserved competitors' promotions in a pharmaceutical context. Using simulation and actual prescription and sampling data, the authors demonstrate superior performance of the HMM relative to alternative methods commonly used to augment the unobserved promotions.

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. Since 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 driver of the dynamics rather than merely build a model that fits the dynamics in the data. Second, CRM datasets are often collected at the individual level. When

modeling dynamics using individual-level data it is crucial to account for heterogeneity to distinguish a 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.

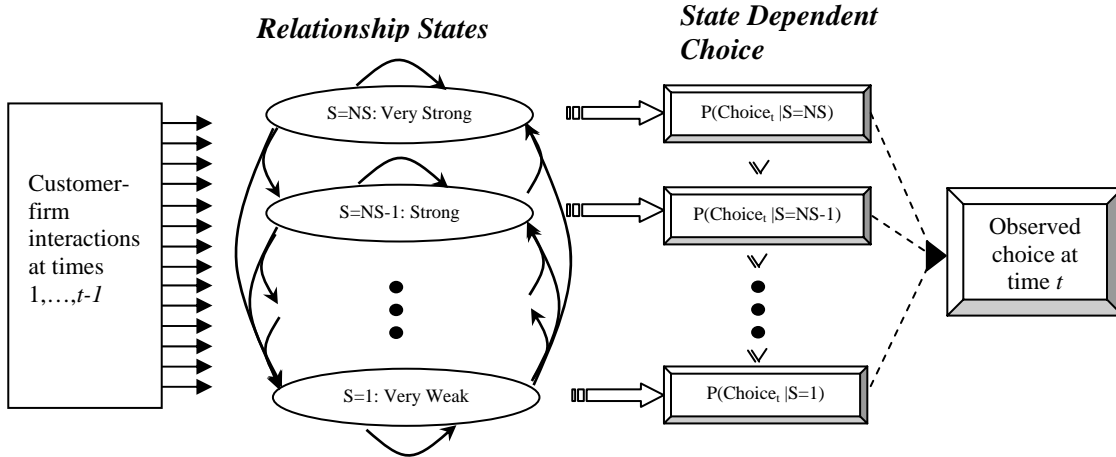
We define a *relationship encounter* as an interaction between the customer and the firm. Such interactions may include purchase transactions, exposure to relationship marketing activities or other non-purchase related exposure to firm. Relationships are made of a longitudinal sequence of relationship 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 a Hidden Markov Model.

A 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. In this paper, we develop a HMM of repeated binary choice that relates

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

transitions between the latent relationship states to the observed buying behavior (see Figure 1 for a graphical representation of the proposed HMM).

Figure 1 - A Hidden Markov Model of Customer Relationships



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* (Q_i) 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: $P(S_{it} = s' | S_{it-1} = s) = q_{its'}$.
- 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,

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

$$\begin{array}{c}
\text{State at } t-1 \\
\mathbf{Q}_{i,t-1 \rightarrow t} = \begin{array}{c} 1 \\ 2 \\ \vdots \\ NS \end{array}
\end{array}
=
\begin{array}{c}
\begin{array}{c} \text{State at } t \\ 1 \quad 2 \quad 3 \quad \dots \quad NS-1 \quad NS \end{array} \\
\begin{array}{|c|c|c|c|c|c|}
\hline
q_{it11} & q_{it12} & q_{it13} & \dots & q_{it1NS-1} & q_{it1NS} \\
\hline
q_{it21} & q_{it22} & q_{it23} & \dots & q_{it2NS-1} & q_{it2NS} \\
\hline
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\hline
q_{itNS1} & q_{itNS2} & q_{itNS3} & \dots & q_{itNSNS-1} & q_{itNSNS} \\
\hline
\end{array}
\end{array}
\quad (1)$$

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 $0 \leq q_{itss'} \leq 1 \quad \forall s, s'$, and $\sum_{s'} q_{itss'} = 1$. In applying our model in the context of alumni-university relationships (see Section 5), we put a structure on the general transition matrix in Equation (1). Specifically, we defined 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 consumer 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 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). Specifically, the terms $q_{iss'}$, in the transition matrix in Equation (1) could be written as:

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

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

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

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

where,

\mathbf{p}_{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 a 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 Section 3.2.1 and all covariates are set to their mean value across individuals and time periods.⁴

⁴ An alternative approach would be to take the stationary distribution of the transition matrix with the covariates set to zero (i.e., the transition matrix with intercept terms only). For the empirical application in Section 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 dataset is not left truncated (i.e., we observe the initial interaction between the customer and the brand), and the stationary distribution of the transition matrix at the mean of the covariates otherwise.

The stationarity conditions above do not guarantee 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. In this paper we model a discrete conditional choice since we believe that the act of choice is a stronger determinant of relationship strength than quantity measures. 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 distinction 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 towards the product or service and those that affect only the short-term choice behavior of the customer. 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.

⁵ 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.

To ensure identification of the states, we restrict the choice probabilities to be non-decreasing in the relationship states. Since both the intercepts and the 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

Since our model is dynamic, one must ensure that the zero-order heterogeneity is fully accounted for in order to distinguish it from time dynamics. Heckman (1981) demonstrated that estimating a random utility aggregate choice model based on a heterogeneous sample may 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} . 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:

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

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:

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

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:

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

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 (see MacDonald and Zucchini 1997 p. 79 for details). 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 Equation (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 $\boldsymbol{\theta}_i = \{\mu(s')_{i1}, \dots, \mu(s')_{iNS-1}, s' \in \{1, \dots, NS\}\}$ and $\boldsymbol{\Psi} = \{\boldsymbol{\rho}_1, \boldsymbol{\rho}_s, \boldsymbol{\rho}_{NS}, \beta_{01}, \beta_{02}, \dots, \beta_{0S}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{NS}\}$. Heterogeneity is introduced into the model for the random-effect parameters by defining diffuse priors $\boldsymbol{\theta}_i \sim MVN(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_\theta)$. 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 $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{\Psi}$. 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 utilize 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 utilizes only the information known up to time t to recover the individual's state at time t , whereas smoothing utilizes the full information available in the data. The filtering approach is more appealing for marketing applications, where decisions are made only based 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 dataset, the

⁸ The complete set of conditional distributions and the iteration sequence of the Markov Chain Monte Carlo simulation appear in Appendix A.

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 dimension 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 dataset contains most of the components suggested for a good CRM dataset. 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 will be discussed further below, this dataset 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 US 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, German and Hunt (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.

4.2 Data Description

The data used to calibrate and validate the model is sampled from the database provided by the alumni association of a large west coast university. Our dataset 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 dataset). From this dataset, 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).

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

<i>Key Characteristics</i>	<i>Percentage/ Average</i>
Overall observations (gift opportunities)	27526
Overall number of alumni	1256
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/us	
Yes	56%
No	44%
Alumni association membership	
Member	69%
Not member	31%

For each alumna/us 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 dataset also contains disaggregate information about

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

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/us (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 non-donation (for the full description of the Run Test results see Netzer 2004).

4.3 Variables Description

The variables of this dataset 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 Equation (3). 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 a valid for these two variables, since donations generally occur post the reunion attendance and volunteering decisions.
2. *Influence attempts* - These alumni-university interactions are postulated to have mainly short-term effects. These covariates determine the vector \mathbf{x}_{it} in Equation (7). 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 may raise the likelihood of giving in the specific year, but not in subsequent years. However, it is actual participation in a reunion that may lead to a change in the strength of relationship between the alumna/us 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 the alumni association, we aggregated gift giving and events data by “calendar year.”

To account for time dynamics that are not related to the relational marketing encounters, a linear time trend variable, representing the number of years since graduation, is included in the state dependent vector. Additionally, the dataset contains alumni characteristics such as year and major of graduation, gender, marriage to a university alumna/us and membership in the alumni association.

4.4 Estimated Models

We estimated the HMM of customer relationship dynamics described in Section 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 non-dynamic 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 coefficients across individuals. A limitation of this model is that this model cannot distinguish between heterogeneity and dynamics in the gift-giving behavior.

Model 3 – Non-dynamic 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 The probability of choice follows the binary logit formulation:

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.

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

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 state dependence parameters using hierarchical Bayes estimation.

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 below 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 density¹² and 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 (9), using the hold-out sample (the years 1999-2001 in the dataset). 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 factor (see Table 2).

Table 2 - Choosing the Number of States

Number of states	-2 Marginal log density	Log Bayes factor ¹⁴	DIC	MSC	Validation -2 log-likelihood
1	26708.5	-----	26493.0	50476.5	4768.5
2	25557.1	575.7	25378.8	48806.6	3689.1
3	25397.1	80.0	25176.3	48332.8	3583.6
4	25689.1	-146.2	25658.2	48639.5	3717.5

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 represent 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.

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

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

¹⁴ The log Bayes factors compare the model with NS state 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 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	25397.1		25565.9	

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

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

The identification 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%. 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 respondents) of reunion attendance and volunteering on the

state transitions. Attending a reunion increases the likelihood that an alumna/us in the dormant state moves to the occasional state from 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 several time periods and therefore should be affected by reunion attendance.

Table 4 – The Mean Posterior Transition Matrices*

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

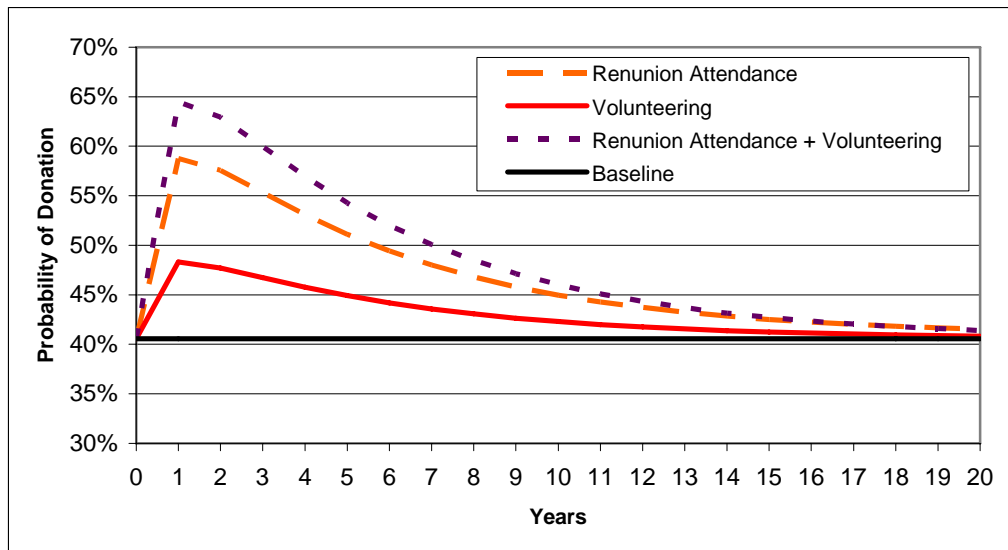
* 95% Confidence interval in parenthesis

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/us, thus one should be careful about treating these as decision variables in the context of our dataset. Nevertheless, given the observed behavior, one could still use these alumni-university interactions to dynamically segment alumni to the relationship states. More generally, 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 datasets in which the time-varying covariates are also decision variables.

The transition probabilities in Table 4 suggest that reunion participation may 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 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/us who is at initial state distribution at time 0 (average donation rate of 41%) and attended a reunion or volunteered to a university role at year 1.

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



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 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 $t+1$. Using Equation (11) we calculated the state membership probability at time t following a donation at time t .

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

where,

$\Pr(S_{it-1})$ 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.

Table 5 – Mean Posterior Transition Following a Donation at Time t

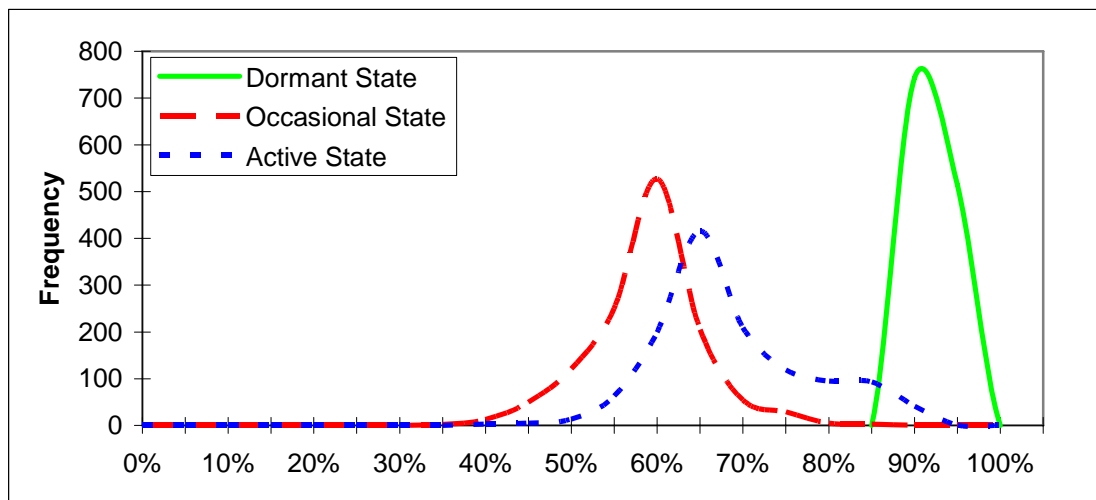
Donation at t			
t-1 \ t	Dormant	Occasional	Active
Dormant	53% [47%-58%]	47% [42%-53%]	0% [--- - ---]
Occasional	1% [1%-2%]	47% [43%-51%]	51% [47%-55%]
Active	0% [0%-0%]	16% [14%-18%]	78% [71%-86%]

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 that state in the previous period (the diagonal elements of the first transition matrix in Table 4).

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



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 Allenby and Ginter 1995 for details). Table 6 presents the parameter estimates of the observed heterogeneity covariates.

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/us	-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.

Several demographic 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.

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) x 3 (years) = 3,768 observations of possible gift giving in the validation period.

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

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 non-donation periods) as well as 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).

Table 7 - Fit Prediction Ability Measures

Model Measure	Model 1 HMM with heterogeneity	Model 2 HMM no heterogeneity	Model 3 Non-dynamic latent class	Model 4 Recency frequency model
Calibration: -2log marginal density/ -2 log likelihood	25397.1	25565.8	27399.8	26045.1
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 -2log likelihood	3583.6	3597.0	4484.2	4148.6

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. All models significantly improve upon a random choice rule, based on the RMSPE

The HMM predicts the hold-out choices significantly better than the non-dynamic latent class model ($z=9.5$; $p\text{-value}<0.001$) and the dynamic, observed states, recency frequency model ($z=5.1$; $p\text{-value}<0.001$). 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 non-donation years.

While the fit of 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 since 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.

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

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. In order to investigate further the difference between the dynamic and the static segmentation models, we analyzed the prediction 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 the calibration sample. The latent class solution suggested three segments which 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 choice 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.

Table 8 – Predictive Ability by Segmentation Method

Segment allocation	N	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	1256	67.9%	77.6%

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 outpredicted the latent class model for alumni in the more transient occasional state. This may suggest that the HMM segmentation matches the “true” alumni segmentation better than the static latent class model.

In summary, the previous section 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

In order to broaden the applicability of the HMM to most common CRM datasets, the model utilizes only observed behavioral measures (such as choice) to elicit the relationship states. This may 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, German and Hunt 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 customer feedback (Morgan and Rego 2006) 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 dataset of 17,000 alumni. For this subset of 128 alumni, we calibrated the HMM (Model 1) and calculated, using Equation (10), the alumni probability of membership in each of the three relationship states in the year of the survey.¹⁷ We then assigned each alumna/us 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.

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

On all relationship measures, except affinity to the graduating class, the average ratings towards 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.

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>serving</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	\$1382	
Sample Size (N)		64	29	35	128

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

This analysis revealed that even after controlling for the effect of typical state dependence, alumni in the active and occasional states had stronger feelings towards 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 towards 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 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. Further, 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, German and Hunt (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 some insights into the drivers of these dynamics.

The main contribution of this research is in suggesting a behaviorally structured 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 choices 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 model 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, pp. 176) point out that the issue of non-stationarity in marketing segmentation in general, and specifically non-stationarity 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 dataset 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. Since 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 using simulation and actual marketing data.

To broaden the applicability of the proposed model to most common CRM datasets, 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 relationships, 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 the study of 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.

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 section 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 Section 3, the random-effect vector of parameters includes:

$\theta_i = \{\mu(s')_{i1}, \dots, \mu(s')_{iNS-1}\}, s' \in \{1, \dots, NS\}$. The vector ψ includes:

$\psi = \{\rho_1, \rho_s, \rho_{NS}, \beta_{01}, \beta_{02}, \dots, \beta_{0S}, \beta_1, \beta_2, \dots, \beta_{NS}\}$.

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,

$$(A.1) \quad \theta_i = Z_i' \delta + \varepsilon_{\theta i},$$

where Z_i is a vector of individual characteristics for individual i , δ is a vector of covariates, relating the individual characteristics to the random effect parameters (θ_i), and $\varepsilon_{\theta i}$ is an error term.

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

$$\{\theta_i\} | Y_i, X_i, A_i, Z_i, \psi, \delta$$

$$\delta | \{\theta_i\}, Z, \Sigma_\theta$$

$$\Sigma_\theta | \{\theta_i\}, Z, \delta$$

$$\psi | Y, X, A, \{\theta_i\}$$

Where Y_i , X_i and A_i are the vectors of donations, covariates in the state dependent choice, and covariates in the transition matrix, respectively, as described in Section 4.3.

1) Generate θ_i

$$(A.2) \quad f(\{\theta_i\} | Y_i, X_i, A_i, Z_i, \psi, \delta) \propto N(\{\theta_i\} | \psi, Z_i, \delta) L(Y_i) \propto |\Sigma_\theta|^{-1/2} \exp[-1/2(\theta_i - \delta' Z_i)' \Sigma_\theta^{-1} (\theta_i - \delta' Z_i)] L(Y_i)$$

where $L(Y_i)$ is the likelihood function from Equation (8).

Since (A.2) does not have a closed form, the Metropolis-Hastings algorithm is used to draw from the conditional distribution of θ_i . The Metropolis-Hastings algorithm proceeds as follows; let's define $\theta_i^{(k)}$ as the accepted draw of θ_i in iteration k and $\theta_i^{(k+1)}$ as the draw in iteration $k+1$. Then, the sequence of draws is given by: $\theta_i^{(k+1)} = \theta_i^{(k)} + \Delta\theta$, where $\Delta\theta$ 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 $\theta_i^{(k+1)}$ is,

$$(A.3) \Pr(\text{acceptance}) = \min \left\{ \frac{[\exp(-1/2(\theta_i^{(k+1)} - \delta'Z_i)' \Sigma_0^{-1}(\theta_i^{(k+1)} - \delta'Z_i))]L(Y_i | \theta_i^{(k+1)})}{[\exp(-1/2(\theta_i^{(k)} - \delta'Z_i)' \Sigma_0^{-1}(\theta_i^{(k)} - \delta'Z_i))]L(Y_i | \theta_i^{(k)})}, 1 \right\},$$

2) Generate δ

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

$$[v\delta | \{\theta_i\}, Z, \Sigma_0] = MVN(u_n, V_n)$$

where,

$$V_n = \left[(Z'Z \otimes \Sigma_0^{-1}) + V_0^{-1} \right]^{-1},$$

$$u_n = V_n \left[(Z' \otimes \Sigma_0^{-1}) \Theta^* + V_0^{-1} u_0 \right],$$

$Z = (Z'_1, Z'_2, \dots, Z'_N)$ is an $N \times nz$ matrix of covariates,

$\Theta = (\theta'_1, \theta'_2, \dots, \theta'_N)$ is an $N \times n\theta$ matrix which stacks $\{\theta_i\}$,

$$\Theta^* = \text{vec}(\Theta'),$$

V_0 and u_0 are prior hyperparameters,

$n\theta = \dim(\theta_i)$, and

$nz = \dim(z_i)$,

We define diffuse priors by setting u_0 to a $n\theta \times 1$ vector of zeros, and $V_0 = 100I_{n\theta}$.

3) Generate Σ_0

$$(A.4) [\Sigma_0 | \{\theta_i\}, Z, \delta] \sim IW_{n\theta} \left(f_0 + N, G_0^{-1} + \sum_{i=1}^N (\theta_i - \delta'Z_i)'(\theta_i - \delta'Z_i) \right),$$

where f_0 and G_0 are prior hyperparameters, f_0 is the degrees of freedom, and G_0 is the scale matrix of the Inverse Wishart distribution. We define diffuse priors by setting $f_0 = n\theta + 5$, and

$$G_0 = I_{n\theta}.$$

4) Generate ψ

Similar to (A.2) the conditional distribution of ψ can be defined by,

$$(A.5) [\psi | Y, X, A, \{\theta_i\}] \propto L(Y)N(\psi_0, V_{\psi_0}) \propto |V_{\psi_0}|^{-1/2} \exp[-1/2(\psi - \psi_0)'V_{\psi_0}^{-1}(\psi - \psi_0)]L(Y),$$

where ψ_0 and V_{ψ_0} are diffused priors and $L(Y)$ is the likelihood function from Equation (8).

Since (A.5) does not have a closed form, the M-H algorithm is used to draw from the conditional distribution of ψ . The acceptance probability at step $k+1$ is defined by,

$$(A.6) \Pr(\text{acceptance}) = \min \left\{ \frac{[\exp(-1/2(\psi^{(k+1)} - \psi_0)'V_{\psi_0}^{-1}(\psi^{(k+1)} - \psi_0))]L(Y | \psi^{(k+1)})}{[\exp(-1/2(\psi^{(k)} - \psi_0)'V_{\psi_0}^{-1}(\psi^{(k)} - \psi_0))]L(Y | \psi^{(k)})}, 1 \right\},$$

We define diffuse priors for the conditional distribution of ψ by setting ψ_0 to a $n\psi \times 1$ vector of zeros, and $V_{\psi_0} = 30I_{n\psi}$, where $n\psi = \dim(\psi)$.

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 therefore using 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, in order 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:

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

Where,

$-2\log\left(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi})\right)$ is the maximized log-likelihood value,

$\hat{T}_s = \text{trace}(\hat{W}_s)$, $\hat{W}_s = \text{diag}(\hat{P}_s)$ and $\hat{P}_s = (\text{prob}(S_1 = s), \dots, \text{prob}(S_t = s), \dots, \text{prob}(S_T = s))$,

$\lambda_s = NS$; 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 modify the different components in (B.1).

The first term in Equation (B.1), $-2\log\left(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi})\right)$, captures the model's fit. Since 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,

$$(B.2) \quad -2\log\left(f(\mathbf{Y}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\psi})\right) = \sum_{j=1}^J \sum_{i=1}^I -2\log\left(f(\mathbf{Y}_i, \hat{\boldsymbol{\theta}}_i^j, \boldsymbol{\psi}^j)\right)$$

where,

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

I is the number of individuals,

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

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

The term \hat{T}_s in (B.1) 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 SNT is the summation over individuals in our model.

Finally, we need to count for 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 since the parameters that capture the effect of the covariates are common across individuals we can simply “count” the number of parameters. If the 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, Jennifer, Susan Fournier, and Adam S. Brasel (2004), "When Good Brands Do Bad," *Journal of Consumer Research*, 31 (June), 1-18.
- Allenby, Greg M. and James L. Ginter (1995), "Using Extremes to Design Products and Segment Markets," *Journal of Marketing Research*, 32 (November), 392-403.
- Allenby, Greg M., Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of American Statistics Association*, 93 (446), 365-374.
- Andrews, Rick L. and Imran S. Currim (2003), "A Comparison of Segment Retention Criteria for Finite Mixture Models," *Journal of Marketing Research*, 39 (May), 235-243.
- Arnett, Dennis B., Steve German and Shelby D. Hunt (2003), "The Identity Salience Model of Relationship Marketing Success: The Case of Nonprofit Marketing," *Journal of Marketing*, 67 (2), 89-105.
- Atchadé, Yves F. (2006), "An Adaptive Version for the Metropolis Adjusted Langevin Algorithm with a Truncated Drift," *Methodology and Computing in Applied Probability*, 8 (2), 235-254.
- Baade, Robert A. and Jeffrey O. Sundberg (1996), "What Determines Alumni Generosity?" *Economics of Education Review*, 15 (1), 75-81.
- Böckenholt, Ulf and William R. Dillon (2000), "Inferring Latent Brand Dependencies," *Journal of Marketing Research*, 37 (February), 72-87.
- Böckenholt, Ulf and Rolf Langeheine (1996), "Latent change in recurrent choice data," *Psychometrika*, 61 (2), 285-302.
- Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: the Role of Satisfaction," *Marketing Science*, 17 (1), 45-65.
- Bucklin, Randolph E. and James M. Lattin (1991), "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, 10 (1), 24-39.
- Bult, Jan R. and Tom Wansbeek (1995), "Optimal selection for direct mail," *Marketing Science*, 14 (4), 378-394.
- Dekimpe, Marnik G. and Dominique M. Hanssens (2000), "Time-series Models in Marketing: Past, Present and Future," *International Journal of Research in Marketing*, 17 (2-3), 183-193.
- Dillon, William R., Ulf Böckenholt, Melinda Smith de Borrero, Ham Bozdogan, Wayne DeSarbo, Sunil Gupta, Wagner Kamakura, Ajith Kumar, Venkatram Ramaswamy, and Michael Zenor (1994), "Issues in the Estimation and Application of Latent Structure Models of Choice," *Marketing Letters*, 5 (4), 323-334.
- Djurić, Petar M. and Joon-Hwa Chun (2002), "An MCMC Sampling Approach to Estimation of Nonstationary Hidden Markov Models," *IEEE Transaction on Signal Processing*, 50,(5), 1113-1123.

- Du, Rex and Wagner A. Kamakura (2006), "Household Lifecycles and Lifestyles in America," *Journal of Marketing Research*, 43 (February), 121-132.
- Dwyer, Robert F., Paul H. Schurr, and Sejo Oh (1987), "Developing Buyer-Seller Relationships," *Journal of Marketing*, 51 (April), 11-27.
- Erdem, Tülin and Baohong Sun (2001) "Testing for Choice Dynamics in Panel Data," *Journal of Business and Economic Statistics*, 19 (2), 142-152.
- Fader, Peter S., Bruce G.S. Hardie and Chun-Yao Huang (2004), "A Dynamic Change-point Model for New Product Sales Forecasting," *Marketing Science*, 23 (1), 50-65.
- Fader, Peter S. and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12 (3), 304-317.
- Flanagan, John C. (1954), "The Critical Incident Technique," *Psychological Bulletin*, 51 (4), 327-359.
- Fournier, Susan (1998), "Consumers and Their Brands: Developing Relationship Theory in Consumer Research," *Journal of Consumer Research*, 24 (March), 343-373.
- Frank, Roland E. (1962), "Brand Choice as a Probability Process," *Journal of Business*, 35 (January), 43-56.
- Gelman, Andrew and Donald B. Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457-511.
- Greene, William H. (1997), *Econometric Analysis*, Prentice Hall, Third edition.
- Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203-238.
- Gupta, Sunil and Valarie A. Zeithaml (2006), "Customer Metrics and Their Impact on Financial Performance," *Marketing Science*, 25 (6), 718-739.
- Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57 (2), 357-384.
- Harrison, William B., Shannon K. Mitchell, and Steven P. Peterson (1995), "Alumni Donations and Colleges' Development Expenditures: Does Spending Matter?" *The American Journal of Economics and Sociology*, 54 (October), 397-413.
- Heckman, James J. (1981), "Heterogeneity and State Dependence," in S. Rosen ed. *Studies in Labor Markets*, Chicago: University of Chicago Press, 91-139.
- Hinde, Robert A. (1979), *Towards Understanding Relationships*, London: Academic.
- Hughes, James P. and Peter Guttorp (1994) "A Class of Stochastic Models for Relating Synoptic Atmospheric Patterns to Regional Hydrologic Phenomena," *Water Resources Research*, 30 (5), 1535-1546.
- Jain, Dipak and Siddhartha S. Singh (2002), "Customer Lifetime Value Research in Marketing: A Review and Future Direction," *Journal of Interactive Marketing*, 16 (2), 34-46.

- Jeuland, Abel P., Frank M. Bass, and Gordon P. Wright (1980), "A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes," *Operations Research*, 28 (2), 255-277.
- Kahneman, Daniel and Dale T. Miller (1986), "Norm Theory: Comparing Reality to its Alternatives," *Psychological Review*, 93 (2), 136-153.
- Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379-390.
- Keane, Michael P. (1997), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, 15 (July) 310-327.
- Kim, Chang-Jin and Charles R. Nelson (1999), *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*, M.I.T. Press, Cambridge, Massachusetts.
- Lewis, Michael (2004), "The Influence of Loyalty Programs and Short-term Promotions on Customer Retention," *Journal of Marketing Research*, 41 (August), 281-292.
- Libai, Barak, Das Narayandas, and Clive Humby (2002), "Toward an Individual Customer Profitability Model: A Segment-Based Approach," *Journal of Service Research*, 5 (1), 69-76.
- Liechty, John C., Michel Wedel, and Rik Pieters (2003), "The Representation of Local and Global Scanpaths in Eye Movements Through Bayesian Hidden Markov Models," *Psychometrika*, 68 (4), 519-541
- MacDonald, Iain L. and Walter Zucchini (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*, London: Chapman and Hall.
- McAlister, Leigh, Rajendra Srivastava, Joel Horowitz, Morgan Jones, Wagner Kamakura, Jack Kulchitsky, Brian Ratchford, Gary Russell, Fareena Sultan, Tetsuo Yai, Doyle Weiss, and Russell Winer (1991), "Incorporating Choice Dynamics in Models of Consumer Behavior," *Marketing Letters*, 2 (3), 241-252.
- Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Predicting Online Purchase Conversion Using Web Path Analysis," *Marketing Science*, 23, (4), 579-595.
- Moon, Sangkil, Wagner A. Kamakura, and Johannes Ledolter (2007), "Estimating Promotion Response When Competitive Promotions Are Unobservable," *Journal of Marketing Research*, forthcoming.
- Morgan, Neil A. and Lopo L. Rego (2006), "The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance," *Marketing Science*, 25 (5), 426-439.
- Morgan, Robert M. and Shelby D. Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing*, 58 (3), 20-38.
- Morrison, Donald G., Richard D.H. Chen, Sandra L. Karpis, and Kathryn E.A. Britney (1982), "Modeling Retail Customer Behavior at Merrill Lynch," *Marketing Science*, 1 (2), 123-141.

- Naik, Prasad A., Murali K. Mantrala, and Alan G. Sawyer (1998), "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17 (3), 214-235.
- Netzer, Oded (2004), "A Hidden Markov Model of Customer Relationship Dynamics," Ph.D. dissertation, Stanford University.
- Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, B*, 56 (1), 3-48.
- Okunade, Albert A. (1993), "Logistic Regression and Probability of Business School Alumni Donations: Micro-data Evidence," *Education Economics*, 1, 243-258.
- Okunade, Albert A and Robert L. Berl (1997), "Determinants of Charitable Giving of Business School Alumni," *Research in Higher Education*, 38 (April), 201-214.
- Oliver, Richard L. (1997), *Satisfaction: A Behavioral Perspective on the Consumer*, New York: McGraw-Hill.
- Pauwels, Koen, Imran Currim, Marnik G. Dekimpe, Eric Ghysels, Dominique M. Hanssens, Natalie Mizik and Prasad Naik (2004), "Modeling Marketing Dynamics by Time Series Econometrics," *Marketing Letters*, 15 (4) 167-183.
- Pfeifer, Philip E. and Robert L. Carraway (2000), "Modeling Customer Relationships using Markov Chains," *Journal of Interactive Marketing*, 14 (2), 43-55.
- Poulsen, Carsten Stig (1990), "Mixed Markov and Latent Markov Modeling Applied to Brand Choice Behavior," *International Journal of Research in Marketing*, 7 (1), 5-19.
- Rabiner, Lawrence R. (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, 77 (2), 257-286.
- Rabiner, Lawrence R. and Biing-Hwang 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," *International Journal of Research in Marketing*, 14 (1), 57-80.
- Reinartz, Werner and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (1), 77-99.
- Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304-328.
- Rust, Roland T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relationships," *Marketing Science*, 25 (6), 560-580.
- Rust, Ronald T., Katherine N. Lemon, and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (1), 109-127.
- Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (1), 41-67.
- Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6 (2), 461-464.

- Scott, L. Steven (2002) "Bayesian Methods for Hidden Markov Models, Recursive Computing in the 21st Century," *Journal of the American Statistical Association*, 97 (475), 337–351.
- Smith, Aaron, Prasad A. Naik, and Chih-Ling Tsai (2006), "Markov-switching Model Selection Using Kullback-Leibler Divergence," *Journal of Econometrics*, 134 (2), 553-577.
- Soukup, J. David (1983), "A Markov Analysis of Fund-Raising Alternatives," *Journal of Marketing Research*, 20 (August), 314-319.
- Spiegelhalter, David J., Nicola G. Best., Bradley P. Carlin, and Angelika van der Linde (2002), "Bayesian measures of model complexity and fit" *Journal of the Royal Statistical Society, Series B*, 64(3):583-639
- Srinivasan V. and R. Kesavan (1976), "An Alternative Interpretation of the Linear Learning Model of Brand Choice," *Journal of Consumer Research*, 3 (September), 76-83.
- Storm, Stephanie (2003), "After Years of Cash Flow, Universities Hit an Ebb" 13 March, *The New York Times*.
- Taylor, Alton L. and Joseph C. Martin, Jr. (1995), "Characteristics of Alumni Donors at a Research I Public University," *Research in Higher Education*, 36 (3), 283-302.
- Thomas, Jacquelyn S. (2001), "The Importance of Linking Customer Acquisition to Customer Retention," *Journal of Marketing Research*, 39 (May), 262-268.
- Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Dordrecht: Kluwer.
- Willemain, Thomas R., Anil Goyal, Mark Van Deven and Inderpreet S. Thukral (1994), "Alumni Giving: The Influences of Reunion, Class and Year," *Research in Higher Education* 35 (2), 201-214.
- Winer, Russell S. (2001), "A Framework for Customer Relationship Management," *California Management Review*, 43 (Summer), 89-105.
- Xie, Jinhong X, Michael Song, Marvin Sirbu, and Qiong Wang (1997), "Kalman Filter Estimation of New Product Diffusion Models," *Journal of Marketing Research*, 34 (August), 378-393.