# A HIDDEN MARKOV MODEL OF CUSTOMER RELATIONSHIP DYNAMICS

Oded Netzer*
Columbia University

James M. Lattin
Stanford University

V. Srinivasan
Stanford University

March 2005

* 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: lattin_james@gsb.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).

# A HIDDEN MARKOV MODEL OF CUSTOMER RELATIONSHIP DYNAMICS

ABSTRACT

This paper addresses the issue of modeling and understanding the dynamics of customer relationships. The proposed model facilitates using typical transaction data to evaluate the effectiveness of relationship marketing actions as well as other customer-brand encounters on the dynamics of customer relationships and the subsequent buying behavior. Our approach to modeling relationship dynamics is structurally different from the models in the existing literature.

Theories in behavioral relationship marketing suggest that relationships evolve over time as a consequence of encounters between the customer and the company (or organization). Accordingly, in the proposed model, customer-brand encounters, such as exposure to relationship marketing activities and past buying behavior, may have an enduring impact by shifting the customer to a different (unobservable) relationship state.

We construct and estimate a hidden Markov model (HMM) to relate the latent relationship states to the observed buying behavior. The HMM enables the marketer to assess the evolution of customer relationships over time. Moreover, since the relationship states are determined, in part, by exposure to marketing actions, it is possible to examine methods by which the firm can alter the customer's relationship level and consequently affect the long-term buying behavior. To account for unobserved heterogeneity across customers, we specify a random-effect model estimated using a hierarchical Bayes procedure.

We calibrate the proposed model using simulated data, as well as using longitudinal alumni gift giving data. This empirical application demonstrates the value of the proposed model in understanding the dynamics of alumni-university relationships and predicting donation behavior. Using the proposed model, we are able to identify three relationship states, probabilistically classify the alumni base into these different states, and estimate the marginal impact of different interactions between the alumni and the university on moving the alumni between these states. The application of the model for marketing program decisions is illustrated using a "what-if" analysis of a reunion attendance marketing campaign. Additionally, using a validation sample, we show that the proposed model improves the ability to predict future donations relative to several benchmark models.

# 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. A study by Jupiter Media Metrix predicted that CRM spending in the U.S. will rise to $16.5 billion in 2006, up from $9.7 billion in 2001 (*Jupiter Media Metrix* – press release 2002). Despite the wide adoption of CRM in the business world, the academic community has been lagging in developing models that could help businesses analyze transaction data to assess customer relationships and put forward a support system for marketing decisions. Recently, marketing modelers started to address this gap by developing models of customer lifetime value (e.g., Blattberg and Deighton 1996; Gupta, Lehmann, and Stuart 2002; Reinartz and Kumar 2003). However, far less attention has been given to modeling the *dynamics* of customer relationships and the effect of relationship marketing actions on customer-brand relationships and the customer's choice behavior.

Firms invest heavily in developing relationship-oriented marketing activities to enhance customers' loyalty. Various relationship marketing activities have emerged recently. For example:

- *Harrah's Entertainment's* Total Reward loyalty program is considered a success story in terms of developing a strong tie between the casino and its gamblers. Similar to other loyalty programs, the idea behind Total Reward is to increase Harrah's share of wallet (the share of the customer's total gambling done in Harrah's casinos). Total Reward includes 25 million members, which generate half of Harrah's Las Vegas revenues (*Brandweek* 2002).

- *Stanford University* launched in November 2001 a 12-city tour of a Hollywood-style video centerpiece called "Think Again." Think Again brings the life of students and faculty at Stanford to the University's alumni across the U.S. The short-term goal of Think Again is to help raise $1 billion. The long-term goal is to strengthen the alumni-university network (*San Jose Mercury News* 2002).

- *Amazon.com* has developed a "learning" agent that recommends music and books to the site's customers based on their past purchases and stated preferences. Using the recommendation system, Amazon.com hopes to create higher switching costs and increase customers' loyalty.

These three relationship-marketing tools share something in common; they are all aimed at creating an enduring impact on the relationship between the customers and the brand[1] and on the customers' buying behavior. This type of marketing activity is more than just a short-term incentive to choose (such as a temporary price change or a special promotional display) designed to have its principal impact on behavior at the point of purchase. In each of the three cases above, marketers engage in activity 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 change the nature of subsequent interactions.

The objective of this research is to capture the dynamics of customer relationships. We suggest a modeling framework for estimating and understanding these dynamics – one that captures the impact of the series of customer-brand interactions in a way that it is structurally different from the models in the existing literature. The proposed model allows one to probabilistically identify the customer's state of relationship at any given time and enables comparing the impact of the alternative customer-brand encounters on moving the customer to a higher state of relationship.

We propose a simple hidden Markov model in which the Markovian states are a finite set of relationship states. The transitions between the states are determined by the history of interactions between the customer and the brand. The *relationship-state dependence* is defined by the effect of the relationship state on the likelihood of the customer's purchase behavior. The number of relationship states is determined by the complexity of dynamics in the data at hand. To distinguish between relationship-state dependence and zero-order heterogeneity (c.f., Fader and Lattin 1993), unobserved heterogeneity is captured through a set of random-effect coefficients. The individual-level parameters are estimated using a Markov Chain Monte Carlo hierarchical Bayes procedure.

---

[1] Since the model proposed in this paper applies to the customer's relationship with brands of consumer goods, services, or non-profit organizations, we use the terms "brand" and "firm" interchangeably to describe the entity the customer develops the relationship with.

As a first step, we apply the model to simulated data. The Monte Carlo experiment reveals that the HMM can capture well the true dynamic behavior and is able to identify the customer's true state of relationship over time. Next, we apply the model to a university-alumni customer relationship data set. This empirical application stresses the value of the model for CRM marketers. Analyzing the model's estimates can help marketers to evaluate the effectiveness of different marketing actions on moving the customer to a higher relationship level with the firm. We use a "what-if" analysis to demonstrate the use of the dynamic classification of alumni into relationship states to choose the most effective targeting strategy in the context of a direct marketing campaign.

The reminder 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 account for unobserved heterogeneity using hierarchical Bayes estimation procedure. In section 4 we demonstrate the application of the model using simulated data. Section 5 describes 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 6 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; Fournier and Yao 1997) and from the empirical modeling perspective (e.g., Bolton 1998; Bolton and Lemon 1999; Thomas 2001).

Several theoretical models have been suggested for the evolution of relationships over time (e.g., Dwyer, Schurr and Oh 1987; Ford 1980). These studies suggest that relationships evolve (not always monotonically) through several discrete levels. In particular, it has been 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 and 1999) suggests that a discrete shift in the relationship occurs if the

4

aggregate satisfaction from a sequence of critical incidents is strong enough to move the customer to a different "conceptual plane" of loyalty. Thus, the transitions between the relationship stages may be triggered by discrete encounters between the relationship parties. For example, offering an airline traveler an upgrade to business class could serve as a critical incident (Flanagan 1954). 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, if the act of upgrade itself and the experiences of the traveler during the business class flight pass the customer's satisfaction threshold.

Overtime, a sequence of critical encounters between the customer and the brand is postulated to become a long-term relationship using behavioral bonds such as commitment and trust (Hinde 1979; Dwyer, Schurr and Oh 1987; Morgan and Hunt 1994). Accordingly, we define a *relationship* as a sequence of discrete encounters between the customer and the brand. Such encounters include: transactions, service encounters, or exposure and response to marketing actions initiated by the firm.

From the managerial point of view, the interest in customer relationships stems from the link between customer relationships and the idea of customer equity and customer lifetime value. The notion of the customer base as an asset to the firm (Rust, Zeithaml and Lemon 2000) is the main driver behind the industry shift towards CRM. Several methods have been suggested for evaluating customer lifetime duration (e.g., Allenby, Leone and Jen 1999; Bolton 1998; Reinartz and Kumar 2003; Schmittlein and Peterson 1994, Thomas 2001) and customer lifetime value (Blattberg and Deighton 1996; Gupta, Lehmann and Stuart 2002; Rust, Lemon and Zeithaml 2004). With the exception of Reinartz and Kumar (2003), these models do not take into consideration the dynamics in the long-term effect of the interactions between the customer and the brand, which is the main focus of the current study. Indeed, in their review of customer lifetime value models, Jain and Singh (2002) suggest incorporating factors that drive consumer behavior, like marketing activities, in the model of customer lifetime value. In order to capture the dynamics in relationships, one must capture the drivers of these dynamics that are both internal and external to the customer.

## 2.2    Customer-Brand Interactions and the Dynamics in Buying Behavior

Probably the most prominent interactions between the customer and the brand are purchases. In the choice modeling literature, the impact of past behavior on subsequent behavior is commonly known as *state dependence* (Heckman 1981). Choice modelers include state dependence in their

econometric models to capture cross-individual heterogeneity as well as the serial correlation in purchases over time (McAlister et al. 1991; Meyer et al. 1997). The common measures of state dependence have been an exponentially smoothed sum (Guadagni and Little 1983; Srinivasan and Kesavan 1976) or a simple average (Bucklin and Lattin 1991) of past purchases.[2] When the state dependence is significant, it implies a non-zero order buyer behavior process (Bass et al. 1984). While several studies suggested that for many packaged goods the buying behavior follows a zero-order process (e.g., Bass et al. 1984), recent studies found positive and significant state dependence effects across product categories even after controlling for cross-individual heterogeneity (Erdem and Sun 2001; Keane 1997; Seetharaman, Ainslie and Chintagunta 1997).

A limitation of current state dependence 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 most state dependence models is that they often ignore another important source of non-stationarity in buying behavior, namely the enduring effects of marketing stimuli. Indeed, for marketing 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).

The problem of overlooking the enduring impact of marketing actions is likely to be more severe in the context of relationship marketing activities. Relationship marketing actions such as loyalty programs (Bolton, Kannan and Bramlett 2000; Sharp and Sharp 1997) direct mailing, and personalized coupons (Wulf, Odekerken-Schröder and Iacobucci 2001), are put forth with the objective of changing customers' relationships with the brand, and therefore may have an enduring effect on customers' buying behavior. The different studies that have investigated the effect of marketing activities on customer relationships, differ with respect to the incorporation of heterogeneity (aggregate vs. individual level), the measure of behavioral loyalty (observed vs. stated), and the marketing activities included (monetary, social, or structural solution). These studies provided mixed results with respect to the enduring impact of relationship marketing activities.

A limitation of most of the studies that measured the effect of relationship marketing actions is that these studies are static. That is, the researcher collected measures for each individual at only

---

[2] Negative state dependence is commonly referred to as "variety seeking" (McAlister and Pessemier 1982).

one point in time.  Therefore, it is difficult to elicit from such studies the effect of relationship

marketing activities on the dynamics of customer relationships.  Bolton and Lemon (1999) and

Verhoef (2002) demonstrated some (albeit limited) relationship dynamics using a two-period model.

Both studies suggested stronger effects for past loyalty measures in comparison to the effects of

marketing activities.  To capture the multi-period dynamics in behavioral loyalty, one must include

the loyalty measure endogenously into the buyer behavior model, *in addition* to the marketing

variable.  Furthermore, customer heterogeneity must be taken into account.  A model that estimates

the long-term effect of marketing variables, but ignores state dependence, may overestimate the

enduring effect of the marketing variables.

   One of the key concepts in relational marketing is the ability of the marketer to tailor marketing

offerings for each individual in order to enhance the relationship.  Advances in data collection and

statistical methods enable marketers to better track the transactions of their customers over time and

target the marketing mix activities such as customized coupons (Bawa and Shoemaker 1987; Rossi,

McCulloch and Allenby 1996), direct mail (Basu, Basu and Batra 1995; Bult and Wansbeek 1995;

Gönül, Kim and Shi 2000; Gönül and Shi 1998; Gönül and Ter Hofstede 2002), direct e-mail

(Ansari and Mela 2003) and customized pricing (Chen and Iyer 2002; Feinberg, Krishna and Zhang

2002) more effectively.   Perhaps the most commonly used approach to target marketing activities is

the recency, frequency, and monetary value (RFM) approach (e.g., Roberts and Berger 1999).

Alternative approaches include the hazard rate models (Gönül, Kim and Shi 2000), Bayesian hazard

rate model (Gönül and Ter Hofstede 2002), dynamic programming (Gönül and Shi 1998),

generalized Gamma distribution (Allenby, Leone and Jen 1999), random-coefficient choice model

(Rossi, McCulloch and Allenby 1996), and the NBD model (Colombo and Jiang 1999).

   Similar to the studies above, the proposed model exploits the individual history of

transactions in order to target the marketing activities.  However, targeting marketing activities

based on the proposed HMM extends the current targeting methods in several aspects.  First, all

of the studies described above are limited in their ability to capture the enduring impact of the

marketing activity.  For marketing activities such as coupons (and arguably direct mail) the

assumption of merely an instantaneous effect may be justified.  However, as discussed above,

relational marketing activities (e.g., an alumni reunion attendance campaign) may have an

enduring impact on customers' future choices. The dynamic nature of the HMM allows incorporating such long-term effects of the targeted marketing action. Second, most models of direct marketing suggest a static model of choice, which may lead to erroneous estimates of the impact of the targeted marketing activity if dynamics in the buying behavior exist. Finally, by estimating the latent relationship states, this research takes a relationship approach for target marketing. The objective of enhancing customer relationships may lead to different targeting than the commonly used objective of profit maximization over a short or medium horizon.

To summarize, the proposed model, described in the next section, makes several important contributions to the dynamic choice modeling and relationship marketing literature. First, the definition of "*relationship-state dependence*" as the effect of the customer's relationship state on his/her buying behavior ties the notion of state dependence to the idea of the customer-brand interaction, which is the building block of relational behavior. Second, the relationship states are determined by multiple modes of customer-brand interactions. Similar to state dependence, the relationship-state dependence defines the effect of past actions on the current decision behavior. However, the states are determined not only by past purchases, but also by other customer-brand encounters, like marketing actions, which may be initiated by the firm. Thus, the definition of the relationship-state dependence broadens the common one-dimensional definition of state dependence, which is based solely on lagged choices. Thirdly, the proposed model allows for a more flexible state structure by estimating the number of states based on the complexity of dynamics in the data, rather than the ad hoc structure commonly used to define state dependence. Fourthly, this is one of the very few models that simultaneously accounts for the effects of state dependence, the enduring effect of marketing stimuli, and cross-individual heterogeneity. As suggested above, omitting any of these three components from a choice model may result in biased estimates of the other effects. Finally, as demonstrated in the "what-if" analysis in §5.7, estimating the enduring impact of marketing activities at the individual-level opens an opportunity for a more effective direct marketing strategy.

## 3. MODEL DEVELOPMENT

### 3.1    The Hidden Markov Model

The hidden Markov model (HMM) described in this section is an individual-level model of buying behavior. We consider a set of customers each of whom are 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 (e.g., scanner panel data).

A *relationship encounter* is defined as an interaction between the customer and the brand. Such interactions may include purchase transactions, exposure to relationship marketing activities, or other non-purchase related exposure to the brand. Relationships are then constructed by a longitudinal sequence of relationship encounters. Accordingly, we 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 brand and the conditional likelihood of choice given the relationship state. The stochastic relationship-states model can be characterized by a first-order random walk model. The transitions between the Markovian states are probabilistically determined and are affected by relationship encounters. However, the researcher does not observe the actual relationship state membership. At any given time, the researcher only observes the actual buying behavior. This structure of latent states and observed behavior could be modeled by a HMM.

A HMM is a model of stochastic process that is not directly observable, but can only be observed through another set of stochastic processes that produces a set of observations. In the proposed model, the hidden Markov process is a random walk that describes the relationship states' transitions. This stochastic process is then transformed into the observed buyer behavior through the stochastic process of choice (e.g., a binary choice model). 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). The simplest application of HMMs in speech recognition includes estimating a HMM using training data for each speech unit (e.g., word), and then using it to probabilistically identify any stream of acoustic sequence. In econometrics, Hamilton (1989) proposed a HMM to estimate the impact of discrete regime shifts on the growth rates of real GNP.

Within the marketing literature, HMMs are mostly 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 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).[3] Wedel and Kamakura (2000, chapter 10) and Dillon et al. (1994) survey these studies, as well as alternative forms of dynamics in segmentations (e.g., Böckenholt and Dillon (2000); Böckenholt and Langeheine (1996)). Wedel and Kamakura conclude that the issue of non-stationarity in marketing segmentation should be further investigated.

The HMM of customer relationships that we develop subsequently pushes forward the marketing literature related to dynamic latent class models in several aspects. First, since we are interested in understanding the means by which the firm could influence these dynamics we relax the commonly used assumption of stationary transitions between the latent states. Specifically, we allow the Markovian transitions between the states to be a function of different interactions between the customer and the brand. Additionally, in the CRM environment, interactions between the customer and the brand may be targeted at the individual level. Consequently, we extend the current dynamic latent class models by allowing for cross customer heterogeneity in the transitions between the latent states. Capturing the heterogeneity in the transition matrix also mitigates the possible confound between time dynamics and cross sectional heterogeneity. Finally, the application of the HMM in the context of alumni university relationships allows us to capture up to 26 possible choices as well as transitions between the latent states, in comparison to up to five transitions in Poulsen's application and only one possible transition in Ramaswamy's model.

The HMM we develop relates the relationship-states Markov chain process to the observed buying behavior. The conditional choice probabilities, given the states, are restricted to be increasing in the state's level (i.e., number) to ensure identification of the underlying relationship states. In this study, we describe the most simple HMM of repeated binary choice (see Figure 1 for a graphical representation of the proposed HMM). The model could be easily extended to multinomial choice or quantity decisions.

## Figure 1 - A Hidden Markov Model of Customer Relationships

---

[3] An alternative recent application of HMMs in marketing used semi-HMMs to study web path analysis (Montgomery et al 2004; Li, Liechty, and Montgomery 2002).

*Relationship States*     *State Dependent Choice*
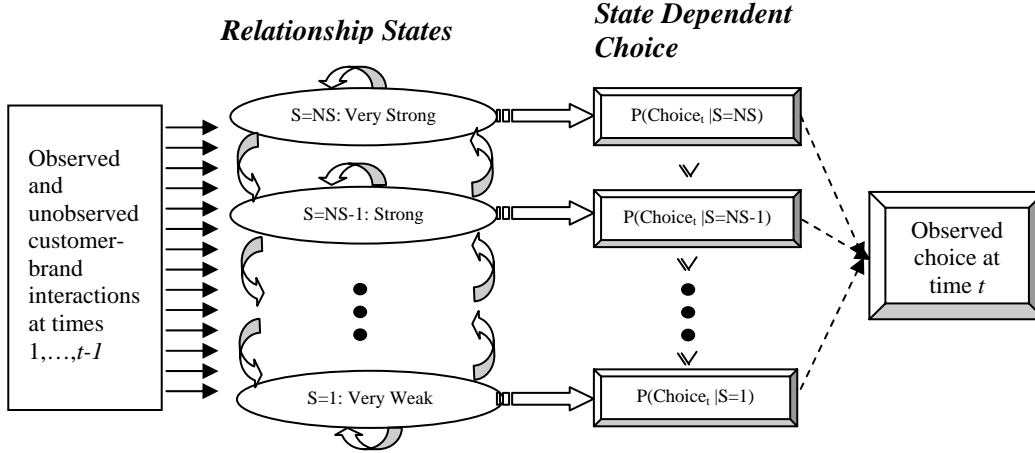
The proposed HMM consists of three main components: (1) The *initial state distribution* ($\boldsymbol{\pi}$), (2) a sequence of random walk *transitions* ($\mathbf{Q}$) that express, in a probabilistic manner, the likelihood that the series of customer-brand interactions in the previous time period were strong enough to transition the customer to an adjacent state, and (3) a vector of *state dependent choice* probabilities given the relationship states ($\mathbf{m}$).[4] The probability of observing a sequence of choices is defined by:

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

where,

$S_{it}$ is customer $i$'s state at time $t$ in a Markov process with NS states, and

$Y_{it}$ is customer $i$'s choice at time $t$.

Following Equation (1) the three components of the HMM can be defined as follows:

1) *The initial state distribution* – the probability that customer $i$ is in state $s$ at time 1 can be defined as:

   $P(S_{i1} = s) = \pi_{is}$.

2) *The transitions* – the probability that a customer transitions from state $s_{t-1}$ at time $t$-1 to state $s_t$ at time $t$ can be defined as: $P(S_{it} = s_t \mid S_{it-1} = s_{t-1}) = q_{is_t, s_{t-1}}$.

3) *The state dependent choice* - the probability that the customer will choose the product at time $t$ conditioned on her state can be defined as: $P(Y_{it} = 1 \mid S_{it} = s_t) = m_{is_t}$.

According to these definitions we can rewrite Equation 1 as:

---

[4] In what follows, we describe the state dependent choice as a binary choice, where no information is available about the competition. We take this approach since this is generally the case in CRM transaction data sets, where the firm does not observe transactions with the competition. Nevertheless, if information about the competition exists, one could easily extend the model by incorporating these data as covariates in the model.

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

## 3.2    The Model's Components

### 3.2.1    The Markov Chain Transition Matrix

We model the transitions between states as a random walk process. The random walk is a special case of a Markov process in which at each transition it is only possible to move to an adjacent state or stay in the current state. Behaviorally this means that at each time period, the customer can stay in the current relationship state or move to an adjacent relationship state.[5] The random walk transition matrix is defined as,

State at t

| State at t-1 | 1 | 2 | 3 | $\cdots$ | $NS-1$ | $NS$ |
|---|---|---|---|---|---|---|
| 1 | $q_{it11}$ | $q_{it12}$ | $0$ | $\cdots$ | $0$ | $0$ |
| 2 | $q_{it21}$ | $q_{it22}$ | $q_{it23}$ | $\cdots$ | $0$ | $0$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $NS$ | $0$ | $0$ | $0$ | $\cdots$ | $q_{itNSNS-1}$ | $q_{itNSNS}$ |

$\mathbf{Q_{it\text{-}1t}} = $ (the above matrix),    (2)

where, $q_{itss'} = P(S_{it} = s' \mid S_{it-1} = s)$ is the conditional probability that individual $i$ moves to state $s'$ at time $t$ given in state $s$ at time $t$-1, and where $0 \leq q_{itss'} \leq 1$ $\forall s, s'$ and $\sum_{s'} q_{itss'} = 1$. Following the random walk specification, all terms that are two states or more away from the diagonal are set to zero.

Each one of the matrix elements in Equation (2) represents a probability of transition to an adjacent state or a probability of staying in the current state. We assume that underlying this probabilistic representation, there is a continuous measure of propensity for transition, which is affected by the aggregate impact of relationship encounters. 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, Olivia, Oliver and MacMillan 1992).

---

[5] The random walk assumption was made merely to keep the model parsimonious. This assumption could be easily relaxed by allowing positive probabilities for transitions to non-adjacent states. Indeed, we estimated a more flexible model that allows for positive transition probabilities to state 1 from every state. The performance of this model was similar to that of the parsimonious random-walk model.

The norm theory of Kahneman and Miller (1986) postulates that past experiences create a norm, against which current experiences are judged. Thus, following the Markovian process, current relationship encounters are judged relative to the status quo, as characterized by the current relationship state, which may in turn change the propensity for transition from the current state. If the cumulative experience from the encounters between the customer and the brand is highly negative (e.g., service failure), it is likely to shift the propensity for transition value 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 value above the threshold needed for a transition to a higher state. If the cumulative effect of 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, we define $TR_{its}$ as individual $i$'s propensity for transition from state $s$ at time $t$. The propensity for transition is a function of the systematic effect of relationship encounters and factors that are unobservable to the researcher. The overall propensity for transition from state $s$ can be written as,

$$TR_{its} = \mathbf{A_{it}}\boldsymbol{\rho_s} + \varepsilon_{its} \ , \tag{3}$$

where,

$\boldsymbol{\rho_s}$     is the effect of relationship encounters on the propensity for transition from state $s$,

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

$\varepsilon_{its}$     is a random error which represents the effect of unobserved factors on the transition propensity.

Additionally, we define two state-specific threshold terms: one for movement to a higher state ($\mu(hi)_s$), and one for dropping down to a lower state ($\mu(lo)_s$). A transition occurs if the cumulative impact of the encounters at time $t$ is strong enough to pass the threshold needed to create a transition. Figure 2 compares two distributions of $TR_{its}$ for two values of $\mathbf{A_{it}}$. The expected value of $TR_{its}$ in curve C2 is higher than the expected value of $TR_{its}$ in curve C1 (this could be a result of a positive encounter between the customer and the brand; for example, an advertisement exposure).

Following the positive encounter, the likelihood that the propensity for transition would cross the threshold for transition to a higher state (crossing $\mu(hi)_s$) in curve C2 is higher than that of curve C1.

**Figure 2 - The Distribution of the Propensity for Transition**
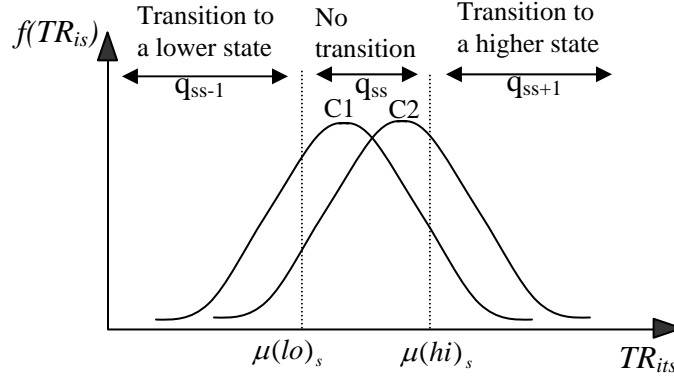


Assuming that the unobserved component of the propensity for transition, $\varepsilon_{its}$ in Equation (3), is independently and identically distributed (IID) of the extreme value type, then the probability that individual $i$ will transition from state $s$ to state $s$-1, stay in state $s$, or transition to state $s$+1, follows an ordered logit model (Greene 1997). Specifically, the terms $q_{itss'}$, $s' \in \{s$-1, $s$, $s$+1$\}$ in the transition matrix in Equation (2) could be written as:

$$q_{itss-1} = \Pr(\text{transition from } s \text{ to } s\text{-1}) = \frac{\exp(\mu(lo)_s - TR^*_{its})}{1 + \exp(\mu(lo)_s - TR^*_{its})}, \tag{4}$$

$$q_{itss} = \Pr(\text{staying in } s) = \frac{\exp(\mu(hi)^*_s - TR^*_{its})}{1 + \exp(\mu(hi)^*_s - TR^*_{its})} - \frac{\exp(\mu(lo)_s - TR^*_{its})}{1 + \exp(\mu(lo)_s - TR^*_{its})}, \tag{5}$$

$$q_{itss+1} = \Pr(\text{transition from } s \text{ to } s\text{+1}) = 1 - \frac{\exp(\mu(hi)^*_s - TR^*_{its})}{1 + \exp(\mu(hi)^*_s - TR^*_{its})}, \tag{6}$$

where, for $s \in \{2,\ldots,NS\text{-1}\}$,

$TR^*_{its}$ is the deterministic component of $TR_{its}$, and $\mu(hi)^*_s = \mu(lo)_s + \exp(\mu(hi)_s)$ to guarantee that the higher threshold is larger than the lower threshold.

From state 1 and state NS the customer could transition to only one other state. Thus, the transition from state 1 to 2 ($q_{it12}$) is defined by Equation (6), and the probability of staying in state 1 is simply: $1$-$q_{it12}$. Similarly, the transition from state $NS$ to state $NS$-1 is defined by Equation (4), and the probability of staying in state NS is simply: $1$-$q_{itNSNS-1}$. To keep the model parsimonious, the parameters of transitions from states $s \in \{2,\ldots,NS\text{-1}\}$ are restricted to be identical for all $s$. Thus, overall there are three sets of transition parameters: the parameters of the transition from state 1

$(\mathbf{\rho_1}, \mu(hi)_1)$, one set of parameters that represents the transitions from states $2,\ldots,$NS-1 $(\mathbf{\rho_s}, \mu(lo)_s, \mu(hi)_s)$, and the parameters that determine the propensity for staying in the highest state, state NS $(\mathbf{\rho_{NS}}, \mu(lo)_{NS})$. Under this specification, the transition matrix $\mathbf{Q_{it\text{-}1t}}$ in Equation (2) is characterized by $4+3 \times Na$ parameters $\forall NS \geq 3$ ($2+2 \times Na$ parameters for $NS{=}2$), where $Na$ is the dimensionality of the vector $\mathbf{A_{it}}$.

### 3.2.2    The Initial State Distribution

For HMMs with stationary 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, no such stationary distribution exists in the HMM proposed above. Instead, we use the stationary distribution at the mean of the time-varying covariates.[6] The initial state distribution is analytically derived from solving the equation $\mathbf{\pi} = \mathbf{\pi}\overline{\mathbf{Q}}$, where $\overline{\mathbf{Q}}$ is the transition matrix at the mean of the covariates.

### 3.2.3    The State Dependent Choice

If the relationship states at every time interval were observed, one could have directly estimate the Markov model defined by the transition matrix in Equation (2). However, such data would require quantifying each one of the relationship states, and collecting individual-level measures of relationship at every time interval. Such data are rarely available. As a result researchers often use observed behavior to define a set of Markovian states. For example, Pfeifer and Carraway (2000) used the recency since last purchase as their measure of the customer's state. Soukup (1983) used an ad hoc dichotomization of past donations (donor, non contributor) to define the customer's state of donation behavior. Following the HMM approach, we infer the latent relationship states from the observed longitudinal buying behavior and marketing data. Thus, given relationship state $s$, we model the probability of a dichotomous choice following the well-known binary logit model,

$$m_{it|s} = \frac{\exp(\tilde{\beta}_{0s} + \mathbf{x'_{it}}\mathbf{\beta_s})}{1 + \exp(\tilde{\beta}_{0s} + \mathbf{x'_{it}}\mathbf{\beta_s})} \; ; \qquad s{=}1,\ldots,NS, \qquad (7)$$

where,

---

[6] 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 5, the two approaches yielded very similar results. In general, one should use the stationary distribution of the transition matrix with the covariates set to zero if the data set is not left truncated (i.e., we observe the initial interaction between the customer and the brand), and the stationary distribution of the transition matrix at the mean of the covariates otherwise.

$\tilde{\beta}_{0s}$ 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 $\mathbf{m_{it}}$ is, $\mathbf{m'_{it}} = [m_{it|s=1}, m_{it|s=2}, ..., m_{it|s=NS}]$.

The distinction between the vectors of covariates to be included in the transition matrix ($\mathbf{A_{it}}$) and in the state dependent choice vector ($\mathbf{x_{it}}$) is noteworthy. In the transition matrix one should include covariates that are hypothesized to have an enduring impact on the customer's relationship with the brand (e.g., advertisement, service encounters, or targeted marketing activities) and consequently have an enduring impact on the customer's buying behavior. On the other hand, the covariates in the state dependent choice vector are assumed to primarily have an immediate effect on the customer's choice given the relationship state (e.g., price and display promotions).

One restriction that is imposed on the binary choice model, to ensure identification of the states, is that the choice probabilities are 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 $\tilde{\beta}_{01} \le \tilde{\beta}_{02} \le ... \le \tilde{\beta}_{0NS}$ is imposed by,

$$\tilde{\beta}_{0s} = \beta_{01} + \sum_{s'=2}^{s} \exp(\beta_{0s'}); \quad s=2,...,NS. \tag{8}$$

## 3.3 The Likelihood of an Observed Sequence of Choices

The assumption of a zero-order process, commonly assumed in choice models, simplifies the likelihood of sequential choice to a simple product of the individual choice likelihoods. However, such a simplification *cannot* be employed in the case of HMMs. 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 (Equation 1a) sums over all possible routes the individual could take over time between the underlying states.

A problem with Equation (1a) is that it has $NS^T$ elements and is therefore computationally intractable for even modest values of T. Following a slight re-arrangement (MacDonald and Zucchini 1997), we can re-write Equation (1a) in a more functional matrix products form:

$$L_{iT} = P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}...Y_{iT} = y_{iT}) = \boldsymbol{\pi_i}(\tilde{\mathbf{m}}_{i1} \otimes \mathbf{I_{NS}})\mathbf{Q_{i12}}(\tilde{\mathbf{m}}_{i2} \otimes \mathbf{I_{NS}})...\mathbf{Q_{iT\text{-}1T}}(\tilde{\mathbf{m}}_{iT} \otimes \mathbf{I_{NS}})\mathbf{1'}, \tag{9}$$

where,

16

$\tilde{m}_{it|s} = m_{it|s}{}^{y_{it}} (1 - m_{it|s})^{(1-y_{it})}$ and $\tilde{\mathbf{m}}'_{\mathbf{it}} = [\tilde{m}_{it|s=1}, \tilde{m}_{it|s=2}, ..., \tilde{m}_{it|s=NS}]$,

$\mathbf{I_{NS}}$ is an identity matrix of dimension NS, and $\mathbf{1}$ is a NS×1 vector of ones.

The log-likelihood function across individuals is simply the sum of the logarithms of Equation (9) over $i \in \{1,...,N\}$.

### 3.4 Recovering the State Membership Distribution

An attractive feature of the HMM is the ability to probabilistically recover the individual's state at any given time period. This measure could be directly derived from the likelihood function in Equation (9). Two approaches have been suggested for recovering the state membership distribution: "filtering" and "smoothing" (see Hamilton 1989 for a discussion). The first approach utilizes only the information known up to time $t$ to recover the individual's state at time $t$, while the later approach 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:

$$P(S_{it} = s \mid Y_{i1}, Y_{i2}, ..., Y_{it}) = \boldsymbol{\pi_i}(\tilde{\mathbf{m}}_{\mathbf{i1}} \otimes \mathbf{I_{NS}})\mathbf{Q_{i12}}(\tilde{\mathbf{m}}_{\mathbf{i2}} \otimes \mathbf{I_{NS}})...\mathbf{Q_{it\text{-}1t\cdot s}}\tilde{m}_{it|s} / L_{it}, \qquad (10)$$

where,

$\mathbf{Q_{it\text{-}1t.s}}$ is the $s^{th}$ column of the transition matrix $\mathbf{Q_{it\text{-}1t}}$, and,

$L_{it}$ is the likelihood of the observed sequence of choices up to time $t$ from Equation (9).

### 3.5 Estimation Procedure

In this section, we describe the procedure used to estimate the model described above. In choosing the estimation procedure, we focus on properly accounting for observed and unobserved heterogeneity. As mentioned previously, 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 were not correlated over time. Similarly, a model that accounts for heterogeneity but ignores state dependence may overestimate the degree of heterogeneity (Keane 1997). The model described in this section addresses the second problem by offering a flexible specification of state dependence. However, the parameters of the

HMM, as described in Equations (2)-(8), do not vary across individuals. In incorporating heterogeneity into the HMM, we rely on significant improvements achieved in this area in the past decade. In recent years, the hierarchical Bayes approach has become the state-of-the-art method for accounting for observed and unobserved heterogeneity (Rossi and Allenby 2003).

The vector of conditional probabilities $\mathbf{m_{it}}$ already incorporates a discrete layer of heterogeneity similar to the latent class models (Kamakura and Russell 1989). However, unlike the traditional latent class model, the probability of state membership in the proposed HMM is dynamic. To allow for heterogeneity in the individual's propensity for dynamics, we incorporate random-effect parameters in the transition matrix $\mathbf{Q_{it\text{-}1t}}$. We incorporate heterogeneity by allowing the threshold parameters ($\mu(lo)_s$ and $\mu(hi)_s$) in Equations (4)-(6) to vary across individuals.[7] This specification allows for heterogeneity in the "stickiness" of different states, since the distance between the low and high thresholds could vary across individuals. We define normal priors for the threshold parameters, such that $\boldsymbol{\theta_i} = \{\mu(hi)_{i1}, \mu(lo)_{is}, \mu(hi)_{is}, \mu(lo)_{iNS}\}$ has a prior distribution of $\boldsymbol{\theta_i} \sim MVN(\overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma_\theta})$. The heterogeneity in the transition matrix also implies heterogeneity in the initial state distribution, since this distribution is determined by the stationary distribution of the transition matrix at the mean of the covariates.

Observed individual characteristics, such as demographics, can be introduced into the model in a hierarchical manner (cf. Allenby and Ginter 1995). Thus, the heterogeneous parameters vector ($\boldsymbol{\theta_i}$) can be written as a function of observed individual characteristics,

$$\boldsymbol{\theta_i} = \mathbf{z_i'}\boldsymbol{\delta} + \boldsymbol{\varepsilon_{\theta i}} \tag{11}$$

where,

$\mathbf{z_i}$    is a vector of observed individual characteristics,

$\boldsymbol{\delta}$    is a matrix that relates the vector $\mathbf{z_i}$ to the parameter vector $\boldsymbol{\theta_i}$, and

$\boldsymbol{\varepsilon_{\theta i}}$    is the unobserved heterogeneity component, such that, $\boldsymbol{\varepsilon_{\theta i}} \sim N(\mathbf{0}, \boldsymbol{\Sigma_\theta})$.

In the hierarchical Bayes procedure, we set prior conditional distributions for the parameters that vary across individuals and obtain the posterior conditional distributions from the estimation

---

[7] A full random-effect specification of the transition matrix $\mathbf{Q_{it\text{-}t}}$ would not only increase the number of parameters in the model, but also complicate the estimation of random-effect parameters for the parameters of covariates with low variance in the data set. Indeed, this is the case for some of the alumni-university interactions in the empirical application in Section 5.

procedure by pooling data across individuals. Given the set of conditional distributions and priors, we draw recursively from the posterior distribution of the model's parameters. The complete set of conditional distributions and the iteration sequence of the Markov Chain Monte Carlo simulation is described in Appendix A.

## 4. MONTE CARLO SIMULATION

Before we turn to applying the HMM described in §3 to a real world data set, it is instructive to estimate the HMM using simulated data. Estimating the HMM using a Monte Carlo experiment can produce several insights with respect to the ability of the model to capture dynamic behavior, which could not be elicited from most real world empirical applications. First, in a Monte Carlo simulation one can "observe" the latent state membership at any given time. This allows us to test how well the model traces the state membership using the observed choice behavior. Second, in a Monte Carlo experiment it is possible create a long sequence of choices per individual, which allows estimating the models' parameters at the individual level. Estimating individual-level models makes it possible to control for cross-individual heterogeneity and focus the attention on time dynamic effects. Finally, simulating data in a controlled environment overcomes the problem of unobserved effects, which may add noise to the HMM estimation.

### 4.1    The Simulated Data Structure

Following the process described in §3, in creating the sequence of choices, we assume that the dynamics in choices is governed by the customer's Markovian transition between several states of relationship, which correspond to different propensities to choose the brand. We then estimate the model described in §3 for each individual (i.e., without the heterogeneity specifications §3.5). Estimating an individual-specific model ensures that the model captures the "true" time dynamic behavior rather than spurious dynamics due to cross individual heterogeneity.

We construct two sequences of 1000 choices, which correspond to two individuals who differ with respect their dynamic choice behavior: (1) an individual with two relationship states, and (2) an individual with three states of relationship. These two scenarios are governed by the transition and conditional choice probabilities in Table 1.

### Table 1- Two Data Sampling Scenarios

19

| Data Sampling Scenario | Initial state | Transitions | Conditional Choice Probabilities |
|---|---|---|---|
| Two-state scenario | $\pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ | $Q = \begin{bmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{bmatrix}$ | $m = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$ |
| Three-state scenario | $\pi = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ | $Q = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.1 & 0.9 \end{bmatrix}$ | $m = \begin{bmatrix} 0.05 \\ 0.5 \\ 0.95 \end{bmatrix}$ |

The two scenarios described in Table 1 correspond to two different complexity levels in the dynamic behavior. The two-state scenario has only a moderate degree of dynamics. The individual in this scenario transitions between two states of the world: a state in which the individual chooses the brand only 20% of the time and a state in which the individual chooses the brand 80% of the time. The individual in the three-state scenario has three possible states of the world: which corresponds to a choice probability of 5%, 50%, and 95% of the time periods in the low, medium, and high states, respectively. To keep the simulation simple, we assume stationary transition matrices with no covariates. The high diagonal element probabilities (80-90%) are consistent with the strong empirical evidence for state dependence in choice behavior (Keane 1997). We assign both individuals to the lowest relationship state at time period 1. The initial state membership becomes irrelevant when the sequence of observations is long, as in this Monte Carlo analysis.

## 4.2    Estimation and Results

We estimated the HMM using the maximum likelihood procedure MAXLIK in the GAUSS statistical software.[8] For each of the scenarios, we estimated a model with two and three latent states. Indeed, the simulation results demonstrated that the congruent models (the two-state model in the two-state scenario and the three-state model in the three-state scenario) captured the true parameter values fairly well. All the parameter estimates for the congruent models are *not* significantly different from the true parameter values at the 0.05 level. Additionally, the Schwarz Bayesian Information Criterion favored the congruent models over the incongruent models. It is interesting to note that even though the choice space is only two dimensional (a binary choice), a

---

[8] For the full details of the simulation estimation results see Author (2004).

higher dimension of states may be needed to capture the choice process if the process is generated by a Markovian process with a higher state dimensionality.

Using the Monte Carlo simulated data, we can also examine the ability of the HMM to track the transitions between the latent states over time. After estimating the models, we calculated the predicted state distribution using the smoothing approach (see §3.4). The congruent models recovered the underlying states very well (see Figure 3). The hits rates for the state recovery are 83.7% and 82.1% for the two and three state scenarios, respectively.

**Figure 3 - Actual and Predicted State Probabilities for the Two-State Model**[9]
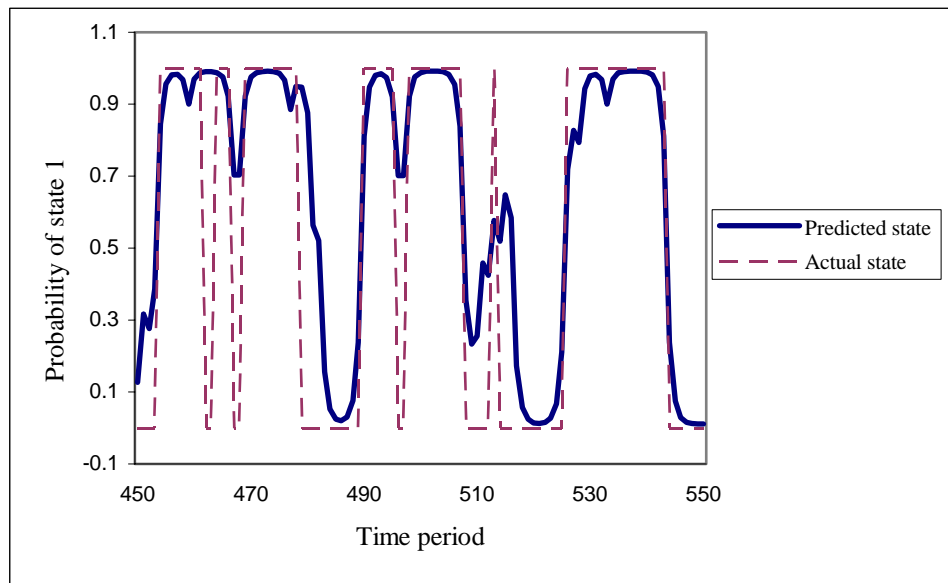


The Monte Carlo experiment provides strong support for the ability of the hidden Markov model to capture the dynamics in the choice process and the transitions between the states for choice data that was generated by a first order latent Markovian process of transitions between latent relationship states. In order to assess the external validity of the model, the next section explores the application of the proposed model to a real world customer relationship environment using secondary data.

---

[9] For ease of readability, we present here the predicted and actual state membership probabilities only for the center time periods 450-550 of the two-state scenario. Similar patterns exist throughout the data for time periods 1-1000 for the two-state and the three-state scenarios.

# 5. EMPIRICAL APPLICATION

In this section, we describe the empirical application of the proposed HMM in the context of the relationship between alumni and their alma mater. We describe, in order, the data set, the alternative models estimated, the estimation results, the comparison of the alternative models in terms of their prediction ability on a validation sample, an illustrative "what-if" analysis that demonstrates the use of the estimated model for strategic decisions, and a survey-based analysis of the behavioral dimension underlying the alumni-university relationship states.

## 5.1 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 to create a marketing decision support system, (i.e., assess what activities are most effective in moving 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 gift giving behavior, due to the strong relationship underlying its construct, would show stronger state dependence than is commonly found in typical scanner panel data applications (Dekimpe, Hanssens and Silva-Risso 1999; Keane 1997). Second, this data set contains most of the components suggested for a good CRM data set. Specifically, it includes the 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 in §6 this data set is somewhat limited in terms of tracking the exposure and response to marketing activities. 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) relate alumni-university relations to the behavioral identity salience model. To our knowledge, no study has previously investigated the dynamics of gift giving behavior and the exogenous and endogenous (to the customer) factors that can alter this dynamic behavior.

## 5.2    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 data set consists of over 17000 randomly sampled alumni who graduated with an undergraduate degree in or before 1996, 10% of the total university alumni base (See Appendix B in Author (2004) for a detailed description of the data set).[10] From this data set, we randomly sampled 2000 alumni for calibration and validation purposes (see Table 2 for descriptive statistics of the calibration and validation sample).

For each alumna/us the data provide aggregate information on total gift giving since the time of graduation, as well as detailed disaggregate data about his/her gift giving since 1976 (or time of graduation, whichever is more recent). The data set also contains disaggregate information about different alumni-university interactions for the years 1976-2001. These interactions include: participation in university events and volunteering for alumni roles.

**Table 2 - Descriptive Statistics of the Calibration and Validation Sample**

---

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

| Key Characteristics | Percentage/Average |
| --- | --- |
| Overall observations (gift opportunities) | 38451 |
| Overall number of alumni | 2000 |
| Mean observations per individual | 19.2 |
| Proportion of donation years | 19.0% |
| Mean donation | $308 |
| Q25 yearly donation | $30 |
| Median yearly donation | $75 |
| Q75 yearly donation | $180 |
| Alumni that never donated | 26% |
| Gender | |
| Female | 38% |
| Male | 62% |
| School | |
| Earth | 2% |
| Humanities | 18% |
| Engineering | 14% |
| Other (e.g., Sciences, Undeclared) | 66% |
| Degree | |
| Undergraduate | 83% |
| Undergraduate + Graduate | 17% |
| Spouse is university alumna/us | |
| Yes | 46% |
| No | 54% |
| Alumni association membership | |
| Member | 56% |
| Not member | 44% |

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 (for the full description of the Run Test results see Author (2004)).

## 5.3    Variables Description

Reflecting upon the HMM, described in §3, the variables of this data set can be divided into four categories:

1. *Individual characteristics* – These variables explain the observed heterogeneity across alumni (the vector $z_i$ in Equation (11)). These variables are known at time 0, which for the specific data set is the most recent of the year 1976 or graduation year.[11] The variables used are,

    $FEMALE_i$ = 1 if alumna, and 0 otherwise,

---

[11] Individual characteristics such as marital status and marriage to an alumna/us from the same institution were excluded from this analysis, because we were unable to verify the status of the alumna/us for these characteristics at time 0.

YSG76$_i$ = number of years since graduation at time 0 (0 if graduated in or after 1976),

EARTH$_i$ = 1 if the alumnus graduated with an Earth science major, and 0 otherwise,[12]

HUMAN$_i$ = 1 if the alumnus graduated with a Humanities major, and 0 otherwise,

ENG$_i$ = 1 if the alumnus graduated with an Engineering major, and 0 otherwise,

UG$_i$ = 1 if the alumnus had only an undergraduate degree from the university, and 0 otherwise,[13]

AA$_i$ = 1 if the alumnus is a member of the Alumni Association, and 0 otherwise.[14]

2. *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 three types of interactions (in addition to past donations),

REUNP$_{it}$ = 1 if the alumnus participated in a reunion in year $t$, and 0 otherwise,

EVENT$_{it}$ = 1 if the alumnus participated in an event (other than a reunion) in year $t$, and 0 otherwise,[15] and

VLNTR$_{it}$ = 1 if the alumnus volunteered for a university role in year $t$, and 0 otherwise.

3. *Influence attempts* - These are short-term effect marketing activities initiated by the university, which determine the vector $\mathbf{x}_{it}$ in Equation (7). In the scanner data context, price promotion, and display are appropriate examples for such short term influence attempts. In the current application, we use a reunion year as such a covariate. We consider reunion year is 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 the actual participation in a reunion, which may lead to a change in the relationship strength between the alumni and the university and to a long-term impact on subsequent donation behavior. Thus, the reunion year is included in the state dependent choice (the vector $\mathbf{x}_{it}$), and the participation in a reunion in the transition matrix (the vector $\mathbf{A}_{it}$). Reunion year is defined as,

---

[12] The degree major variables (EARTH, HUMAN, and ENG) are defined relative to the other majors, which were set to zero.

[13] The undergraduate degree variable is defined relative to having both undergraduate and graduate degrees from the university, which was set to zero.

[14] We assume that most alumni that are members of the Alumni Association joined the Alumni Association at graduation.

[15] This covariate includes university-organized events such as lecture series, alumni dinners, and university sporting events.

REUNY$_{it}$    = 1 if year $t$ is a reunion year for alumnus $i$, and 0 otherwise.[16]

4.  *Choice behavior* - Gift giving behavior - This is the dependent variable, which is captured by the incident of donation (0 or 1).[17]

GIFT$_{it}$    = 1 if the alumnus made any contribution in year $t$, and 0 otherwise.

Consulting with the Alumni Association, the gift giving and events data was aggregated into increments of "calendar years." To account for time dynamics that are not related to the relational marketing encounters, a linear time trend variable, which represents the years since graduation, is included in the state dependent vector,

YSG$_{it}$    = years since graduation of individual $i$ at year $t$.

## 5.4    Estimated Models

We estimated the full HMM of customer relationship dynamics described in §3, using the MCMC hierarchical Bayes estimation described in §3.5. Additionally, we estimated two restricted versions of this model, with no demographics or no heterogeneity, as well as four dynamic and non-dynamic benchmark models commonly used in the literature.

*Model 1* – Full HMM – This is the full HMM

*Model 2* – HMM with heterogeneity but no demographics – This model is similar to Model 1 but without the individual characteristics specification in Equation (11)

*Model 3* – HMM with no heterogeneity and no demographics – This model is similar to Model 1, but with common coefficients across individuals. A limitation of this model is that what may appear to be movement of alumni to different states over time is simply movement from the aggregate population initial state to the alumnus-specific relationship state, given the observed history of donations and interaction with the university for each alumna/us.

*Model 4* – 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 5* – Loyalty model – This model is based on the model proposed by Guadagni and Little (1983) (Hereafter GL). Following GL, the loyalty covariate is defined as,

---

[16] A reunion year is any multiple of five years since the alumnus' graduation.
[17] This variable could be replaced by the actual amount donated using an ordered logit or a tobit model. However, we believe that the actual act of giving is a stronger determinant of relationship than the amount given.

$$LOY_{it} = \lambda * LOY_{it-1} + (1-\lambda) * LAGIFT_{it-1}, \tag{12}$$

where $\lambda$ is a decay parameter.

The probability of choice follows the binary logit formulation:

$$P(Y_{it} = 1) = \frac{\exp(\mathbf{x}'_{\mathbf{it}}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_{\mathbf{it}}\boldsymbol{\beta})}, \tag{13}$$

where,

$\boldsymbol{\beta}$ is a vector of fixed-effect covariates, and

$\mathbf{x}_{\mathbf{it}}$ is a vector of time-varying covariates.

The vector $\mathbf{x}_{\mathbf{it}}$ includes the set covariates described in §5.3 (REUNY$_{it}$, REUNP$_{it}$, VLNTR$_{it}$, EVENT$_{it}$, YSG$_{it}$), and the loyalty covariate (LOY$_{it}$) described in Equation (12).

*Model 6* – Heterogeneous state dependence model – A special case of Model 5 is a model where the decay parameter in Equation (12) is set to zero. This leads to the familiar state dependence model (cf. Seetharaman, Ainslie, and Chintagunta 1999). In this model the dependence of current choice on past choices is captured through the effect of the lagged choice. We incorporate heterogeneity in this model by incorporating observed and unobserved heterogeneity in the intercept and the state dependence parameters following Equation (11).[18]

*Model 7* – Heterogeneous loyalty model – A limitation of the GL model (Model 5) is that the loyalty term captures both time dynamics and cross-individual heterogeneity. Thus, this model is unable to distinguish between the two effects. Keane (1997) demonstrates that introducing heterogeneity into the GL model's parameters improves its predictive ability. Therefore, in addition to Model 5, we estimate a "stronger" GL model, with heterogeneity in the intercept and loyalty parameters.

The distinction between the HMMs and Models 5-7 is noteworthy. Models 5-7 account for time dynamics in an ad hoc fashion where the structure of the effect of past choices on the current choice is defined *a priori* (one lagged choice in the case of the state dependence model and a weighted sum of past choices in the case of the loyalty model). The HMM offers a more behaviorally structured approach for state dependence through a Markovian composition of behavioral relationship states. Furthermore, although some structure is imposed on the transitions between the states, the choice

---

[18] We avoid a full random-effect model, since the variation of some covariates is relatively low, which limits a robust estimation of individual-level posterior distributions, even with the hierarchical Bayes estimation.

of the number of states allows to determine the structure of dynamics based on the complexity of the dynamics of the data at hand.

## 5.5    Estimation Results

Models 3, 4 and 5 were estimated using the maximum likelihood procedure MAXLIK in the GAUSS statistical software.  Models 1, 2, 6, and 7 were estimated using a MCMC hierarchical Bayes procedure, using the Gibbs sampler and the Metropolis Hastings algorithm.

In the hierarchical Bayes estimation, the first 40000 iterations were used as a "burn-in" period and the last 10000 iterations were used to estimate the conditional posterior distributions and moments.  Convergence was assessed by inspecting the time series of draws.[19]

*5.5.1 Selecting the Number of States*

The number of discrete states (NS) could be estimated from the data or defined based on theoretical grounds.  Methodologically, one could treat the number of states as a parameter for estimation.  Thus, the first stage in estimating the HMM is selecting the number of states.  Model selection procedures can then be used to choose the number of states (cf. Kamakura and Russell 1989).  Due to the sensitivity of alternative Bayesian model selection criteria, such as the Bayes factor to the specified priors (Rossi and Allenby 2003), we compare alternative model selection criteria.  Specifically, we contrast the Bayes factor with the log marginal density[20] and the marginal validation log-likelihood measure (Andrews and Currim 2003).  The validation log-likelihood is calculated following Equation (9), using the validation sample (the years 1999-2001 in the data set) at the mean of the individual-level posterior distributions.  The marginal log density and the Bayes factor are calculated from the output of the Metropolis-Hasting sampler.  The marginal log density is calculated using a harmonic mean of the individual likelihoods across iterations (Newton and Raftery 1994).  The marginal density is given by:

$$\tilde{\ell}(\mathbf{Y_{it}} \mid \text{model}) = \left[ 1/K \sum_{k=1}^{K} \left[ \ell(\mathbf{Y_{it}} \mid \{\boldsymbol{\theta_i}\}, \boldsymbol{\psi}) \right]^{-1} \right]^{-1} \quad , \tag{14}$$

where,

$\ell(\mathbf{Y_{it}} \mid \{\boldsymbol{\theta_i}\}, \boldsymbol{\psi})$ is calculated following Equation (9), and

---

[19] See Appendix D in Author (2004) for the posterior means in the last 10000 iterations

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

28

K is the number of MCMC iterations.

Based on the three measures, the optimal model is the model with three states. This model maximizes the log marginal density and the validation LL, and shows a favorable Bayes factor in comparison to the models with two and four states (see Table 3).

**Table 3 - Choosing the Number of States**

| Number of states | Number of parameters | Marginal log density | Validation log-likelihood | Bayes factor[21] |
|---|---|---|---|---|
| 1 | 3 | -18360.8 | -3388.5 | ----- |
| 2 | 31 | -11455.3 | -2030.3 | > 1e300 |
| 3 | 60 | **-11339.4** | **-1951.1** | **2.19e50** |
| 4 | 63[22] | -11360.6 | -1952.8 | 5.46e-10 |

*5.5.2  HMMs' Estimates*

Table 4 reports the posterior means and posterior standard deviations (parameter estimates and standard errors for Model 3) of the three variants of the HMM based on the calibration sample. The identification of the three states is primarily determined by the state-specific intrinsic propensity to donate (the parameters $\beta_{01}$, $\beta_{02}$, and $\beta_{03}$). At the mean of the covariates: "years since graduation" and "reunion year", the conditional probability of donation given state 1 is approximately 0, given state 2 is 32%, and given state 3 is 99%. Consequently, we define these three states as "*dormant*," "*occasional*," and "*active*" states, respectively.

**Table 4 - Estimation Results for the Hidden Markov Models**[23]

| Parameter | Model 1 - HMM with hetero. and demog. | | Model 2 - HMM with hetero. no demog. | | Model 3 - HMM no hetero. no demog. | |
|---|---|---|---|---|---|---|
| $\beta_{01}$ | -6.460 | (0.566) | -6.216 | (0.566) | -6.324 | (0.693) |
| $\beta_{02}$ | 1.743 | (0.257) | 1.657 | (0.212) | 1.646 | (0.128) |

---

[21] The Bayes factor compares the model with *s* states to the model with *s*-1 states. Since the Bayes factor is a ratio-based measure, a value greater than 1 favors the less parsimonious model.

[22] The number of parameters does not increase substantially as the number of states increases beyond three states because of the assumption that the transition parameters are identical for all intermediate states.

[23] Numbers in parentheses are posterior standard deviations for Models 1 and 2, and standard errors for Model 3.

| | | | | | | |
|---|---|---|---|---|---|---|
| $\beta_{03}$ | 1.780 | (0.213) | 1.697 | (0.242) | 1.148 | (0.036) |
| $\beta_1(REUNY)$ | 0.664 | (0.454) | 0.802 | (0.435) | 0.926 | (0.625) |
| $\beta_2(REUNY)$ | 0.227 | (0.275) | 0.214 | (0.292) | 0.226 | (0.082) |
| $\beta_3(REUNY)$ | -0.085 | (0.438) | 0.148 | (0.496) | 0.015 | (0.124) |
| $\beta_1(YSG)$ | -0.099 | (0.184) | 0.066 | (0.160) | 0.085 | (0.037) |
| $\beta_2(YSG)$ | 0.047 | (0.094) | 0.056 | (0.090) | 0.063 | (0.007) |
| $\beta_3(YSG)$ | 0.402 | (0.140) | 0.346 | (0.156) | 0.120 | (0.016) |
| $\mu(hi)_1$ | 3.319 | (1.492) | 3.292 | (1.602) | 3.538 | (0.082) |
| $\mu(lo)_2$ | -1.515 | (0.572) | -1.378 | (0.347) | -2.330 | (0.095) |
| $\mu(hi)_2$ | 1.357 | (0.735) | 1.220 | (0.565) | 1.578 | (0.028) |
| $\mu(lo)_3$ | -1.286 | (1.270) | -1.283 | (1.087) | -1.989 | (0.084) |
| $V(\mu(hi)_1)$ [24] | 1.895 | (0.186) | 2.558 | (0.201) | ------------- | |
| $V(\mu(lo)_2)$ | 0.235 | (0.057) | 0.111 | (0.035) | ------------- | |
| $V(\mu(hi)_2)$ | 0.430 | (0.152) | 0.316 | (0.052) | ------------- | |
| $V(\mu(lo)_3)$ | 1.332 | (0.158) | 1.168 | (0.106) | ------------- | |
| $\rho_{1(EVENT)}$ | 0.469 | (0.553) | 0.153 | (0.510) | 0.040 | (1.025) |
| $\rho_{2(EVENT)}$ | 0.633 | (0.570) | 0.833 | (0.514) | 1.067 | (0.372) |
| $\rho_{3(EVENT)}$ | 0.228 | (0.667) | 0.384 | (0.476) | -0.087 | (0.448) |
| $\rho_{1\,(VLNTR)}$ | 1.107 | (0.373) | 1.150 | (0.601) | 0.333 | (0.383) |
| $\rho_{2\,(VLNTR)}$ | 0.699 | (0.317) | 0.642 | (0.464) | 0.863 | (0.236) |
| $\rho_{3\,(VLNTR)}$ | 0.261 | (0.436) | 0.458 | (0.491) | 0.357 | (0.260) |
| $\rho_{1\,(REUNP)}$ | 2.596 | (0.429) | 2.989 | (0.457) | 2.850 | (0.231) |
| $\rho_{2\,(REUNP)}$ | 1.698 | (0.574) | 0.810 | (0.512) | 1.392 | (0.352) |
| $\rho_{3\,(REUNP)}$ | 0.018 | (0.523) | 0.479 | (0.380) | 1.423 | (1.654) |
| Log marginal density[25] | -11339.4 | | -11474.7 | | -11650.9 | |

One alumni-university interaction that has a strong impact on moving alumni from the occasional to the active state is reunion attendance. The middle matrix in Table 5 demonstrates the effect of reunion attendance on the relationship-states transitions. Reunion participation increases the likelihood that an alumna/us in the dormant state would move to the occasional state from 3% to 33%. Reunion attendance also has a strong impact on moving alumni from the occasional to the active state; it increases this likelihood from 9% to 34%. More importantly, reunion attendance decreases the likelihood of dropping from the occasional state to the "sticky" dormant state from 18% to only 4%. Thus, 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

---

[24] $V()$ refers to the posterior variance across individuals.
[25] For Models 1 and 2, the log marginal density was calculated following Equation (14). For Model 3, the log marginal density is the log-likelihood at the ML parameter estimates.

30

and by increasing the likelihood of a transition to the relatively "sticky" active state. In contrast, the effect of reunion participation on keeping alumni in the active state is minimal. Of course, given that the likelihood that an alumna/us in the active state stays in this state in the next period is 78%, the opportunity for increasing the likelihood of staying in the active state is limited.

**Table 5 – 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 | 97% | 3% | 0% | Dormant | 67% | 33% | 0% | Dormant | 90% | 10% | 0% |
| Occasional | 18% | 73% | 9% | Occasional | 4% | 62% | 34% | Occasional | 10% | 74% | 16% |
| Active | 0% | 22% | 78% | Active | 0% | 21% | 79% | Active | 0% | 18% | 82% |

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 occasional state is consistent with the theory of intermittent reinforcement, which suggests that the effect of intermittent reinforcement lasts longer than the effect of continuous reinforcement. In marketing contexts it has been suggested that purchases and store visits can be encouraged through consumer intermittent reinforcement (Nord and Peter 1980). Due to the probabilistic nature of the transitions, active alumni are likely to transition into the occasional state once in several time periods and are therefore likely to be considered for reinforcement such as reunion attendance reinforcement. In other words, the transition matrix analysis above does not suggest to ignore the active alumni altogether, but to reinforce them on an intermittent basis as they are probabilistically classified into the occasional state.

Attendance in a reunion seems to have the strongest impact on alumni in the occasional state. However, the occasional state and its alumni are not directly observable. The HMM allows the researcher to empirically explore the existence of such a hidden state, dynamically classify alumni into this state, and asses the impact of different alumni-university interactions on moving alumni from and to this state. This type of inference could not be made using the traditional state dependence model that defines only two states based on one lagged choice, or by a static latent class model that does not allow for movements between states due to customer-brand interactions.

Similar transition matrices can be drawn for other interactions between the alumni and the university. The right most matrix in Table 5 describes the effect of volunteering for university roles

on the states' transitions. The effect of volunteering is less strong than the effect of reunion attendance on alumni in the dormant and the occasional states.[26]

One interesting product of the transition matrices in Table 5 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 76%, 17%, and 7% 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 (15), we calculated the state membership probability at time $t$ following a donation at time $t$ and no other alum-university interaction between time $t-1$ and time $t$.

$$P(S_{it} = s \mid Y_{it} = 1, \mathbf{S}_{\mathbf{it\text{-}1}}) = (\mathbf{Pr}(\mathbf{S}_{\mathbf{it\text{-}1}})\mathbf{Q}_{\mathbf{it\text{-}1t.s}}m_{its}) / (\mathbf{Pr}(\mathbf{S}_{\mathbf{it\text{-}1}})\mathbf{Q}_{\mathbf{it\text{-}1t}}\mathbf{m}_{\mathbf{it}} \otimes \mathbf{I}_{\mathbf{NS}}\mathbf{1}'), \tag{15}$$

where,

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

$\mathbf{Q}_{\mathbf{t\text{-}1t.s}}$ is the $s^{th}$ column of the transition matrix $\mathbf{Q}_{\mathbf{t\text{-}1t}}$.

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

**Table 6 – Mean Posterior Transition Following a Donation at Time $t$**

| Donation at t | | | |
|---|---|---|---|
| t-1　＼　t | Dormant | Occasional | Active |
| Dormant | 14% | 86% | 0% |
| Occasional | 0% | 72% | 28% |
| Active | 0% | 8% | 92% |

The transition matrices in Tables 5 and 6 present the mean of the posterior distribution across alumni. Next, we describe the heterogeneity in these matrices across alumni.

*5.5.3 Posterior Distributions and Observed Heterogeneity*

In Models 1 and 2, alumni are allowed to differ with respect to their propensity to switch between states due to the individual-specific transition threshold parameters. Figure 4 depicts the

---

[26] Since the parameters of participation in events (other than reunions) were not significant (see Table 4), we do not elaborate on these alumni-university interactions.

distribution of the alumni's posterior propensity to stay in each one of the states given this state in the previous period (the diagonal elements of the transition matrix with no interactions in Table 5).

**Figure 4 - Posterior Distribution of the Propensity to Stay in Each State**
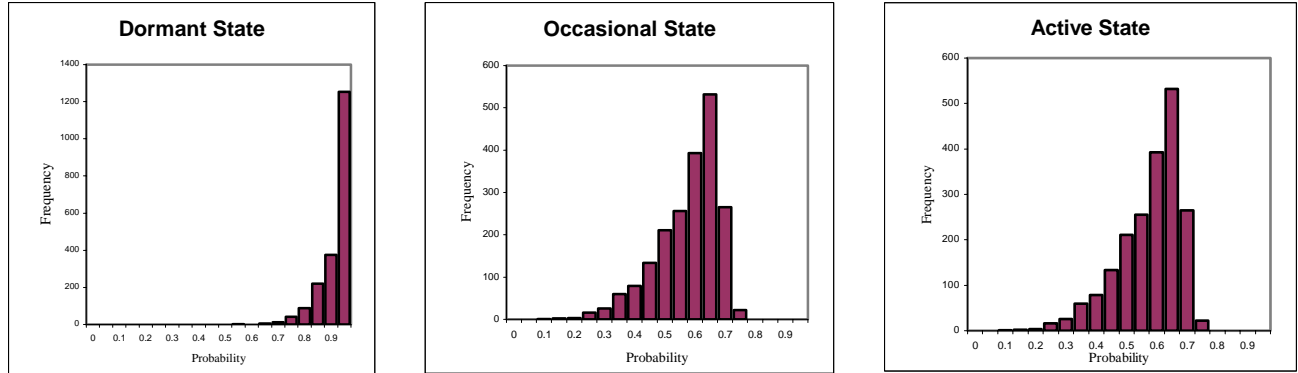


Not only is the dormant state the most "sticky" on average, as suggested by Table 5, but also alumni in this state are most homogenous in terms of their likelihood of staying in this state. On the other hand, alumni in the occasional and active states are relatively heterogeneous in terms of their propensity to stay in these states.

The random-effect parameters are also related to observed heterogeneity. Table 7 presents the parameter estimates of the hierarchical representation of heterogeneity (Equation (11)).

**Table 7 - Mean Posteriors and STDs for Observed Heterogeneity Parameter[27]**

| Variable | $\mu(hi)_1$ High Threshold - Dormant | $\mu(lo)_2$ Low Threshold - Occasional | $\mu(hi)_2$ High Threshold - Occasional | $\mu(lo)_3$ Low Threshold - Active |
|---|---|---|---|---|
| Intercept | 3.755** (0.331) | -1.014** (0.277) | 1.405** (0.370) | -1.382** (0.420) |
| AA | -0.852** (0.096) | -0.223** (0.096) | -0.123 (0.111) | -0.259 (0.171) |
| YSG76 | -0.060** (0.009) | 0.011 (0.012) | -0.022** (0.010) | -0.003 (0.015) |
| FEMALE | -0.081 (0.113) | -0.263** (0.074) | 0.029 (0.100) | 0.189 (0.198) |
| EARTH | -0.409 (0.432) | -0.099 (0.384) | -0.682 (0.437) | 1.324** (0.541) |
| HUMAN | 0.118 (0.326) | -0.347 (0.282) | 0.414* (0.292) | -0.740** (0.391) |
| ENG | -0.165 (0.312) | 0.138 (0.261) | -0.220 (0.318) | -0.163 (0.369) |
| UG | 0.320** (0.150) | -0.067 (0.208) | -0.233 (0.157) | 0.948** (0.247) |

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

Several demographic characteristics have a significant impact on the heterogeneity in the propensity for transition between the states. Specifically, membership in the Alumni Association increases the likelihood of transitioning away from the dormant state. Furthermore, graduates of

---

[27] Numbers in parentheses are posterior standard deviations.

the earth and humanities schools are less likely to remain active in comparison to other alumni. Alumni with dual degrees were significantly more likely to transition away from the dormant state, and significantly less likely to transition away from the active state in comparison to alumni that have only an undergraduate degree.

### 5.6.    Prediction Ability

For each alumna/us in the sample, the last three years of possible gift giving in the data (1999-2001) are saved for validation. This validation data is used to assess the prediction ability of the HMMs and compare it to the four benchmark models. In addition to the 3 HMMs (Models 1-3) we estimated Models 4-7 on the calibration period. Maximum likelihood estimation was used to estimate Models 4 and 5, and a MCMC hierarchical Bayes procedure, using the Gibbs sampler and the Metropolis Hastings algorithm was used to estimate Models 6 and 7.[28] The parameters estimated based on the calibration period[29] are used to predict the 2,000 (alumni) x 3 (years) = 6,000 observations of possible gift giving in the validation period.

We use two measures of prediction ability to compare the alternative models: 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. The RMSPE of each model is compared to the RMSPE of a random choice rule, which is based on the mean aggregate probability of donation over time and alumni.[30] All models significantly improve upon a random choice rule, based on the RMSPE (see Table 8). The RMSPE, and the validation log likelihood suggest that the dynamic HMMs can predict donation periods marginally better than the state dependence and loyalty models. The non-dynamic latent class model with three segments predicts the holdout choices significantly worse than the dynamic models. Overall, the HMMs outperform the alternative models in terms of prediction ability.

### Table 8 - Prediction Ability Measures

| Model | Model 1 HMM with hetero. | Model 2 HMM with hetero. | Model 3 HMM no hetero. | Model 4 Non-dynamic latent | Model 5 Loyalty no hetero. | Model 6 SD with hetero. | Model 7 Loyalty with hetero. |
|---|---|---|---|---|---|---|---|

[28] The parameter estimates of Models 4-7 can be obtained from the authors.

[29] For models 1, 2, 6, and 7 we used the mean of the individual-level posterior distributions.

[30] The random choice rule's RMSPE was calculated based on the aggregate donation probability in the calibration period (19.0%). The random choice rule's RMSPE is 0.393.

| Measure | and demog. | no demog. | | class | | | |
|---|---|---|---|---|---|---|---|
| # of parameters | 60 | 32 | 22 | 11 | 8 | 24 | 25 |
| RMSPE | 0.3158 | 0.3152 | 0.3183 | 0.3641 | 0.3197 | 0.3336 | 0.3213 |
| Improvement in RMSPE over the Random RMSPE | 19.6% | 19.8% | 19.0% | 7.3% | 18.6% | 15.1% | 18.2% |
| Validation -LL | -1951.1 | -1946.3 | -1987.0 | -2464.5 | -2047.2 | -2167.9 | -2032.2 |

The prediction ability of the parsimonious HMM with no demographics (Model 2) is similar to that of the full HMM (Model 1). This is consistent with the relatively small effect of the individual characteristics in Table 7. The similarity in the prediction ability between Models 1 and 3 is also not surprising, since in Model 3 heterogeneity in the validation period is elicited from the history of choices in the calibration period even when the model's parameters do not vary across individuals. By the same token, it is not surprising that the prediction ability of Models 5 and 7 is similar.

The improvement in the prediction ability of Model 1 in comparison to Model 6 is noteworthy. These models are similar in the sense that both models include a structure of state dependence. However, these models differ in the dimensionality of the states' construction. The state dependence in Model 6 is determined solely based on one lagged choice, while the relationship-state dependence is a function of both lagged choices and other alumni-university interactions. The previous section demonstrated the advantage of including alternative customer-brand interactions in the states' transitions to determine the marginal impact of these interactions on the states' transitions. The validation sample analysis suggests that an additional advantage of the relationship HMM is in improving the ability to predict future choices over the state dependence model.

In addition to improved prediction ability, the proposed HMM provides insights into the impact of alternative customer brand interactions on the dynamics in the buying behavior. In the next section, we demonstrate the value of the HMM in identifying the most effective targeting of such customer-brand interactions using a "what-if" analysis.

## 5.7 "What-if" Analysis

The proposed model could be used to answer the following question: What if we employed a selectively targeted marketing program based on a dynamic segmentation of customers into relationship states? In this section, we illustrate such a targeted marketing campaign to persuade

alumni to attend reunions. This type of analysis demonstrates the value of utilizing the proposed HMM as a decision support system for marketing mix strategy.

Assume that the university would like to contact alumni (e.g., via personal phone calls), in an attempt to persuade them to attend the upcoming reunion in the year 1999. This is based on the estimation results which suggest that reunion participation is a significant driver of donation behavior. Since for approximately 20% of the alumni, the year 1999 is a reunion year, we assume that it is too costly to contact all the alumni that are approaching a reunion year. For illustrative purposes, let us assume that the university contacts 10% of these alumni.[31] Which alumni should the university approach to maximize the impact of the marketing campaign? The history of alumni-university interactions and the proposed model could be used to estimate the alumni's relationship state and then use the alumni's state of relationship as a basis for a one-to-one targeting strategy.

The data set, described in §5.2, is limited in the sense that it does not include marketing activities initiated by the Alumni Association. Therefore, for the purpose of this illustrative analysis, we are making several assumptions. First, we avoid the simultaneity problem (the alumna/us simultaneous decision to attend in a reunion and to donate) by assuming that the impact of reunion attendance is similar for alumni that attended the reunion partially due to the marketing activity and alumni that attended the reunion without the marketing campaign. Second, we assume that the marketing campaign increases the likelihood of attending a reunion by 50% from the alumna/us intrinsic propensity to attend a reunion (e.g., an alumnus with a 20% probability of attending a reunion without the campaign has a 30% probability of attending a reunion following a targeted contact during the campaign). This assumption allows for a stronger point share response to the campaign for loyal customers, since we use the state-specific intrinsic propensity of attending a reunion, which is 18%, 31%, and 47% in the dormant, occasional, and active states, respectively. A simple field experiment in which a subset of the alumni is contacted using a non-targeted campaign, could provide estimates that relax these assumptions.

We contrast the impact of alternative marketing campaigns. Previously, we demonstrated that the HMM (which uses latent states) outperforms, in terms of prediction ability, the state dependence

---

[31] In the case of the alumni association that provided this data, this amounts to approximately 3,500 alumni.

model which utilizes observed behavior. In this section, we compare the HMM, latent state-based targeting with targeting that is based solely on the observed behavior.

To target alumni based on their latent relationship states, we first calibrated the model and estimated the alumni relationship state in the year 1998, following Equation (10). Subsequently, we use the estimated state distribution in 1998 to select 10% of the alumni that are approaching a reunion year and have the highest probability of being in each one of the alternative states. We create three such marketing campaigns: *dormant campaign, occasional campaign*, and *active campaign*, which target alumni in the dormant, occasional, and active latent states, respectively. Recall that based on the transition matrices in Table 5, the impact of reunion participation is most significant for alumni in the occasional state (and somewhat in the dormant state). Therefore, we predict that the occasional campaign would maximize the impact of the marketing effort.

A practical alternative for the above HMM targeting is to use directly observed behavior (e.g., past two donation years) to segment the alumni and target marketing activities (cf. Soukup 1983). Targeting that is based on the observed choice behavior is consistent with the RFM approach and is commonly used in target marketing applications (e.g., catalog mailing). Specifically, marketers tend to focus their marketing efforts on customers who purchased the product frequently in the previous periods. We created three observed behavior campaigns which correspond to the three HMM-based campaigns: an *observed non-donors campaign*, which targets alumni that did not donate in the past two years, an *observed switchers campaign*, which targets alumni that donated once in the past two years, and an *observed loyals campaign*, which targets alumni that donated in each of the past two years. In each one of these campaigns, we randomly sampled 10% of the alumni that matched the campaign observed behavior in the years 1997-1998 and are approaching a reunion year in the year 1999.

Additionally, we compare the six targeted campaigns with a naïve *uniform campaign*. In the uniform campaign, we randomly sampled 10% of the alumni that are approaching a reunion year. It is important to note that the number of alumni contacted in all seven campaigns is identical. Due to the proportional responsiveness assumption, the responsiveness to the campaign (in terms of reunion attendance) is higher for the active and observed loyals campaigns relative to the other campaigns. Thus, a superior performance of the occasional campaign could not be attributed to the differences in the reunion attendance rates between the different campaigns.

For each campaign, we calculated the predicted donations in the three years following the campaign (1999-2001), based on the HMM parameters from § 5.5.2 and compared them to the actual donation behavior in the absence of a campaign. For each campaign, we repeated this simulation 10000 times. The average improvements (over the simulation iterations and individuals) in donation rates and donation amounts over no-campaign are presented in Table 9.

**Table 9 - "What-if" Analysis for a Reunion Attendance Marketing Campaign**

| Marketing Campaign | Increase in Donation Rates and Amounts per Alumni over No-Campaign | | | |
|---|---|---|---|---|
| | The year of the campaign | One year after the campaign | Two years after the campaign | Cumulative effect for the three years |
| Dormant | 0.5% $0.84 | 0.4% $0.75 | 0.3% $0.67 | 1.3% $2.26 |
| Occasional | 2.3% $10.97 | 1.8% $8.30 | 1.5% $6.85 | 5.6% $26.12 |
| Active | 0.4% $2.37 | 0.3% $1.46 | 0.2% $1.05 | 0.9% $4.88 |
| Observed non-donors | 0.9% $2.19 | 0.8% $2.29 | 0.7% $2.00 | 2.4% $6.47 |
| Observed switchers | 2.0% $10.21 | 1.6% $7.45 | 1.3% $6.04 | 4.9% $23.71 |
| Observed loyals | 0.7% $4.14 | 0.5% $2.22 | 0.4% $1.82 | 1.6% $8.18 |
| Uniform | 1.1% $3.83 | 0.8% $2.92 | 0.7% $2.54 | 2.5% $9.27 |

In calculating the donation amounts, we use the mean donation amount in the state the alumna/us belongs to in the specific time period. To keep the monetary estimates conservative, we excluded the top 1% of donations (donations over $10,000).[32] The truncated mean donations are $0, $151 and $463 for the dormant, occasional, and active state, respectively. Though we use two independent models for the donation incident and amount estimations, we relax the independence assumption commonly assumed in incident-amount applications (e.g., Schmittlein and Peterson 1994). In this application donation incidents and amounts are correlated, since they are both a function of the alumni underlying state.

As expected, the occasional campaign is the most effective campaign both in terms of generating donation incidents and donation amounts. Over the three years, the occasional campaign created an

---

[32] These top donors receive special care from the university and would not be part of a general alumni marketing campaign.

increase of 5.6% point shares in donation rates and revenue of $26 per individual contacted. Across the 3,500 alumni in the campaign, this amounts to over $90,000 incremental donations due to the campaign over the 3 years. The revenue per contact of this campaign over the three years following the campaign was over twice that of the uniform campaign, over three times that of the loyals campaign, and over five times that of the active campaign. The effectiveness of the occasional campaign was also better than its observed behavior matching campaign, the observed switchers campaign. The increase in donation incidents and amounts in the occasional campaign are 14% and 10% higher than those of the observed switchers campaign, respectively. The average overlap between the target population in the occasional and observed switchers campaign is 80%.

It is only through the use of the HMM and the analysis of the transition matrices (Table 5) that one could have identified the existence of an occasional or switchers state of behavior and the strong effect of reunion attendance on alumni in this state. Hence, using the latent states rather than observed behavior may be of substantial financial value to the university. Specifically, targeting the observed loyal alumni (customers), an approach that has been suggested in the marketing literature (Blattberg and Deighton 1996) and used in practice (e.g., airline loyalty programs), performs poorly in this analysis both in terms of donation rates and donation amounts (the overall revenue of $8 per contact in the observed loyals campaign is not likely to justify the campaign's cost).

Though diminishing over time, the cumulative effect of the reunion campaign two years following the campaign was still large. Moreover, the occasional campaign not only had a superior immediate impact on donations, but also had a more enduring effect in comparison to the observed behavior and uniform campaigns. This suggests that customer-brand interactions such as reunions may have an enduring impact on customers' medium and long-term behavior, that the Markovian nature of the model allows the estimation of such long-lasting effects, and that firms could take advantage of this enduring impact by selective targeting.

A noteworthy limitation of the "what-if" analysis described above is that the proposed model is not a structural forward-looking model, in which the parameter estimates are invariant to policy changes (Erdem and Keane 1996). For these reasons and others, this "what-if" analysis is suggestive of the marginal value of dynamic targeting of alternative marketing activities rather than a formal policy tool. Nevertheless, this "what-if" analysis demonstrates that the estimation of the HMM not

only enables dynamic classification of customers into relationship states, but also a dynamic assessment of the effectiveness of alternative marketing actions. Such a utilization of the HMM could improve the ability of marketers to send effective and timely marketing messages.

From a theoretical point of view, it would be instructive to understand the behavioral dimensions underlying the relationship states. Thus, in the next section, we use survey-based data to contrast the three alumni states on attitudinal dimensions, in order to validate the behavioral basis underlying the HMM's relationship states.

## 5.8    Behavioral and Attitudinal Dimensions

In order to broaden the applicability of the HMM to most common CRM data sets, 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 differences in the attitudinal relationship behavior. 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; Dick and Basu 1994; Fournier 1998; Morgan and Hunt 1994). To explore the underlying attitudinal dimensions of the three relationship states, we use survey-based data.

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 the sub-sample of alumni for which we had the full gift giving behavior data. For this subset of alumni, we calibrated the HMM (Model 1) and estimated, using Equation (10), the alumni probability of membership in each of the three relationship states in the year of the survey.[33] The highest probability rule was used to classify the alumni into the relationship states.

---

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

Table 10 compares the mean responses to the relationship questions across alumni in the three relationship states. On all relationship measures, except affinity to the graduating class, the average ratings towards the university are increasing in the states from dormant to occasional to active. The last column in Table 10 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 10 - Behavioral Dimensions of Relationship in the Three States[34]

| Question | Scale | Dormant | Occasional | Active | ANCOVA P-values |
|---|---|---|---|---|---|
| How *satisfied* are you with your university experience? | 1-5 | 4.51 | 4.75 | 4.80 | 0.220 |
| How strong is your *feeling* about the university? | 1-5 | 4.31 | 4.50 | 4.60 | 0.098 |
| Do you feel *proud* of your university degree? | 1-4 | 3.47 | 3.62 | 3.69 | 0.833 |
| Do you feel your university experience helped *shape your life*? | 1-4 | 2.89 | 3.24 | 3.43 | 0.069 |
| Do you feel a strong *emotional connection* to the university? | 1-4 | **2.72** | **3.14** | **3.22** | **0.022** |
| Do you feel a strong *responsibility* to help the university? | 1-4 | **2.28** | **2.66** | **3.03** | **0.015** |
| Do you feel strong *affinity* with your graduating class? | 1-4 | **1.94** | **2.52** | **2.26** | **0.009** |
| How strongly would you *recommend* 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 *serving* your needs as an alum? | 1-4 | 2.74 | 2.93 | 2.96 | 0.223 |
| To what extent do you feel the university *values* its alumni? | 1-3 | **2.25** | **2.41** | **2.55** | **0.047** |
| Do your *parents/grandparents* have a degree from this university? | Yes/No | 19% | 18% | 12% | 0.109 |
| Did you receive *financial aid* 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 |

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

---

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

41

dimensions of commitment, responsibility, and self-connection (Fournier 1998; Morgan and Hunt 1994), are important factors of customer-brand relational behavior.

These attitudinal measures give behavioral support to the projection from higher propensity to donate to a higher state 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 attitude towards donation, 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.

## 6. GENERAL DISCUSSION

In this paper, we use the alumni-university gift giving data to estimate a hidden Markov model of customer relationship dynamics. The HMM enables the estimation of the alternative drivers of these dynamics. We account for observed and unobserved heterogeneity using a hierarchical Bayes MCMC estimation procedure.

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 interactions between the customer and the brand on the customer's relationship state and consequent buying behavior.

Using simulated data, we show that the HMM is able recover the customer's true nature of dynamic behavior and accurately identify the customer's relationship state. The empirical application to the problem of university-alumni relationships demonstrates the use of the model to a dynamic relationship problem. Examining the transition probabilities we find that the impact of reunion attendance is the strongest for alumni in the transitory, occasional state. Accordingly, we demonstrate through a "what-if" analysis that targeting the alumni in the occasional state is financially superior to several competing targeting strategies that are based on observed and latent

behaviors. Specifically, we find that targeting alumni in the latent occasional state is superior to the common practice of targeting the most loyal or highest lifetime value alumni (or customers). 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. Furthermore, the HMM can be used to understand which marketing activities are most effective in building customer-brand relationships and driving actionable behavior, and to determine the optimal targeting of such marketing activities. Additionally, the empirical application suggests that the proposed model predicts future choices better than several benchmark models.

The proposed model extends the marketing literature by suggesting a Markovian framework for estimating the dynamics in customer relationships. The HMM enables the use of observed buyer behavior to estimate the long-term impact of customer-brand interactions on customer relationships, an issue that has been largely neglected in the customer relationship literature. Methodologically, the proposed model extends the specification of state dependence to a more behaviorally structured relationship-state dependence, which incorporates a dynamic response to marketing actions 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 the dynamic structure in the loyalty and state dependence models.

The HMM also extends the family of latent class models (Kamakura and Russell 1989) by allowing for dynamic latent segmentation. To our knowledge this is the first model that investigates the impact of covariates, such as customer-brand 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.

This research is not without limitations. The main purpose of the empirical application is to demonstrate the application of the proposed customer relationships dynamics model. Thus, in order to keep the modeling endeavor generalizable, we avoided modifying the general model of

customer relationships to the specific empirical application. However, the empirical application in the context of alumni-university relationships raises several issues that are *specific* to this application, such as modeling donation amounts, and the existence of multiple simultaneous actions that may result from the relationship state (e.g., donation, volunteering, and attendance in events). A worthwhile application of the proposed model would be to estimate a comprehensive model that is tailored to modeling alumni gift giving behavior. Such a study could explore a modified HMM in which the state-dependent choice is replaced with a component that captures the monetary donation amounts as well as a multivariate choice component which captures multiple actions taken by the alumni, such as the decision to participate in university events. More generally, future research could explore the application of multivariate HMMs to modeling multiple modes of actions that stem from a single underlying state of behavior.

To keep the model parsimonious, we made several simplifying assumptions with respect to the model's parameters. Using a richer data set, one may be able to relax the random walk assumption and allow for a full transition matrix between each and every relationship state. Additionally, a richer data set may allow estimating state-specific parameters for the intermediate states (the states 2 to NS-1) and random-effect parameters for all the parameters of the transition matrix.

To broaden the applicability of the proposed model to most common CRM data sets, we used only observed buying behavior to elicit the relationship states. Using survey data, we show that the relationship states, which were estimated based on observed buying behavior, are also different in terms of behavioral dimensions of 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 observed behavioral variables would be helpful in providing insight into the mediating factors of the projection from observed buying behavior to 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 (Roy, Chintagunta, and Haldar 1996).

The HMM framework could also be extended to the study of alternative measures of customer relationships. For example, using the HMM framework, one could explore the connection between service encounters, satisfaction, and actionable behavior (Bolton 1998). The HMM could also be applied to purchase frequency analysis, where the observed behavior is time until next purchase rather than probability of purchase in the next period (e.g., Allenby, Leone, and Jen 1999).

To summarize, we believe that from the modeling perspective, marketing researchers have not armed CRM practitioners with adequate metrics and models for evaluating customer relationships, their evolution over time, and the effect of marketing activities on altering this evolution. These factors are necessary to transform a CRM system from an information system into a decision support system. In this research, we take a step towards filling this gap.

## APPENDIX A – ESTIMATION ALGORITHM

The MCMC procedure recursively generates draws from the conditional distribution of the model's parameters. The parameters could be divided into two groups: (1) parameters that vary across individuals (random-effect parameters) and, (2) parameters that do not vary across individuals (fixed parameters). The set of parameters in each group is determined by the model's heterogeneity specification as described in §3.6.

We denote by $\theta_i$ the set of random-effect parameters and by $\psi$ the set of fixed parameters. For the HMM described in Section 3, the random-effect vector of parameters includes $\theta_i = \{\mu(hi)_{1i}, \mu(lo)_{si}, \mu(hi)_{si}, \mu(lo)_{NSi}\}$. The dimension of $\theta_i$ is 4. The vector $\psi$ includes, $\psi = \{\rho_1, \rho_s, \rho_{NS}, \beta_{01}, \beta_{02}, ..., \beta_{0S}, \beta_1, \beta_2, ..., \beta_{NS}\}$. The dimension of the vector $\psi$ is $3 \times Na + NS \times (Nx+1)$, where $NS \geq 3$ is the number of states, $Na$ is the dimension of the vector ($A_{it}$) and $Nx$ is the dimension of the vector ($x_{it}$).[35] Note that all the parameters in $\theta_i$ and $\psi$ are defined from $-\infty$ to $\infty$. Therefore, the Normal prior is appropriate for these two vectors. Individual characteristics are introduced hierarchically to the vector $\theta_i$ following Equation (11).

### The Joint PDF

First, let us define the joint probability density function for the observed data and the model's parameters.[36]

$$(A.1) \quad L(\mathbf{H}, \{\theta_i\}, \psi) = \prod_{i=1}^{N} [\mathbf{H}_{T_i} | \{\theta_i\}, \psi] \prod_{i=1}^{N} [\{\theta_i\} | \delta, \Sigma_\theta][\delta][\Sigma_\theta][\psi]$$

The first term in Equation (A.1) above is the likelihood of observed sequences of choices given the models' parameters as defined in Equation (9). The second term represents the distribution of the model parameters across individuals, and the last term represents the prior distributions.

### Generate $\theta_i$

The full conditional distribution of $\theta_i$ is,

$$(A.2) \quad [\{\theta_i\} | \mathbf{H}, \psi, \mathbf{z}, \delta] \propto L(\mathbf{H}) N(\{\theta_i\} | \psi, \mathbf{z}, \delta) \propto |\Sigma_\theta|^{-1/2} \exp[-1/2(\theta_i - \delta'\mathbf{z})'\Sigma_\theta^{-1}(\theta_i - \delta'\mathbf{z})] L(\mathbf{H}),$$

where $L(\mathbf{H})$ is the likelihood function from Equation (9).

Since (A.2) does not have a closed form, the Metropolis-Hastings algorithm is used to draw from the conditional distribution of $\theta_i$. The M-H algorithm proceeds as follows; let's define $\theta_i^{(k)}$ as the draw of $\theta_i$ 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 $N(0, 0.15I)$. The step size was chosen to yield an acceptance rate of 30%-70%.
The probability of accepting this draw is,

$$(A.3) \quad \text{Pr(acceptance)} = \min\left\{ \frac{[\exp\left(-1/2(\theta_i^{(k+1)} - \delta'\mathbf{z}_i)'\Sigma_\theta^{-1}(\theta_i^{(k+1)} - \delta'\mathbf{z}_i)\right)] L(\mathbf{Y} | \theta_i^{(k+1)})}{[\exp\left(-1/2(\theta_i^{(k)} - \delta'\mathbf{z}_i)'\Sigma_\theta^{-1}(\theta_i^{(k)} - \delta'\mathbf{z}_i)\right)] L(\mathbf{Y} | \theta_i^{(k)})}, 1 \right\},$$

---

[35] For NS=2 there are only $2 \times Na + 2 \times (Nx+1)$ parameters.
[36] The square brackets [ ] correspond to a probability density function.

**Generate δ**

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

$[\mathbf{v\delta} \,|\, \{\mathbf{\theta_i}\}, \mathbf{z}, \mathbf{\Sigma_\theta}] = MVN(\mathbf{u_n}, \mathbf{V_n})$,

where,

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

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

$Z = (\mathbf{z_1'}, \mathbf{z_2'}, ..., \mathbf{z_N'})$ is an *Nxnz* matrix of covariates,

$\mathbf{\Theta} = (\mathbf{\theta_1'}, \mathbf{\theta_2'}, ..., \mathbf{\theta_N'})$ is an *Nxnθ* matrix which stacks $\{\mathbf{\theta_i}\}$,

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

$\mathbf{V_0}$ and $\mathbf{u_o}$ are prior hyperparameters,

$n\theta = \dim(\mathbf{\theta_i})$, and

$nz = \dim(\mathbf{z_i})$,

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

**Generate $\mathbf{\Sigma_\theta}$**

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

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

**Generate ψ**

Similar to (A.2) the conditional distribution of $\mathbf{\psi}$ can be defined by,

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

where $\psi_0$ and $V_{\psi_0}$ are diffused priors and L(Y) is the likelihood function from Equation (9).

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

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

We define diffuse priors for the conditional distribution of $\mathbf{\psi}$ by setting $\psi_0$ to a $n\psi \times 1$ vector of zeros, and $\mathbf{V_{\psi0}} = 30\mathbf{I_{n\psi}}$, where $n\psi = \dim(\mathbf{\psi})$.

**REFERENCES**

Author (2004), "A Hidden Markov Model of Customer Relationship Dynamics," Ph.D. dissertation.

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.

———, 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.

Ansari, Asim and Carl F. Mela (2003), "E-customization," *Journal of Marketing Research*, 40, 2 (May), 131-145.

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.

Baade, Robert A. and Jeffrey O. Sundberg (1996), "What Determines Alumni Generosity?" *Economics of Education Review,* 15 (1), 75-81.

Bass, Frank M., Moshe M. Givon, Manohar U. Kalwani, David Reibstein and Gordon P. Wright (1984), "An Investigation into the Order of the Brand Choice Process," *Marketing Science*, 3 (4), 267-287.

Basu, Amiya K., Atasi Basu and Rajeev Batra (1995), "Modeling The Response Pattern to Direct Marketing Campaigns," *Journal of Marketing Research*, 32 (May), 204-212.

Bawa, Kapil and Robert W. Shoemaker (1987), "The Effect of Direct Mail Coupon on Brand Choice Behavior," *Journal of Marketing Research*, 24 (November), 370-376.

Beirne, Mike (2002), "Dollars in the Desert," April 1st, *Brandweek*.

Blattberg, Robert C. and John Deighton (1996), "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, July-August, 136-144.

Böckenholt, Ulf and William R. Dillon (2000), "Inferring Latent Brand Dependencies," *Journal of Marketing Research*, 37 (February), 72-87.

——— 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.

———, P.K. Kannan and Matthew D. Bramlett (2000), "Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value," *Journal of the Academy of Marketing Science*, 28 (1), 95-108.

——— and Katherine N. Lemon (1999), "A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction," *Journal of Marketing Research*, 36 (May), 171-186.

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 Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378–394.

Chen, Yuxin and Ganesh Iyer (2002), "Consumer Addressibility and Customized Pricing," *Marketing Science*, 21 (2), 197-208.

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing,* 13 (3), 2-12.

Dekimpe, Marnik G., Dominique M. Hanssens and Jorge M. Silva-Risso (1999), "Long-Run Effects of Price Promotions in Scanner Markets," *Journal of Econometrics*, 89 (1/2), 269-291.

Dick, Alan S. and Kunal Basu (1994), "Customer Loyalty: Toward an Integrated Conceptual Framework," *Journal of the Academy of Marketing Science*, 22 (2), 99-113.

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.

Dwyer, Robert F., Paul H. Schurr, and Sejo Oh (1987), "Developing Buyer-Seller Relationships," *Journal of Marketing*, 51 (April), 11-27.

Erdem, Tülin and Michael P. Keane, (1996), "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1-20.

——— and Baohong Sun (2001) "Testing for Choice Dynamics in Panel Data," *Journal of Business and Economic Statistics*, 19 (2), 142-152.

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.

Feinberg, Fred M., Aradhna Krishna and Z. John Zhang (2002), "Do We Care What Others Get? A Behaviorist Approach to Targeted Promotions," *Journal of Marketing Research*, 39 (August), 277-291.

Flanagan, John C. (1954), "The Critical Incident Technique," *Psychological Bulletin***,** 51 (4), 327-359.

Ford, David (1980), "The Development of Buyer-Seller Relationships in Industrial Markets," *European Journal of Marketing*, 14 (5/6), 339-353.

Fournier, Susan (1998), "Consumers and Their Brands: Developing Relationship Theory in Consumer Research," *Journal of Consumer Research* (March), 343-373.

——— and Julie L. Yao (1997), "Reviving Brand Loyalty: A Reconceptualization within the Framework of Consumer-Brand Relationships," *International Journal of Research in Marketing,* 14 (December), 451-472.

Frank, Roland E. (1962), "Brand Choice as a Probability Process," *Journal of Business,* 35 (January), 43-56.

Gönül, Füsun and Frenkel Ter Hofstede (2003) "A New Empirical Approach to Catalog Mailing Decisions," *Working paper*, McCombs School of Business, University of Texas at Austin.

———, Byung-Do Kim, and Meng Ze Shi (2000) "Mailing Smarter to Catalog Customers," *Journal of Interactive Marketing*, 14 (2), 2-16.

——— and Meng Ze Shi (1998), "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science*, 44 (9), 1249-1262.

Greene, William H. (1997), *Econometric Analysis*, Prentice Hall, Third edition.

Guadagni, Peter and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science,* 2 (3), 203-238.

Gupta, Sunil, Donald R. Lehmann and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (February), 7-18.

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.

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.

Jupiter Media Metrix (2002), "Corporate Spending on Customer Service Technology Will Outpace All Other IT Initiatives Over The Next Two Years, Reports Jupiter Media Metrix," Press release, February 26, New York.

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.

Li, Shibo, John C. Liechty and Alan L. Montgomery (2002), "Modeling Category Viewership of Web Users with Multivariate Count Models," *Working Paper,* Carnegie Mellon University, GSIA.

MacDonald, Iain L. and Walter Zucchini (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*, London: Chapman and Hall.

McAlister, Leigh and Edgar Pessemier (1982), "Variety Seeking Behavior: An Interdisciplinary Review," *Journal of Consumer Research*, 9 (December), 311-322.

———, 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.

Meyer, Robert, Tülin Erdem, Fred Feinberg, Itzhak Gilboa, Wesley Hutchinson, Arahdna Krishna, Steven Lippman, Carl Mela, Amit Pazgal, Drazen Prelec, and Joel Steckel (1997), "Dynamic Influences on Individual Choice Behavior," *Marketing Letters*, 8 (3), 349-360.

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.

Morgan, Robert M. and Shelby D. Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing*, 58 (3), 20-38.

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.

Nord, Walter R. and Paul J. Peter (1980), "A Behavior Modification Perspective on Marketing," *Journal of Marketing*, 44 (Spring), 36-47.

Okunade, Albert A. (1993), "Logistic Regression and Probability of Business School Alumni Donations: Micro-data Evidence," *Education Economics,* 1, 243-258.

——— and Robert L. Berl (1997), "Determinants of Charitable Giving of Business School Alumni," *Research in Higher Education,* 38 (April), 201-214.

Oliva, Terence A., Richard L. Oliver and Ian C. MacMillan (1992), "A Catastrophe Model for Developing Service Satisfaction Strategies," *Journal of Marketing*, 56 (July), 83-95.

Oliver, Richard L. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*, New York: McGraw-Hill.

——— (1999), "Whence Customer Loyalty," *Journal of Marketing*, 63 (October), 33-44.

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.

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

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science,* 22 (3), 304-328.

——— , Robert E. McCulloch and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321–340.

Roy, Rishin, Pradeep K. Chintagunta and Sudeep Haldar (1996), "A Framework for Investigating Habits, 'The Hand of the Past,' and Heterogeneity in Dynamic Brand Choice," *Marketing Science*, 15 (3), 280-299.

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.

———, Valarie A. Zeithaml and Katherine N. Lemon (2000)*, Driving Customer Equity*, The Free Press.

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.

Seetharaman, P.B., Andrew Ainslie and Pradeep K. Chintagunta (1999), "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research*, 36 (November), 488-500.

Sharp, Byron and Anne Sharp (1997), "Loyalty Programs and their Impact on Repeat-Purchase Loyalty Patterns," *International Journal of Research in Marketing,* 14 (5), 473-486.

Soukup, J. David (1983), "A Markov Analysis of Fund-Raising Alternatives," *Journal of Marketing Research*, 20 (August), 314-319.

Srinivasan V. and R. Kesavan (1976), "An Alternative Interpretation of the Linear Learning Model of Brand Choice," *Journal of Consumer Research*, 3, 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 (2001), "The Importance of Linking Customer Acquisition to Customer Retention," *Journal of Marketing Research,* 39 (May), 262-268.

Verhoef, Peter C. (2002), "The Joint Effect of Relationship Perceptions, Loyalty Program and Direct Mailings on Customer Share Development," *Working Paper*, Erasmus Research Institute in Management.

Vo, Kim (2002), "Stanford University Nears $1 Billion Mark in Fundraising Campaign," May 28, *San Jose Mercury News.*

Wedel, Mihcel 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.

Wulf, Kristof, Gaby Odekerken-Schröder, and Dawn Iacobucci (2001), "How and Why Retailers Should Invest in Consumer Relationships: A Cross-Country and Cross-Industry Exploration," *Journal of Marketing*, 65 (4), 33-50.