MARKETING SCIENCE
Vol. 32, No. 4, July–August 2013, pp. 570–590
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2013.0786

A Joint Model of Usage and Churn in Contractual Settings

Eva Ascarza
Columbia Business School, New York, New York 10027, ascarza@columbia.edu
Bruce G. S. Hardie
London Business School, London NW1 4SA, United Kingdom, bhardie@london.edu

As firms become more customer-centric, concepts such as customer equity come to the fore. Any serious attempt to quantify customer equity requires modeling techniques that can provide accurate multiperiod forecasts of customer behavior. Although a number of researchers have explored the problem of modeling customer churn in contractual settings, there is surprisingly limited research on the modeling of usage while under contract. The present work contributes to the existing literature by developing an integrated model of usage and retention in contractual settings. The proposed method fully leverages the interdependencies between these two behaviors even when they occur on different time scales (or "clocks"), as is typically the case in most contractual/subscription-based business settings.

We propose a model in which usage and renewal are modeled simultaneously by assuming that both behaviors reflect a common latent variable that evolves over time. We capture the dynamics in the latent variable using a hidden Markov model with a heterogeneous transition matrix and allow for unobserved heterogeneity in the associated usage process to capture time-invariant differences across customers.

The model is validated using data from an organization in which an annual membership is required to gain the right to buy its products and services. We show that the proposed model outperforms a set of benchmark models on several important dimensions. Furthermore, the model provides several insights that can be useful for managers. For example, we show how our model can be used to dynamically segment the customer base and identify the most common "paths to death" (i.e., stages that customers go through before churn).

Key words: churn; retention; contractual settings; access services; hidden Markov models; RFM; latent variable models
History: Received: April 10, 2012; accepted: February 9, 2013; Preyas Desai served as the editor-in-chief and Scott Neslin served as associate editor for this article. Published online in *Articles in Advance* May 22, 2013.

1. Introduction

Faced with ever-increasing competition and more demanding customers, many companies are recognizing the need to become more customer-centric in the way they do business. At the heart of customer centricity is the concept of customer equity (Blattberg et al. 2001, Rust et al. 2001, Fader 2012). Any serious attempt to quantify customer equity requires modeling techniques that can provide accurate multiperiod forecasts of customer behavior. Numerous researchers working in the areas of marketing, applied statistics, and data mining have developed a number of models that attempt to either explain or predict customer churn at the next contract renewal occasion. However, customer retention is not the only dimension of interest in the customer relationship; there are other behaviors that influence the value of a customer as well. This is the case in contractual business settings or so-called access services (Essegaier et al. 2002) where we observe customer usage while "under contract."[1] In such businesses, predictions of future usage are an important input into any analysis of customer value. In contrast to the work on retention, researchers have been surprisingly silent on how to forecast customers' usage in contractual settings (Blattberg et al. 2008).

One characteristic of contractual businesses is that usage and retention are, by definition, interconnected processes: customers need to renew their contracts/memberships/subscriptions in order to have continued access to the associated service. Furthermore, given that usage and renewal are decisions made by the same customer, there might be factors (e.g., satisfaction, commitment to the organization, service quality) that simultaneously influence both behaviors. For example, let us consider a gym that offers a monthly membership providing customers with unlimited use

[1] The term "usage" refers to whatever quantity is relevant to the business setting being studied, be it the number of transactions, purchase volume, total expenditure, etc.

of its facilities. And let us consider Susan, who currently has certain fitness goals and is committed to exercise; as such, she is a member of her local gym. Susan's commitment will be reflected in the number of times she exercises in a particular week. If she feels very committed, she is likely to exercise very often. However, if her commitment decreases over time, her propensity to use the gym facilities will also decrease. Similarly, Susan's decision of whether or not to renew her membership at the end of each month is likely to be influenced by her level of commitment at that particular point in time. Thus, although there might be factors affecting usage or renewal decisions exclusively (e.g., work travel means fewer visits to the gym in a given week), it is easy to think of a common factor (e.g., commitment) affecting both decisions. As a consequence, if we want to understand and predict customer usage and renewal behaviors, we should model them jointly.

Another common characteristic of contractual settings is that usage and renewal behaviors typically do not occur on the same "clock" (or time scale). In the gym example above, membership is monthly, whereas usage is typically summarized on a daily or weekly basis. Similarly, cable/satellite TV subscriptions may run on a monthly basis, whereas we observe the number of movies purchased each day or week. Given the existing modeling tools at her disposal, the analyst will either need to model each behavior independently—hence not capturing the interdependencies between these two parallel processes—or aggregate the usage observations and set both processes to the same clock. The problem with such an aggregation exercise is that data that could enrich the analyst's understanding of customer relationship dynamics—and, consequently, the predictions of future behavior—are wasted. For example, suppose Susan's commitment to exercise drops a few weeks before her monthly contract comes up for renewal; we would expect to see this reflected in a reduction in her usage of the gym's services. If intramonth usage is ignored—or aggregated up to the monthly level—the analyst will not be able to detect any drop in Susan's commitment, and hence the fact that she is at risk of churning until it is too late and she has cancelled her membership. (On the flip side, an increase in usage could signal an increase in commitment with the associated opportunities for customer development.) By acknowledging the two-clock nature of most contractual settings in which usage while under contract is observable, one can leverage the dynamics in usage to better understand and predict customer behavior (including renewal).

The present work contributes to the existing literature by building an integrated model of usage and retention for subscription-based settings. We propose a dynamic latent variable model that captures the interdependencies between these two behaviors, even when they occur on different clocks. The proposed model fully leverages the two-clock nature of most contractual settings by allowing usage and retention to occur on two different time scales. This approach makes it possible to generate accurate multiperiod forecasts of usage and renewal behavior and provides multiple insights into customer behavior that are managerially useful. For example, this model can be used to dynamically segment the customer base on the basis of underlying commitment (as opposed to usage alone), which has implications for customer development. It can also be used to identify the most common "paths to death" (i.e., stages that customers go through before cancellation). The data requirements of the model are minimal; the data needed to implement the model are readily found in a firm's customer database.

In the next section, we review the relevant literature and introduce our general modeling approach. Section 3 formalizes the assumptions and specifies our joint model of usage and renewal behavior. We present the empirical analysis in §4 and then discuss the managerial insights of the proposed model in §5. We conclude with a summary of the methodological and practical contributions of this research, as well as a discussion of directions for future research in §6.

2. Proposed Modeling Approach

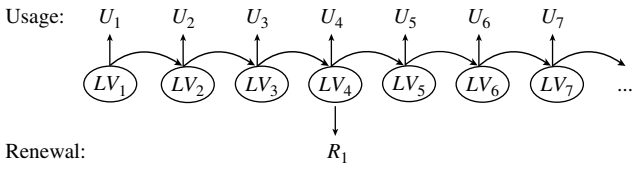
Researchers working in the areas of marketing, applied statistics, and data mining have developed a number of models that attempt to either explain or predict churn (e.g., Bhattacharya 1998, Mozer et al. 2000, Parr Rud 2001, Lemon et al. 2002, Lu 2002, Larivière and Van den Poel 2005, Lemmens and Croux 2006, Schweidel et al. 2008b, Risselada et al. 2010). This stream of work has typically modeled churn at the next renewal opportunity as a function of, among other things, past usage behavior. (See Blattberg et al. 2008 for a review of these various methods.) Although these methods typically provide good predictions of churn at the next renewal opportunity, they are generally not up to the task of making multiperiod forecasts of renewal behavior. The task of predicting future usage in contractual setting has received less attention (Blattberg et al. 2008). The basic approach has been to model current usage as a function of past usage (e.g., Bolton and Lemon 1999).

Consider data of the form presented in Figure 1, where we observe usage (e.g., number of transactions, purchase volume, total expenditure) every period ($t = 1, 2, 3, \ldots$) and renewal opportunities every four usage periods. Having calibrated a model in which churn is a function of past usage behavior (e.g., $R_1 = f(U_1, U_2, U_3)$), we can forecast churn at

Figure 1 Basic Data Structure

Usage:	U_1	U_2	U_3	U_4	U_5	U_6	U_7
Renewal:				R_1			

Figure 2 Assumed Data-Generating Process



$t = 8$ as we have usage data up to and including $t = 7$ (e.g., $\hat{R}_2 = f(U_5, U_6, U_7)$). However, we cannot forecast renewal at $t = 12$ because that would require usage data for periods 8–11, which are currently the future.[2]

One possible solution is to create a usage submodel in which current usage is modeled as a function of past usage. The churn and usage models can be cobbled together to provide multiperiod forecasts of both behaviors (i.e., use the usage model to predict future usage, which then serves as an input into the churn model to predict future churn).[3] Examples of this approach in the marketing literature include Borle et al. (2008), who combine an interpurchase time model for usage with a conditional hazard for retention, and Bonfrer et al. (2010), who combine a geometric Brownian motion process for usage with a first-passage time model for defection. Both models provide multiperiod forecasts of usage and churn, but they assume that churn can occur each usage period (i.e., usage and renewal decisions occur on the same clock), thus limiting the settings in which they can be applied.

Our proposed approach to the modeling problem is illustrated in Figure 2. We assume the existence of a common latent variable, the level of which influences a customer's usage and her likelihood of renewing her contract at each renewal opportunity. As such, changes in usage over time, along with the decision to churn, reflect dynamics in the latent variable. If we can forecast the evolution of the latent variable multiple periods into the future, forecasts of usage and renewal follow automatically. (In others words, there is no need to use predictions of usage as inputs when predicting usage and retention.)

The idea of usage and renewal being driven by a common latent variable is supported by survey-based research in the marketing literature, which shows that contract renewal and service usage are driven by some underlying attitudinal construct. For example, Rust and Zahorik (1993) propose a dynamic framework linking customer satisfaction to retention. They find that changes in retention rates are linked to changes in satisfaction with the service. Bolton (1998) shows that differences in satisfaction levels explain a substantial portion of the variance in contract durations. Similarly, Bolton and Lemon (1999) find a significant relationship between satisfaction and usage. Other researchers have examined alternative attitudinal constructs such as commitment (e.g., Gruen et al. 2000, Verhoef 2003). In this research, we assume the existence of such a latent variable. We seek to model it and its effects on the manifest variables; however, we do not formally define the variable, a position consistent with most latent variable models in the statistics, biostatistics, econometric, and psychometric literatures.[4] For the sake of linguistic simplicity—and given the nature of our empirical setting—we will call it "commitment" going forward.[5]

There are several advantages associated with such a modeling approach. First, it easily accommodates and leverages the fact that usage and renewal typically occur on two different time scales, without the need to ignore (or aggregate) useful information. Second, although the joint estimation of models in which retention and usage are modeled as a function of past usage allows for the possibility of exogenous factors/shocks affecting both decisions simultaneously (e.g., through a correlated error structure), such an approach does not explicitly model systematic dynamics in those factors. In other words, combining the two submodels "controls" for common factors affecting both decisions, but it does not "explain" them. (Explicitly modeling the evolution of a latent variable affecting both decisions not only improves our insights into customer behavior but also implies that we can forecast those systematic changes, resulting in more accurate predictions.[6]) Last, but by no means least, our modeling approach provides a

[2] This problem exists regardless of how many usage periods occur between successive renewal opportunities.

[3] Some of the benchmark models considered in our empirical analysis are based on this logic. An obvious problem with such an approach is that any prediction error propagates through the forecasts.

[4] At an abstract level, the notion of a latent variable driving both behaviors is very similar to the idea of usage and renewal being outcomes of the maximization of a common utility function. We do not use the term "utility" to refer to the latent variable because we are not using a formal utility maximization framework when developing our model.

[5] We acknowledge that the concept of commitment has been defined and previously studied in the marketing literature (e.g., Morgan and Hunt 1994, Garbarino and Johnson 1999, Gruen et al. 2000). Its theoretical definition and measurement are beyond the scope of this paper.

[6] We support this claim in our empirical analysis when we compare our approach to other methods.

compelling and intuitive story of customer behavior in contractual settings that is easy to communicate to a nontechnical audience.

Consistent with much of the modeling literature in the customer relationship management (CRM) area, we assume that the behaviors of interest are measured in discrete time; for example, Kumar et al. (2008) and Venkatesan and Kumar (2004) use months, whereas Borle et al. (2008) and Bonfrer et al. (2010) use weeks. This same literature typically treats usage data in one of two ways. The first focuses on the time between events, with a submodel for the usage (e.g., expenditure, number of units purchased) associated with each event (e.g., Venkatesan and Kumar 2004, Borle et al. 2008). The second sums usage across all the events that occur in a given time interval, giving us the total usage per period (e.g., total expenditure per month, total number of units purchased per week), and an appropriate model for that quantity is specified (e.g., Kumar et al. 2008, Sriram et al. 2012).[7] In this research we take the second approach because it fits naturally with our assumed data generating process.

Before we formally develop the model, let us note other streams of literature potentially relevant to our modeling objective. In essence, the data we seek to model are from a bivariate longitudinal process in which one variable is contingent on the other (i.e., customers need to get access to the service), and one of the variables follows a binary and absorbing mechanism (i.e., churn is absorbing). As such, it appears that a discrete/continuous model of consumer demand (e.g., Hanemann 1984, Krishnamurthi and Raj 1988, Chintagunta 1993) could be used to address this problem. This type of model was proposed in the marketing and economics literature to model joint decisions (binary/continuous) such as "whether to buy" and, if so, "how much to buy." The underlying assumption of such models is that all the involved decisions are consequences of optimizing a common utility function. Although they have been extended to handle dropout (e.g., Narayanan et al. 2007), we cannot automatically make use of these models because they do not accommodate the two different time scales inherent in the data typically at an analyst's disposal.

Another possible starting point is the biostatistics literature on modeling longitudinal data with dropout or a terminal event (e.g., Diggle and Kenwark 1994, Henderson et al. 2000, Xu and Zeger 2001, Hashemi et al. 2003, Cook and Lawless 2007) and, to a lesser extent, related work in the fields of marketing and economics (e.g., Hausman and Wise 1979, Danaher 2002), where different methods have been proposed to model longitudinal data with some type of attrition process. With variations particular to each application area, these longitudinal models see the joint estimation of a measurement process and a survival function. However, such models do not address our research objective. First, they focus on controlling for dropout-induced bias, rather than predicting dropout. And second, these methods do not naturally lend themselves to the two-clock nature of the problem we are addressing.

A number of marketing researchers have proposed various dynamic latent models to capture consumers' evolving behavior. For example, Sabavala and Morrison (1981), Fader et al. (2004), and Moe and Fader (2004a, b) present nonstationary probability models for media exposure, new product purchasing, and website usage, respectively; Netzer et al. (2008) and Montoya et al. (2010) develop models of charitable giving and drug-prescribing behavior. None of these models is directly relevant to our modeling problem because they assume a univariate action given the latent state. Nevertheless, our proposed model is influenced by their work.

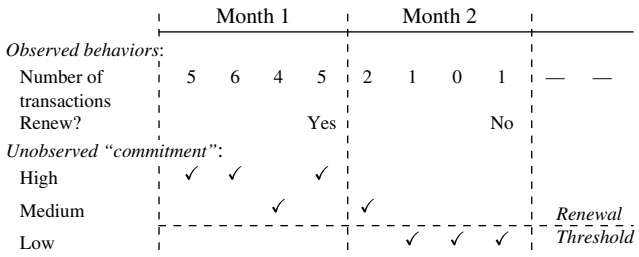
3. The Model

We now turn our attention to the specification of our model, first outlining the intuition of the model. We assume that observed usage and renewal behaviors reflect the individual's commitment at any point in time. We assume that this latent variable is discrete in nature and evolves over time in a stochastic manner. Variations in commitment are reflected in usage behavior (e.g., number of transactions). The decision of whether or not to renew the contract when it comes up for renewal also reflects the individual's commitment at that point in time; if her commitment is below a certain threshold, she will not renew her contract. Consistent with the nature of contractual businesses, we assume that churn is absorbing (i.e., once a customer churns, we do not observe her behavior any more). The fact that an individual is under contract at a particular point in time implies that her commitment had to be above some renewal threshold in all preceding renewal periods.

With reference to Figure 3, where we have a monthly subscription with usage observed on a weekly basis, the fact that this person is active in the second month means that she renewed in week 4, which implies that her unobserved commitment was above the renewal threshold in that particular week. However, it does not tell us anything about the level of the latent variable in weeks $1, 2, 3, 5, \ldots$; this has to be inferred from her usage behavior. As her commitment is below the renewal threshold in week 8,

[7] At one level, the two approaches are equivalent, given the interplay between timing and counting processes.

Figure 3 Model Intuition



she does not renew her contract at the second renewal opportunity.

We now present our model more formally. We start by specifying the model for the basic case where only data on customers' usage and renewal decisions are stored in the firm's transaction database. Following a discussion of parameter identification, we explore how covariate effects (e.g., demographic variables, marketing activities, seasonality) can be incorporated into the model. So far, we have purposefully been vague about how usage is being measured. Our proposed modeling approach is suitable for a variety of usage measures (e.g., the number of transactions, purchase volume, total expenditure). For expositional purposes, and given the nature of our empirical setting, we start by assuming that usage is a count process (e.g., number of transactions). This assumption is relaxed in §3.4, where we discuss further model extensions. In particular, we present the required model adjustments for settings where different usage measures are used.

3.1. Basic Model Specification

The model comprises three processes, all occurring at the individual level: (i) the underlying commitment process that evolves over time, (ii) the usage process that is observed every period, and (iii) the renewal process that is observed every n periods, where n denotes the number of usage periods associated with each contract/subscription agreement. For example, a setting in which there is an annual subscription and usage is observed on a quarterly base implies $n = 4$. (For those special settings where the usage and renewal processes operate on the same clock, $n = 1$.)

3.1.1. The (Unobserved) Commitment Process. We assume the existence of a latent variable—which we label *commitment*—that represents the predisposition of the customer to purchase (or use) the products or services associated with the contract, as well as her predisposition to continue the relationship with the firm. To capture temporal changes in customer behavior, we allow this individual-level latent variable to change over time in a stochastic manner. In particular, we assume that this latent variable is discrete and follows a (hidden) Markov process.[8]

More formally, let t denote the usage time unit (periods) and let i denote each customer ($i = 1, \ldots, I$). Let us assume that there exists a set of K states $\{1, 2, \ldots, K\}$, with 1 corresponding to the lowest level of commitment and K the highest. These states represent the possible commitment levels that each customer could occupy at any point in time during her relationship with the firm. We assume that S_{it}, the state occupied by person i in period t, evolves over time following a Markov process with transition matrix $\mathbf{\Pi}_i = \{\pi_{ijk}\}$. That is,

$$P(S_{it} = k \mid S_{it-1} = j) = \pi_{ijk}, \quad j, k \in \{1, \ldots, K\}. \tag{1}$$

Consistent with past research using hidden Markov models (HMMs), we assume that the number of latent states is common across all customers.

We allow individuals to differ in the probabilities with which they move among the latent states;[9] this is accommodated using a Dirichlet mixing distribution for each row in the transition matrix:

$$f(\mathbf{\Pi}_i \mid \mathbf{A}) = \prod_{j=1}^{K} f(\boldsymbol{\pi}_{ij} \mid \boldsymbol{\alpha}_j), \tag{2}$$

$$\boldsymbol{\pi}_{ij} \sim \text{Dirichlet}(\boldsymbol{\alpha}_j), \quad j = 1, \ldots, K, \tag{3}$$

where $\mathbf{A} = \{\alpha_{jk}\}_{j,k=1,\ldots,K}$ denotes the matrix containing the population parameters determining the transition probabilities, $\boldsymbol{\alpha}_j$ is the jth row of $\mathbf{A}$ ($[\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jK}]$), and $\boldsymbol{\pi}_{ij}$ is the jth row of $\mathbf{\Pi}_i$ ($[\pi_{ij1}, \pi_{ij2}, \ldots, \pi_{ijk}]$).

This choice of mixing distribution is not only parsimonious but also computationally convenient because the Dirichlet distribution is the conjugate prior of the multinomial process governing transitions between the (latent) states (Scott et al. 2005). The Dirichlet specification does not impose any correlation between the rows of the transition matrix. That is, it could be the case that the π_{ijk} are very heterogeneous across the population whereas the π_{ikj} are not.

[8] Alternatively, we could model commitment as a continuous latent variable and use a Kalman filter (e.g., Xie et al. 1997, Naik et al. 1998) to model the evolution of the underlying process. However, consistent with previous work that has modeled the dynamics of the firm–consumer relationship (Netzer et al. 2008, Schweidel et al. 2011), we choose to take a nonparametric approach instead of being tied to a parametric form for the evolution of the latent variable. Moreover, assuming discrete values for the level of commitment facilitates the managerial interpretation of the proposed model. The latter point will become clearer once we discuss the implications of the model results.

[9] This accommodates heterogeneous churn rates and therefore increasing aggregate retention rates, a pattern typically observed when analyzing cohorts of customers in contractual settings (Fader and Hardie 2010).

Finally, we need to establish the initial conditions for the commitment states in period 1. We assume that the probability that customer i belongs to commitment state k at period 1 is determined by the vector $\mathbf{q} = [q_1, q_2, \ldots, q_K]$, where

$$P(S_{i1} = k) = q_k, \quad k = 1, \ldots, K. \tag{4}$$

3.1.2. The (State-Dependent) Usage Process. While under contract, a customer's usage behavior is observed every period. This behavior reflects her underlying commitment—for any given individual, we would expect higher commitment levels to be reflected by higher usage levels. At the same time, we acknowledge that individuals may have different intrinsic levels of usage (i.e., unobserved cross-sectional heterogeneity in usage patterns).

We assume that, for individual i in (unobserved) state k, usage (e.g., the number of transactions) in period t follows a Poisson distribution with parameter

$$\lambda_{it} \mid [S_{it} = k] = \theta_k \beta_i. \tag{5}$$

That is, the usage process is determined by an individual-specific parameter β_i that remains constant over time and a state-dependent parameter θ_k that takes the same value for all customers who belong to commitment level k.[10]

The parameter β_i captures heterogeneity in usage across the population, thus allowing two customers with the same commitment level to have different transactions patterns. In other words, individuals with higher values of β_i are expected, on average, to have a higher transaction propensity than those with lower values of β_i, regardless of their commitment level. The parameter β_i is assumed to follow a lognormal distribution with mean 0 and standard deviation σ_β.[11]

The vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]$ of state-specific "mean usage" parameters measures the change in usage behavior as a result of changes in underlying commitment. To ensure positive values of λ_{it}, we make the restriction that $\theta_k > 0$ for all k. Furthermore, we impose monotonicity between θ_k and the level of commitment (i.e., $0 < \theta_1 < \theta_2 < \cdots < \theta_K$), which implies that for each customer, the expected level of usage is increasing with her commitment level. (Note that allowing for switching between commitment states means we can accommodate overdispersion in individual transaction behavior while still using the Poisson distribution.)

For each customer i we have a total of T_i usage observations. Let y_{it} be customer i's observed usage in period t, and let $\tilde{S}_i = [S_{i1}, S_{i2}, \ldots, S_{iT_i}]$ denote the (unobserved) sequence of states to which customer i belongs during the observation window, with realization $\tilde{s}_i = [s_{i1}, s_{i2}, \ldots, s_{iT_i}]$. The customer's usage likelihood function is

$$\begin{aligned}\mathcal{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \mid \tilde{S}_i = \tilde{s}_i, \text{data}) &= \prod_{t=1}^{T_i} P(Y_{it} = y_{it} \mid S_{it} = s_{it}, \boldsymbol{\theta}, \beta_i) \\ &= \prod_{t=1}^{T_i} \frac{(\beta_i \theta_{s_{it}})^{y_{it}} e^{-\beta_i \theta_{s_{it}}}}{y_{it}!}, \end{aligned} \tag{6}$$

where $\theta_{s_{it}}$ takes the value θ_k when individual i occupies state k at time t (i.e., $s_{it} = k$).

3.1.3. The (State-Dependent) Renewal Process. At the end of each contract period (i.e., when $t = n, 2n, 3n, \ldots$), each customer decides whether or not to renew her contract based on her current level of commitment. We assume that a customer does not renew (i.e., churns) if her commitment state at the renewal occasion is the lowest of all possible commitment levels (i.e., $S_{it} = 1$); otherwise, she renews. In addition, given that in period 1 all customers have freely decided to take out a contract, we restrict the commitment state in the first period to be different from 1 (i.e., we restrict $q_1 = 0$ in (4)).

It is worth noting that churn is an absorbing process. Therefore, if a customer is active in period t, her commitment state in all preceding *renewal* periods ($n, 2n, \ldots \leq t$) must have been greater than 1; otherwise, she would not have renewed her contract and no activity could have been observed at time t.[12] However, nothing is implied about the underlying states she belonged to in the preceding nonrenewal periods (i.e., $t \neq n, 2n, \ldots$).

To illustrate this, let us consider a gym where membership is renewed monthly and individual visits are observed on a weekly basis. The fact that an individual is active in a particular month implies that she was not in the lowest commitment state at the end of all preceding months (i.e., weeks 4, 8, …). Table 1 shows examples of sequences of commitment states (assuming $K = 3$) that, based on our assumptions regarding the renewal process, could or could not occur in such a setting.

[10] Notice that the process governing usage is assumed to be the same for renewal and nonrenewal periods. Differences in usage behavior are due to differences in the state-dependent parameter only.

[11] We use the lognormal—as opposed to, say, the gamma—distribution simply for reasons of computational convenience.

[12] Although we do not observe such behavior in our empirical application, we acknowledge that in some circumstances we could encounter customers who cancel their subscription and then subscribe again (i.e., are reactivated) after a certain number of periods. In such cases, our model specification would treat these individuals as newly acquired customers. To allow for such behavior in this model, we could adapt the state space to accommodate a "dormant" state from which customers could reactivate their subscription.

Table 1 Illustrative Feasible and Infeasible Commitment State Sequences

Month 1				Month 2						
Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9	⋯	Feasible?
1	3	1	2	2	3	2	2	3		×
2	3	1	1	2	3	2	2	3		×
2	3	2	2	2	3	2	1	3		×
2	3	1	1	—	—	—	—	—		✓
2	3	2	2	2	3	2	1	—		✓
2	3	2	2	2	3	2	2	3		✓
2	1	1	2	2	1	2	2	1		✓

The first three sequences are infeasible. The first sequence of states cannot occur given that, for an individual to have become a customer, her commitment state in period 1 must have (by definition) been greater than 1. The following two sequences of states are also infeasible because if a customer is active in week 9 (the third month), her commitment state at the end of the first and second months (weeks 4 and 8) had to have been greater than 1. However, there are no restrictions about her commitment states in any periods other than 4 and 8, which means the next four sequences shown in Table 1 are feasible.

At first glance, the specification for the churn process might seem restrictive given that renewal behavior is assumed to be deterministic conditional on the commitment state. However, membership of this hidden state evolves in a stochastic and heterogeneous manner. As a result, renewal behavior is modeled probabilistically, and customers are allowed to churn at different rates.

3.1.4. Bringing It All Together. We now combine the three processes to characterize the overall model. For each customer i, we have shown how the unobserved sequence $\tilde{S}_i$ determines her renewal pattern over time. Moreover, conditional on her $\tilde{S}_i = \tilde{s}_i$, the expression for the usage likelihood was derived. To remove the conditioning on $\tilde{s}_i$, we need to consider all possible paths that $\tilde{S}_i$ may take, weighting each usage likelihood by the probability of that path:

$$\mathcal{L}_i(\mathbf{\Pi}_i, \mathbf{q}, \boldsymbol{\theta}, \beta_i \mid \text{data}) = \sum_{\tilde{s}_i \in \Upsilon_i} \mathcal{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \mid \tilde{S}_i = \tilde{s}_i, \text{data}) f(\tilde{s}_i \mid \mathbf{\Pi}_i, \mathbf{q}), \quad (7)$$

where Υ_i denotes all possible commitment state paths customer i might have during the observation window, $\mathcal{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \mid \tilde{S}_i = \tilde{s}_i, \text{data})$ is given in (6), and $f(\tilde{s}_i \mid \mathbf{A}, \mathbf{q})$ is the probability of path $\tilde{s}_i$. If there were no restrictions as a result of the renewal process, the space Υ_i would include all possible combinations of the K states across T_i periods (i.e., K^{T_i} possible paths). However, as discussed earlier, the nature of the renewal process places constraints on the underlying commitment process. If $T_i = n, 2n, \ldots,$ and the customer did not renew her contract, Υ_i contains $(K-1)^{T_i/n} K^{T_i - T_i/n - 1}$ possible paths; otherwise, Υ_i contains $(K-1)^{\lfloor (T_i-1)/n \rfloor + 1} K^{T_i - \lfloor (T_i-1)/n \rfloor - 1}$ paths.

Considering all customers in our sample, and recognizing the heterogeneity in β_i, the overall likelihood function is

$$\mathcal{L}(\mathbf{A}, \mathbf{q}, \boldsymbol{\theta}, \sigma_\beta \mid \text{data}) = \prod_{i=1}^{I} \int_0^\infty \int_{\omega(\boldsymbol{\pi}_{i1})} \cdots \int_{\omega(\boldsymbol{\pi}_{iK})} L_i(\mathbf{\Pi}_i, \mathbf{q}, \boldsymbol{\theta}, \beta_i \mid \text{data}) \cdot f(\mathbf{\Pi}_i \mid \mathbf{A}) f(\beta_i \mid \sigma_\beta)\, d\mathbf{\Pi}_i\, d\beta_i, \quad (8)$$

where $\omega(\boldsymbol{\pi}_{ij})$ is the simplex $\{[\pi_{ij1}, \pi_{ij2}, \ldots, \pi_{ijk}] \mid \pi_{ijk} \geq 0; k = 1, \ldots, K; \sum_{k=1}^{K} \pi_{ijk} = 1\}$.

To summarize, we propose a joint model of usage and churn in which the two behaviors (which typically occur on different time scales, but need not) reflect a common dynamic latent variable modeled using a hidden Markov model. Churn is deterministically linked to the latent variable, whereas usage is modeled as a state-dependent Poisson process that incorporates time-invariant cross-sectional heterogeneity.

We estimate the model parameters using a hierarchical Bayesian framework. In particular, we use data augmentation techniques to draw from the distribution of the latent states S_{it} as well as the individual-level parameters β_i and $\mathbf{\Pi}_i$. We control for the path restrictions (because of the nature of the contract renewal process) when augmenting the latent states. As a consequence, the evaluation of the likelihood function becomes simpler, reducing to the expression of the conditional (usage) likelihood function, $\mathcal{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \mid \tilde{S}_i = \tilde{s}_i, \text{data})$. See Web Appendix A (available as supplemental material at http://dx.doi.org/10.1287/mksc.2013.0786) for further details.

3.2. Parameter Identification

This basic model has $K^2 + (K-1) + K + 1$ population parameters, which are the elements of $\mathbf{A}$, $\mathbf{q}$, $\boldsymbol{\theta}$, and σ_β, respectively. The Markov process (determined by parameters $\mathbf{A}$ and $\mathbf{q}$) captures both churn and changes in usage behavior, whereas the Poisson process (determined by parameters $\boldsymbol{\theta}$ and σ_β) links these underlying dynamics to usage behavior alone.

The challenge in estimation is to separate individual dynamics from cross-sectional heterogeneity and the general randomness associated with the Poisson usage process. We have a number of observations of individual usage, but there are typically fewer for renewal behavior. This scarcity in renewal data is due to two factors: renewal only happens every n periods, and the churn process is absorbing (i.e., once a customer churns, we do not observe her behavior

anymore). Nevertheless, given the deterministic link between the latent state and renewal behavior (i.e., in any given renewal period, customers churn if and only if they are in state 1, the lowest commitment state), all renewal observations are very informative and are therefore necessary for identification.

Let us outline the intuition of how the model parameters are identified.[13] We estimate the state-specific mean usage parameters of the Poisson process ($\boldsymbol{\theta}$) mostly from usage behavior during the renewal periods; in each renewal period ($t = n, 2n, \ldots$), we know for certain the underlying state for churners, and we have partial information about the underlying commitment for nonchurners (i.e., they are not in the lowest state). Given that the variance of the Poisson equals its mean and that the number of states is set a priori, observed usage variation across customers allows us estimate cross-sectional heterogeneity in usage behavior (i.e., σ_β).

Dynamics in the latent variable are identified from differences in usage behavior (within customers) as well as in churn decisions over time. The parameters of the Dirichlet distribution ($\mathbf{A}$) jointly determine the mean and variance of the transition probabilities, reducing the burden on the data for identifying the heterogeneity in probabilities across people. Given $\boldsymbol{\theta}$ and σ_β, the mean transition probabilities are identified (mostly) from the dynamics in observed usage behavior. Furthermore, changes in retention rates over time—in particular, the fact that cohort-level retention rates increase over time—help us identify the variance (i.e., unobserved heterogeneity) in the transition probabilities. Finally, the initial states probabilities ($\mathbf{q}$) are primarily identified from differences in usage behavior in the first period.

3.3. Incorporating Covariates in the Model

In many situations the firm will have reliable customer demographic data and/or information about the interactions between the firm and the customers. The latter case might include a wide range of marketing actions, from "untargeted" advertising campaigns or promotional activities aimed at all customers to individual-level direct mail and email communications.

This additional information can be incorporated into the model in different ways. If we expect the customer-level covariates to explain cross-sectional variation in overall usage levels or if the time-varying marketing actions are expected to have a short-term impact on usage alone (and not on other behaviors), we could simply make the usage rate in Equation (5) a function of the available covariates:

$$\lambda_{it} \mid [S_{it} = k] = \theta_k \beta_i \exp(\boldsymbol{\delta}_1 \mathbf{x}_i + \boldsymbol{\delta}_2 \mathbf{z}_{it}), \qquad (9)$$

where $\boldsymbol{\delta}_1$ reflects how mean usage varies as a function of the individual time-invariant covariates denoted by $\mathbf{x}_i$, and $\boldsymbol{\delta}_2$ captures the effects of the time-varying covariates (e.g., marketing activities at time t) denoted by $\mathbf{z}_{it}$.[14]

Moreover, the observed covariates could have a more persistent effect on customer behavior or, in other words, moderate or influence customer commitment. In that case one could easily incorporate covariate effects in the transition probabilities. For reasons of mathematical convenience, a heterogeneous multinomial logit (or probit) model could be used in place of the Dirichlet distribution. Alternatively, we could follow the approach of Netzer et al. (2008) and Montoya et al. (2010) and use an ordered logit (or probit) model to incorporate such covariate effects.

Ideally, one would include covariates in both the usage and latent commitment processes. This approach would allow us to separate the impact of marketing actions on usage versus renewal behavior. Although plausible, such an approach would require a significant amount of within- and between-individual variation in the data to be able to separate the two effects.[15] In a similar manner, one could also include information on competitors' marketing actions when modeling both usage rates and transition probabilities. This approach would be particularly interesting in highly competitive markets, such as telecommunications or financial services, in which churn is generally observed because the customer has switched to a direct competitor. The fundamental challenge faced by the analyst would be how to gain access to the relevant competitive information.

3.4. Variants on the Basic Model

The proposed model assumes that, conditional on the underlying state, usage behavior is characterized by a Poisson distribution. Behaviors for which this specification is appropriate include the number of credit card transactions per month, the number of movies purchased each month in a pay-TV setting, and the number of phone calls made per week. However, in some settings, the usage level has an upper bound, either because of capacity constraints from the company's side or because the time period in which usage is observed is short. Going back to the above-mentioned gym example, if one wants to model the

[13] We have also run simulation analyses to corroborate this. The results are reported in Web Appendix B.

[14] As a particular case of the latter, one can easily incorporate seasonal dummies or any other type of time-varying information to control for seasonality and time trends. This is examined empirically in Web Appendix C.

[15] Experimental data would provide a perfect scenario for measuring such effects, because this would not only solve any identification issues but also alleviate the potential endogeneity of marketing actions.

number of days a member attends in a particular week, the Poisson may not be the most appropriate distribution because there is an upper bound of seven days. Similarly, an orchestra will have a fixed number of performances in any given booking period, and the analyst may wish to acknowledge this upper bound when modeling usage. In these cases the Poisson distribution should be replaced by the binomial distribution in which the upper bound (e.g., the number of days in a week, total number of performances offered) is the number of trials.

There are also situations where usage is not discrete; for example, usage could refer to time (e.g., minutes used in wireless contracts), expenditure (e.g., total amount spent), or other nonnegative continuous quantities (e.g., MBs downloaded in a wireless data plan). The proposed model is easily applied in such settings, provided the distributional assumption of the usage process is modified; we simply replace the Poisson or binomial with, say, a gamma or lognormal distribution. (Details of these alternative specifications are provided in Web Appendix D.)

Finally, there could be cases in which usage and renewal occur each and every period (i.e., operate on the same clock), or alternatively, the firm temporally aggregates the usage data (e.g., a gym offering monthly subscriptions and recording the total number of attendances in each month). The model presented in §3.1 can easily accommodate such cases. One simply needs to set $n = 1$ in the specification of the renewal process, which basically restricts the lowest commitment state to be absorbing. (As a consequence, the transition matrix would have fewer elements because there is no possibility to move from state 1.)

4. Empirical Analysis

4.1. Data

We explore the performance of the proposed model using data from an organization in which an annual subscription/membership is required to gain the right to buy (or use) its products and services, as is the case for some "warehouse clubs" and priority-booking schemes for cultural organizations. Membership of this scheme also provides subscribers with additional benefits, including newsletters and invitations to special events.

In addition to the membership fee, subscribers are an important source of revenue for the organization through their usage behavior. The company generates approximately \$5 million a year from membership fees alone and a further \$40 million from members' transactions. Each year is divided into four "buying" periods; all members receive a catalog each period with information about the products offered and complete an order form. When one's membership is close to expiring (generally one month before the expiration date), the organization sends out a renewal letter. If membership is not renewed, the benefits can no longer be received.

We focus on the cohort of individuals that took out their initial subscription during the first quarter of 2002 and analyze their buying and renewal behavior for the following four years. Expressing these data in terms of periods (as we defined t in §3), we have a total of 17 periods. We observe usage in periods 1–16 and renewal decisions in periods 5, 9, 13, and 17.[16]

4.1.1. Some Patterns in the Data. Of the 1,173 members of this cohort, 884 renewed at the end of their first year (75% renewal rate), 738 renewed at the end of year 2 (83% renewal rate), 634 renewed at least three times (86% renewal rate), and 575 were still active after four renewal opportunities.

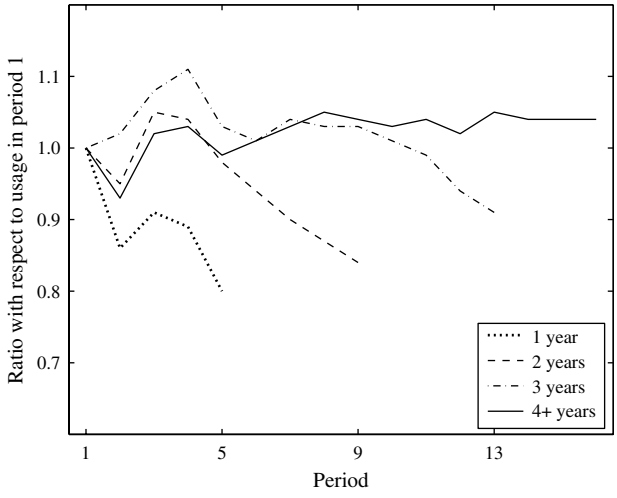
This cohort of customers made a total of 14,255 purchases across the entire observation period. On average, a subscriber made 1.05 purchases per period. However, the transaction behavior was very heterogeneous across subscribers, with the average number of purchases per period ranging from 0 to 41.9. (We use the words "purchase" and "transaction" interchangeably.) There was also variability in within-customer variation in the number of transactions. The individual coefficient of variation for usage ranged from almost 0 (customers whose transaction history was very stable) to 4 (customers with high variation in their period-to-period purchasing behavior).

To examine the observed relationship between usage and renewal behaviors, we split the cohort into four groups depending on how long they were members of the organization (1 year, 2 years, etc.) and look at the evolution of their usage behavior over time. Figure 4 plots the cumulative moving average of the number of transactions by period (indexed against period 1). We observe that, on average, customers decrease their usage prior to churn.

To examine whether this pattern also occurs at the customer level, we analyze individual-level usage behavior at the end of each customer's observation period. We find that for the majority of customers, the number of transactions decreases before churn: for 70% of churners, transaction levels in the last two

[16] In developing the logic of our model, we discussed a contract period of $n = 4$ with renewal occurring at 4, 8, 12, etc. For the case of quarterly periods, this implicitly assumed that customers are acquired immediately at the beginning of the period (e.g., January 1, the first day of Q1) with the contract expiring at the end of fourth period (e.g., December 31, the last day of Q4). In this empirical setting, customers are acquired throughout the first period, which means the first renewal occurs sometime in the fifth period.

Figure 4 **Indexed Cumulative Moving Average of Usage, by Duration of Membership**



periods of their relationship are below their individual averages. We also compute the ratio of usage in each individual's last two observed periods to that of their first two periods. The average ratio for churners is 0.47, versus 0.83 for those individuals who were still subscribers at the end of the observation period.

4.2. Model Estimation and Results

We split the four years of data into a calibration period (periods 1–11) and a validation period (periods 12–17). We first need to determine the number of hidden states in the Markov chain. We estimate the model varying the number of states from two to four and compute (i) the log marginal density, (ii) the deviance information criterion (DIC), and (iii) the in-sample mean square error (MSE) for the predicted number of transactions (per individual, per usage period). As shown in Table 2, the specification with the best log-marginal density and DIC is the model with three hidden states. (The Bayes factor of this specification, compared with a more parsimonious model, also gives support for the three-state model.) We also find that the model with three states has the best individual-level in-sample predictions, with an MSE of 1.45.

Table 3 presents the posterior means and 95% central posterior intervals (CPIs) for the parameters of the usage process under the three-state specification. The first set of parameters (θ_ks) corresponds to the usage

Table 2 **Measures of Model Fit**

No. of states	Log marginal density	DIC	Individual MSE
2	−16,263	80,546	1.93
3	−14,967	73,518	1.45
4	−15,313	75,065	1.53

Table 3 **Parameters of the Usage Process with Three States**

	Parameter	Posterior mean	95% CPI
Usage	θ_1	0.20	[0.18 0.22]
Propensity	θ_2	0.21	[0.19 0.22]
	θ_3	1.20	[1.14 1.28]
Heterogeneity	σ_β	0.90	[0.84 0.96]

parameters common to all customers in each commitment level, and σ_β measures the degree of unobserved heterogeneity in usage behavior within each state.

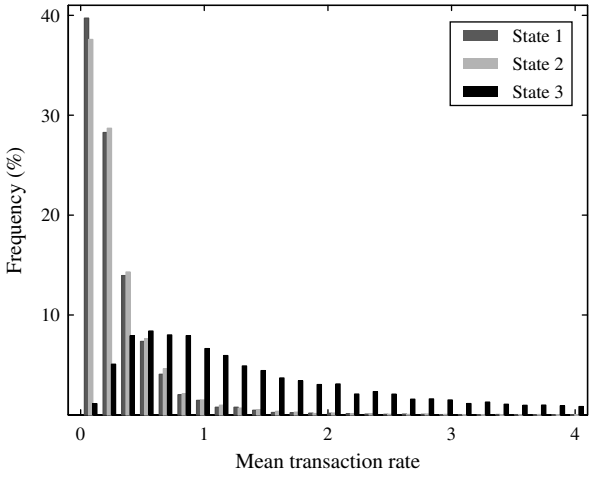
We note that the posterior means for θ_1 and θ_2 are very similar. Recalling (5), the distributions of the state-specific Poisson means for all individuals are reported in Figure 5. Integrating over the distribution of β_i, we find that the average of the state-specific Poisson means (across individuals) are 0.30 for state 1 and 0.32 for state 2. The important difference between these two states is with regard to renewal behavior. The interpretation of each state is determined by the transaction propensity and the renewal behavior. Hence, althouth those individuals in state 1 will, on average, make ever-so-slightly fewer transactions than those in state 2, they will churn if they belong to that state during a renewal period. On the contrary, individuals in state 2 will renew their membership even though they may also make few purchases.[17] For individuals in state 3, the highest commitment level, the average number of transactions per period is 1.80, which translates to more than seven transactions in a year, provided the customer stays in the highest state during the whole year.

Dynamics in the latent variable are captured by the hidden Markov chain. The top part of Table 4 shows the posterior estimate of $\mathbf{q}$, which represents the initial conditions for the commitment states in period 1. This is the distribution of underlying states for a just-acquired member of this cohort. We note that customers were equally distributed between states 2 and 3 when they took out their first subscription. The bottom part of Table 4 shows the posterior estimates of the (Dirichlet) parameters that determine the transition matrix.

For easier interpretation, we report in Table 5 the average and 95% interval of the individual posterior means of the transition probabilities. That is, we obtain the posterior distribution of the elements of the matrix for each individual. We then compute the posterior means of these quantities (for each individual) and report the overall mean and 95% interval across all individuals. For example, the last two rows

[17] Given that θ_1 and θ_2 are close to each other, identification of states 1 and 2 during a nonrenewal period comes mainly from differences in observed usage and renewal behavior in periods other than the current period.

Figure 5 **Distributions of the State-Specific Usage Process Poisson Means**



should be read as follows: for an average individual in state 3, the probability of remaining in state 3 is 0.72, the average probability of switching to state 2 in the next period is 0.23, and the average probability of switching to the lowest commitment state is 0.05. Note that individuals do not switch states with the same propensity; if we look at individuals within the 95% interval of individual posterior means, their (posterior mean) probabilities of switching from state 3 to state 2 range from 0.06 to 0.34.

Care must be taken when interpreting the state 1 transition probabilities ($\bar{\pi}_{11} = 0.60$, $\bar{\pi}_{12} = 0.38$, and $\bar{\pi}_{13} = 0.02$), because these probabilities only apply to nonrenewal periods. The model assumes that customers churn if they are in the lowest commitment state during a renewal period; however, not all periods are renewal occasions. Therefore, it is possible to find individuals who were in state 1 at a particular time (nonrenewal period) and changed their commitment state before the renewal occasion occurred. The fact that the estimate of $\bar{\pi}_{11}$ is less than 1 implies that the data do not want this state to be absorbing in this empirical setting.

Table 4 **Parameters of the Commitment Process with Three States**

Parameter	Posterior mean	95% CPI
q_1	0.00	—
q_2	0.50	[0.44 0.55]
q_3	0.50	[0.44 0.55]
α_{11}	38.17	[32.32 46.80]
α_{12}	24.20	[16.39 31.30]
α_{13}	1.13	[0.75 1.66]
α_{21}	0.25	[0.23 0.28]
α_{22}	0.28	[0.21 0.37]
α_{23}	0.21	[0.19 0.23]
α_{31}	0.13	[0.12 0.15]
α_{32}	0.62	[0.55 0.70]
α_{33}	1.94	[1.73 2.12]

Table 5 **Mean Transition Probabilities and the 95% Interval of Individual Posterior Means**

	To state		
From state	1	2	3
1	0.60	0.38	0.02
	[0.60 0.61]	[0.37 0.38]	[0.02 0.02]
2	0.34	0.38	0.28
	[0.14 0.66]	[0.21 0.69]	[0.10 0.61]
3	0.05	0.23	0.72
	[0.01 0.11]	[0.06 0.34]	[0.58 0.93]

To get a tangible sense of how the model fits the data, we compare the actual and predicted levels of usage in the calibration period. We find that the three-state specification of the proposed model gives a very good fit when predicting total usage: the mean absolute percentage error (MAPE) for the total number of transactions per period is 7.44%. However, being able to track aggregate levels of usage is not enough; we would expect the model to capture cross-sectional differences too. We compute the posterior distributions of the maximum number of transactions, the minimum number of transactions, and five common percentiles of the transaction distribution. Comparing these with the actual numbers, we observe in Table 6 that all the summaries of the actual usage behavior lie within the 95% CPI of the posterior distributions. These results, combined with the good aggregate fit, confirm that the model predictions in the calibration period are accurate.

To further assess the validity of the model, we look at the relationship between the observed behaviors (both usage and renewal) and the latent states to which customers are assigned. First, we group the customers depending on their levels of usage (i.e., whether recent usage is below/above their regular levels). Then we look at both the probability of being assigned to each latent state and the actual renewal behavior at the end of the current contract. Regarding usage, we confirm that the probability of being assigned to lower states increases when a customer's

Table 6 **Comparing the In-Sample Distribution of Transactions with the Model Predictions**

	Actual	Posterior mean	95% CPI
Min	0.00	0.00	[0.00 0.00]
5%	0.00	0.00	[0.00 0.00]
25%	2.00	1.94	[1.00 2.00]
50%	4.00	4.56	[4.00 5.00]
75%	11.00	11.07	[10.25 12.00]
95%	32.85	33.10	[31.00 35.00]
Max	457.00	453.04	[396.00 513.00]

usage in recent period is below their average. Regarding renewal, we observe that individuals assigned to state 1 have much higher churn rates than those assigned to higher states. (Details of this analysis are provided in Web Appendix E).

Now that we have shown the quality of the model predictions in the calibration period, we examine the performance of the model in the holdout validation period.

4.3. Forecasting Performance

Based on the posterior distributions of the model parameters, we can easily predict each customer's future underlying "commitment" in any given period. In particular, we forecast underlying commitment for periods 12 to 17. Once future latent states are "known," predicting usage and renewal behavior follows naturally given the model assumptions about usage and renewal behaviors. First, we forecast usage behavior in period 12 for all those members that were active at the end of our calibration period. (Notice that period 12 is a nonrenewal period. That is, customers do not make renewal decisions until period 13.) Then, conditional on each individual's (predicted) underlying state in period 13, we determine renewal behavior at that particular moment. Finally, conditional on having renewed at that time, we forecast usage behavior for all remaining periods and renewal behavior for the last period of data. (Given our use of a hierarchical Bayesian framework, we perform this customer-level forecasting exercise for each draw of the Markov chain (once it has converged) and report the posterior means.) This time-split structure allows us to analyze separately *usage* forecast accuracy (comparing the actual versus predicted number of transactions in period 12), *renewal* forecast accuracy (comparing renewal rates in periods 13 and 17), and *overall* forecast accuracy (comparing usage levels from period 14 onwards).

4.3.1. Benchmark Models. We compare the accuracy of the model forecasts with those obtained using the following set of benchmarks: (i) heuristics based on the work of Wübben and Wangenheim (2008); (ii) RFM (recency, frequency, and monetary value) methods widely used among researchers and practitioners; (iii) a bivariate econometric model that jointly estimates submodels for the two behaviors of interest; and (iv) two restricted versions of our proposed dynamic latent variable model.

Heuristics. We consider two heuristics for predicting expected usage: Heuristic A ("periodic usage") assumes that each individual repeats the same pattern every year, and Heuristic B ("status quo") assumes that all customers will make as many transactions as their current average.[18] We also consider two heuristics for predicting churn: Heuristic C ("no usage") assumes that churn occurs if there is no usage activity during the last two periods, and Heuristic D ("lower usage") assumes that churn occurs if an individual's average usage over the last two periods is lower than that of the corresponding periods in the previous year. (The latter heuristic is in the spirit of Berry and Linoff's 2004 discussion of how changes in usage can be a leading indicator of churn.)

RFM Models. As previously noted, it is standard practice to develop models that predict next-period customer behavior as a function of past behavior. This past behavior is frequently summarized in terms of recency, frequency, and monetary value. We consider two random-effects Poisson regression models for predicting usage. The first model ("cross-sectional") uses data from period t to predict usage in period $t+1$, and the second model ("panel") uses data from periods $1, \ldots, t$ to predict usage in period $t+1$. We also consider two logistic regression models (cross-sectional and panel) for predicting churn. (Details of these model specifications and the associated parameter estimates are reported in Web Appendix F.)

A major problem with these models is that they cannot automatically be used to forecast customer behavior multiple periods into the future. For example, the RFM Poisson regression model cannot predict usage in periods $14, 15, \ldots$, because such predictions would be conditional of measures of usage behavior in periods $13, 14, \ldots$, which are unobserved in period 11, the time at which the forecasts are made. Similarly, the RFM logistic regression models cannot predict period 17 churn in period 11 because they are conditioned on usage behavior in period 16. To overcome these limitations, we use a combination of the churn and usage models to make such multiperiod predictions. Regarding multiperiod usage behavior, we first predict renewal behavior in period 13 using the RFM logistic regression models. Then, for those individuals who are predicted to renew in period 13, we use the corresponding RFM usage model recursively, simulating individual transactions, updating the RFM characteristics, and then simulating transactions for the next period. Finally, we predict renewal behavior in period 17 based on the simulated usage behavior in the previous periods.[19] Note that predicted usage in period 12 is needed to compute predictions of both usage and renewal in period 13. This

[18] For example, suppose a customer made 2, 4, 2, 4 transactions over the preceding four periods. Under heuristic A, we would predict that this customer makes 2, 4, 2, 4 transactions over the next four periods. Under heuristic B, we would predict a pattern of 3, 3, 3, 3.

[19] As such, the accuracy of these two sets of forecasts will not reflect the performance of the logistic and Poisson regressions individually, but rather the combined performance of the two models.

is unlike the forecasts obtained using the proposed method, where all predictions are based on the evolution of the underlying commitment state.

Bivariate Model. Another way to model our data is to assume two latent variables where one variable determines usage and the other determines retention. Further, one could allow these two variables to be correlated in order to capture possible dependencies between usage and renewal behaviors. More specifically, we consider a modified version of a Type II Tobit model (Wooldridge 2002). This approach, generally used for data with selection effects, models usage (measurement variable) conditional on renewal being positive (censoring variable) while allowing for correlation in the error terms of both processes. Notice that we say "modified" Type II Tobit model, because (a) the standard Tobit model is not suitable for the two-clock structure considered in this research, and (b) the standard specification it is not appropriate for count data. As a consequence, we modify the likelihood function to use a Poisson instead of a Gaussian process and to make the model suitable for the two-clock nature of our setting. To capture the relationship between usage and renewal, we allow the time-varying shocks governing both decisions to be correlated. We also incorporate the effects of past usage in both equations so as to account for nonstationarity in the usage and renewal decisions. (Details of the model specification and the associated parameter estimates are reported in Web Appendix G.)

Restricted Versions of the Proposed Model. Finally, we estimate two restricted versions of the proposed model: a homogeneous usage model in which the members of each commitment state have the same expected purchase behavior and a homogeneous transition model where the state transition probabilities are the same for all individuals.[20]

4.3.2. Usage Forecast. To assess the validity of the usage predictions, we compare the models' forecasts in period 12 with the actual data. The predictive performance is compared at the aggregate level, looking at the percentage error in the predicted total number of transactions; at the disaggregate level, looking at the histogram of the number of transactions; and at the individual level, looking at the MSE computed across individuals. For the disaggregate-level accuracy, we compute how many customers are expected to have zero transactions, one transaction, two transactions, etc., and we compare these values with the actual data. We assess the similarity of the distributions of the actual and predicted number of transactions using the χ^2 statistic.[21]

Table 7 Assessing the Period 12 Predictive Performance of the Usage Models

	Aggregate (% error)	Disaggregate (χ^2)	Individual (MSE)
Heuristic			
A (periodic)	28.8	16.9	3.0
B (status quo)	21.8	137.7	1.9
Poisson regression			
Cross-sectional	−4.8	19.0	8.0
Panel	−40.5	43.5	2.3
Bivariate model	−12.3	20.7	3.6
Proposed model			
Homogeneous usage	−10.2	29.0	3.2
Homogeneous transitions	−3.6	15.9	1.5
Full specification	−7.2	6.5	1.4

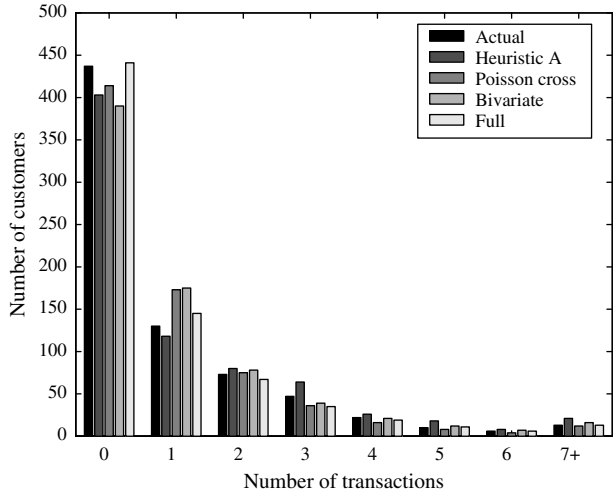
Table 7 shows the error measures for all usage models. If we consider aggregate-level performance, the two best models are the homogeneous transitions specification and the cross-sectional RFM-based random-effects Poisson regression model. However, when we consider predictive performance at the distribution and individual levels, we observe that the full specification of our proposal model outperforms all other models. First, it predicts the distribution of the number of transactions most accurately (having the smallest value of the χ^2 statistic) and has the lowest measure of error in individual-level predictions.

To better understand the meaning of a lower χ^2 (i.e., a better disaggregate fit), we compare in Figure 6 the histogram of the actual number of transactions in period 12 with those predicted by the various models. For the sake of clarity, we select the most accurate method in each set of benchmarks on the basis of the disaggregate predictions in period 12 (see Table 7). The dominance of the full specification of our proposed model is clear from this plot: the height of the columns corresponding to the actual model (first column) and the proposed model (last column) are the closest across the "Number of transactions" bins. Thus, even though the full-model results have a slightly higher percent error at the aggregate level than that associated with the cross-sectional regression model and one of the constrained specifications, these histograms, along with the individual-level MSE numbers, show that it predicts usage more accurately than any of the other methods.

[20] We estimate both specifications varying the number of hidden states. The best-fitting model for the specification with homogeneous usage has three states, whereas that for the homogeneous transition specification has four states.

[21] Note that we are using this statistic as a measure of the "match" between the actual and predicted period 12 histograms, not as a measure of the goodness of fit of the model. As such, we do not report any *p*-values.

Figure 6 Comparing the Predicted and Actual Distributions of the Number of Transactions in Period 12



4.3.3. Renewal Forecast. The predictive performance of all the churn models is presented in Table 8. First, we compare actual versus predicted renewal rates in period 13. As shown in the table, the two dynamic latent variable model specifications with heterogeneous transition probabilities provide the most accurate predictions of the future renewal rate (1.0% and 2.7% error). Both heuristics yield very poor predictions of period 13 churn. The two logistic regression models overestimate future renewal (and therefore the size of the customer base) by more than 10%. We also compute the hit rate (i.e., the percentage of customers correctly classified) for all methods. The full model correctly classifies 78% of customers, the highest among the three dynamic latent variable models. At first glance, it appears that the logistic regression models are better than the proposed model because their hit rates are higher. However, it should be noted that the actual retention rate in the sample is 86%, and these two methods predict that almost every customer renews (98% and 95% predicted renewal rates). As a consequence, the high figures (for hit rate) associated with the logistic regression models are a consequence of their classifying stayers correctly (at the cost of failing to predict churners).

Second, we compare the accuracy of the methods when predicting renewal behavior at the end of our validation period (i.e., period 17). The period 17 predictions for the homogeneous usage specification and the full-model specification are exceptionally accurate at the aggregate level (-0.6% and 0.5%, respectively), with hit rates of 67% and 68%. We note that the predicted renewal rates associated with the homogeneous transition specification are the same for periods 13 and 17; this is a natural consequence of the model specification.

4.3.4. Renewal and Usage Forecast. Finally, we consider the overall forecasting accuracy of the models by examining usage behavior in periods 14–16, which in turn depends on predicted renewal behavior in period 13.

We look at actual versus predicted usage levels in periods 14–16, examining the accuracy of the predictions at the aggregate, disaggregate, and individual levels (see Table 9). Comparing the aggregate MAPE computed across all forecast periods, we find that the full model provides the most accurate predictions over the entire validation period (MAPE = 2.4%). The combination of the RFM-based Poisson and logistic regression models results in very poor estimates of future behavior at the aggregate level. When we consider the disaggregate-level (average χ^2 across the three forecast periods) and individual-level (squared error averaged across the three forecast periods and the 738 customers who were still active at the end of the calibration period) measures of predictive accuracy, we see that the full-model specification is clearly superior.

To conclude, we have shown that our proposed dynamic latent variable model accurately predicts

Table 8 Assessing the Period 13 and Period 17 Predictive Performance of the Renewal Models

	Period 13			Period 17		
	Renewal rate (%)	% error	Hit rate (%)	Renewal rate (%)	% error	Hit rate (%)
Heuristic						
C (no usage)	27	−68.8	37	—	—	—
D (lower usage)	63	−26.5	60	—	—	—
Logistic regression						
Cross-sectional	98	13.6	85	51	−43.6	57
Panel	95	10.9	83	68	−24.7	60
Bivariate model	79	−7.8	71	79	−12.4	68
Proposed model						
Homogeneous usage	87	1.0	77	90	−0.6	67
Homogeneous transition	81	−6.1	73	81	−15.9	60
Full specification	88	2.7	78	91	0.5	68
Actual	86	—	—	91	—	—

Table 9 Assessing the Accuracy of Usage Predictions for Periods 14–16

	Aggregate (MAPE)	Disaggregate (avg. χ^2)	Individual (MSE)
RFM			
Cross-sectional	26.4	291.8	18.5
Panel	68.2	48.2	18.4
Bivariate model	3.0	35.8	5.8
Proposed model			
Homogeneous usage	4.8	84.2	5.0
Homogeneous transition	10.2	26.2	3.5
Full specification	2.4	16.0	3.1

Table 10 Comparing the Present Value of Actual and Predicted Validation-Period Revenue

	Total revenue ($000)	Aggregate % error
RFM cross-sectional	2,386	26
RFM panel	4,548	136
Bivariate model	1,006	−46
Proposed model		
Homogeneous usage	1,825	−2
Homogeneous transition	1,631	−12
Full	1,797	−3
Actual	1,855	

usage and renewal behavior across multiple periods, outperforming a broad set of benchmark methods on a number of dimensions.

4.4. Implications for Customer Valuation

To provide a better sense of how the model accuracy translates into economic terms, we consider validation-period predictions of revenue. In particular, we translate the models' predictions of usage and renewal behavior into the corresponding revenues (in U.S. dollars) generated by each customer during the six periods comprising the validation period.[22] We discount these revenues to the start of the validation period and compare the model-based predictions to the actual numbers. We apply a discount rate of 5% per quarter, which translates to an annual rate of approximately 19%.[23]

The revenue generated by these individuals during the 1.5-year validation period was $1.86 million (see Table 10). Predictions based off the RFM and bivariate benchmark models are inaccurate, whereas the forecast from the proposed model is off by only 3%. (The homogeneous usage specification performs slightly better, $28,000 closer to the actual value.)

The accurate revenue forecasts generated by our proposed model, despite the fact that it assumes "status quo" behavior on the part of the firm, means that it can provide a very useful input to the firm's planning activities as decisions are made about multiperiod investments in customer acquisition activities in order to meet revenue targets.

5. Additional Model Insights

In addition to providing accurate multiperiod forecasts of usage and retention—hence offering a very powerful tool for customer valuation—the dynamic aspect of the model provides additional insights that are managerially useful. Understanding the evolution of customer churn and usage propensity has the potential to provide marketers operating in contractual business with useful information for issues such as segmentation, cross selling, and the design of retention programs. In this section we show how the proposed model can be used to obtain such insights.

5.1. Understanding Individuals' Commitment Patterns

We start by looking at individual-level inferences. Using the estimated model, we can easily compute the distribution of commitment state membership for each customer over her observation window. Recovering the underlying states over time could allow the firm to identify those customers who are likely to have changed (decreased or increased) their commitment state recently. This information would help the marketer differentially target the customers. For example, in our setting, the organization would be interested in knowing, before the membership expiration date, which members have recently suffered a drop in their underlying commitment level, so that preemptive retention activities can be undertaken. (As illustrated by the poor performance of Heuristics C and D in predicting churn, such at-risk customers cannot be identified without the use of a formal model.)

To illustrate how the underlying commitment level relates to observed behavior, we consider the evolution of state membership for three individuals: customer A, who renewed her subscription on all the renewal occasions, and customers B and C, who cancelled their subscriptions after two years (i.e., in period 9). Table 11 shows the observed transaction patterns for these three customers during periods 1–8, and Figure 7 shows how the distributions of state membership vary during periods 5–8.

Let us start by looking at customers A and B. Although both customers have exactly the same usage behavior during periods 5–8 (see Table 11), their inferred commitment patterns are radically different (see Figure 7). Customer A has a very high probability of being in state 3 each period, which we interpret as her being highly committed in all periods. Customer B's probability of belonging to state 1 increases

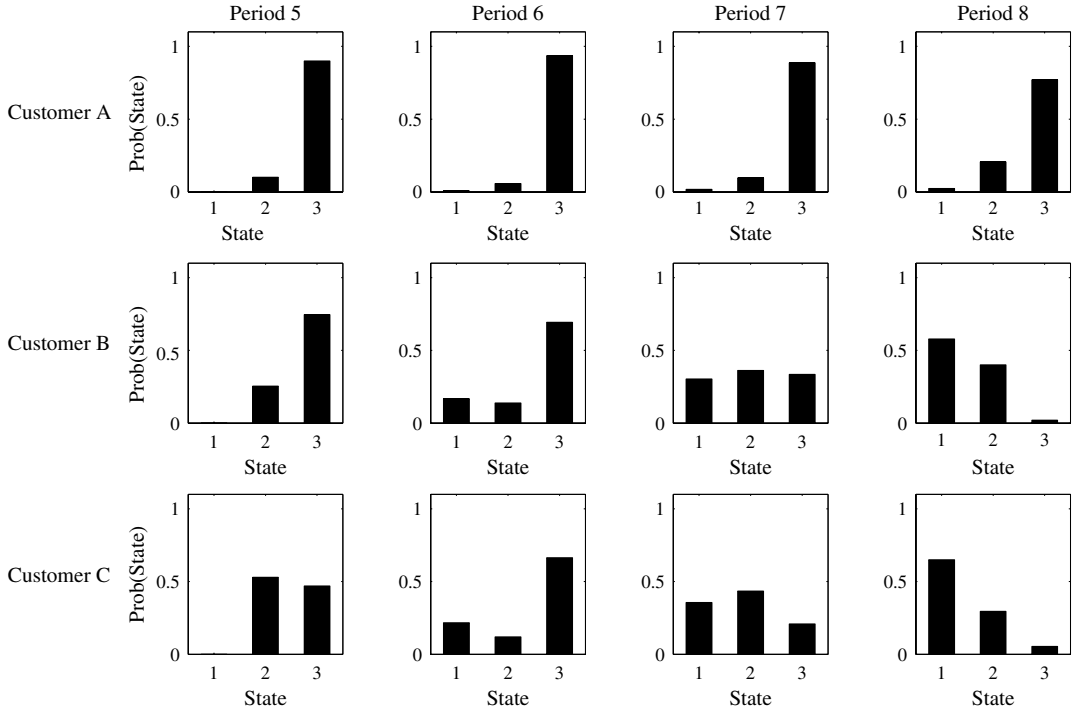
Table 11 Actual Usage Behavior for Three Customers

	Period							
	1	2	3	4	5	6	7	8
Customer A	2	0	2	0	1	2	1	0
Customer B	4	2	3	3	1	2	1	0
Customer C	0	1	1	1	0	2	0	0

[22] We use each individual's average calibration-period spend per transaction when forecasting revenue. Detailed information about the costs incurred by the organization was not available. However, putting aside fixed costs, and given the marketing practices of the organization at the time the data were collected, a constant margin would be an appropriate way to account for the costs of serving those customers.

[23] We replicate the analysis using (quarterly) discount rates ranging from 2% to 7% and obtain qualitatively similar results.

Figure 7 State Membership Dynamics for Three Customers



notably from period 5 to period 8, which we interpret as a drop in her commitment. Why do we see these differences in the underlying states when their behavior in periods 5–8 is the same? The answer is because the state membership probabilities also reflect the differences in their usage over their entire lifetime. As customer B was more active than customer A in year 1, observing one period with zero purchases (period 8) is a likely indicator of a drop in underlying commitment. However, given customer A's past behavior, one period of no purchases does not necessarily indicate a high risk of churn.

Comparing customers B and C, we note that B's purchasing in periods 1–4 is higher than that of C. This is a consequence of the differences in β_i as well as in the commitment levels for those periods; the latter is reflected in the inferred commitment level for period 5—customer B has a high probability of being in the highest commitment state, whereas customer C has an almost equal probability of being in states 2 or 3—as well as in the evolution of commitment in subsequent periods. For example, the jump in customer C's purchasing in period 6 is interpreted as evidence of an increase in commitment. Even though customer B made the same number of purchases in period 6, there is little change in the probability of her being in the highest commitment state because two purchases is not out of the ordinary in light of her transactions in periods 1–4. The inferred probabilities of commitment state membership for these two customers are now basically the same. The subsequent drop in purchasing for customer B and the lack of purchasing by customer C are reflected in the changing inferred probabilities of commitment state membership for the next two periods; the model detects that both customers have decreased their commitment.

Finally, it is worth noting that although customers B and C end up in the same "place" at the time of renewal (i.e., the lowest commitment level), they got there via two different paths. This phenomenon opens an interesting question: Are there particular paths that customers go through before cancelling their subscription?

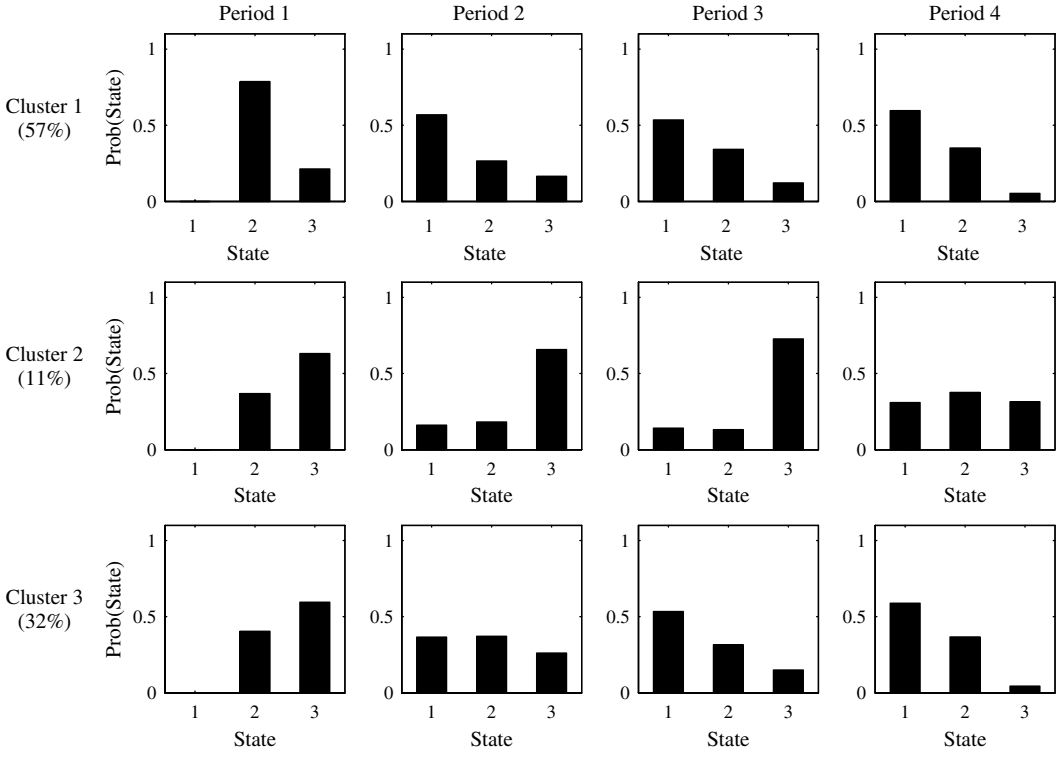
5.2. Identifying "Paths to Death"

We can extend this analysis by calculating the evolution of commitment for all the customers in our sample and then using that information to identify similar "paths to death" (i.e., the most common commitment paths that customers go through before canceling their contracts). Knowledge of these different paths should be of interest to those developing customer retention programs.

For all customers who cancel their membership after one year, we compute the posterior probabilities of belonging to each latent state in each period. These probabilities are computed based on the individual posterior draws of state membership in each period.[24]

[24] We use probabilities, as opposed to posterior states, because their use allows us to account for uncertainty around the posterior estimate of state membership.

Figure 8 Evolution of Centroid Probabilities Before Cancellation



We then perform a *k*-means cluster analysis to identify groups of customers with similar commitment evolution patterns (i.e., customers are grouped based on their probability of belonging to each commitment level in *all* periods before cancellation). Varying the number of clusters from two to four, we find that a three-cluster solution best represents the data. The cluster centroids are plotted in Figure 8.

Cluster 1, representing 57% of the sample, corresponds to those customers whose commitment was very low since the start of their membership. Except for the first period, these customers are always very likely to belong to the lowest commitment state. In contrast, cluster 2's commitment is at its highest level from the moment of acquisition until period 4, the moment at which it decreases to the medium level. (Note that this cluster is much smaller than the first one, representing 11% of the sample.) Finally, cluster 3 represents the remaining 32% of the sample. Customers in this cluster were highly committed when they took out their membership, but then their commitment decreased monotonically over time.

Another way of looking at these data is to assign each cluster to the state with the highest posterior probability of membership on a period-by-period basis. The associated evolution of states is plotted in Figure 9. Three different paths to death emerge from our data. We call cluster 1 the "walking dead" (Schweidel et al. 2008a). Cluster 2 ("sudden death") is a small group of customers who were highly committed for most of the year. Cluster 3 corresponds to those customers who started high but whose commitment to the organization decreased as time went by ("slow death").

5.3. Dynamic Segmentation

We can also group customers on the basis of their commitment level at each point in time, resulting in a dynamic segmentation scheme that could help the marketer better understand her customer base. Figure 10(a) shows how the segment sizes evolve over

Figure 9 Evolution of Commitment States Before Cancellation

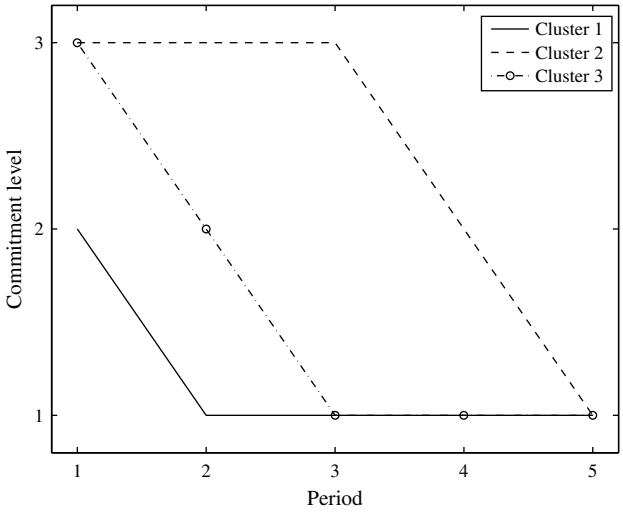
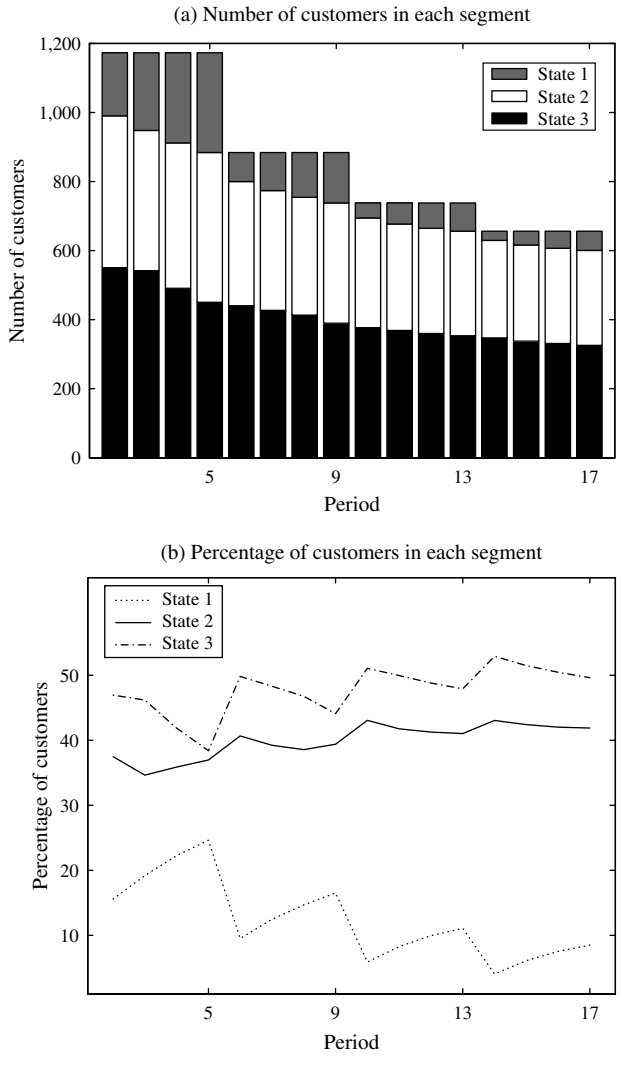


Figure 10 Examining the Dynamics of Segment Sizes



time. (Note that the numbers for periods 12–17 are forecasts.) We observe that the size of the state 1 segment (top gray) increases over time and then radically drops after periods 5, 9, and 13. This is due to the churn process; based on our model assumptions, all customers in state 1 in the renewal period do not renew their membership. Consequently, the total height of the bars decreases after the renewal periods.

A different way of looking at the same pattern is to plot the percentage of customers belonging to each segment over time (see Figure 10(b)). We observe how the share of customers in state 1 increases within each year and then drops right after each renewal opportunity. Moreover, we observe how the overall share of low-commitment customers decreases from year 1 to year 2, from year 2 to 3, etc., whereas the share of states 2 and 3 increases over time. This is an illustration of how the model captures the phenomenon of increasing retention rates observed in most contractual settings when analyzing cohorts of customers.

How would a company benefit from such segmentation scheme? Segmenting customers on the basis of their underlying commitment enables us to not only detect at-risk customers (i.e., potential churners) but also identify highly committed customers. This is in the marketer's interests if, for instance, commitment is correlated with the purchasing or use of additional products and services.

To illustrate this point, we collected additional data to show how the level of underlying commitment (inferred from changes in renewal and usage behavior) could relate to other types of behavior relevant in our empirical setting. In addition to its core offering, the organization under study also runs various educational and special events, which are independent of the set of products and services offered for profit (i.e., the ones used in our empirical analysis).[25] (On average, we would expect the attendance of such events to reflect a customer's commitment.) We obtained information on event attendance for 2004 and extracted the records for those members belonging to the cohort analyzed in this paper. The rate of attendance of these events is low compared to the usage rates; on average, a member attends 0.41 special events a year (0.10 per transaction period), whereas the average number of transactions is 3.8.

Given that period 12 corresponds to the end of year 2004, we select the model predictions about state membership in period 12, as presented in Figure 10. We report in Table 12 the average number of events attended by the individuals assigned to each resulting commitment segment. We observe that the average number of special events attended is higher for the members of the higher commitment segments.[26]

An alternative explanation of this result could be that the attendance of these events simply reflects usage. In turn, given that in our model specification, high levels of commitment are positively associated with high usage levels, finding a monotone relationship between underlying commitment and the attendance of other events might simply be a reflection of individual usage heterogeneity. To examine whether this is the case, we perform a similar analysis in which we segment the sample on the basis of usage behavior alone. We split all active customers into three (similarly sized) groups, depending on the number of transactions up to and including 2004.[27]

[25] These events are offered to all members, without special offers or targeted strategies.

[26] We also examine the percentage of members that attended at least one event and find that this is higher for higher-commitment segments.

[27] The analysis is repeated using historical data from 2004 only, and we also consider usage-based segments with sizes similar to those obtained with the model-based segmentation. The results are robust to these changes.

Table 12 Attendance of Other Events in 2004 by Period 12 Commitment Segment

Commitment segment	No. of customers	Avg. no. of events
1	62	0.53
2	300	0.57
3	376	0.76

Table 13 Attendance of Other Events in 2004 by Usage Segment (Segment 1 = Lowest Usage, Segment 3 = Highest Usage)

Usage segment	No. of customers	Avg. no. of events
1	273	0.71
2	226	0.46
3	239	0.79

Table 13 summarizes the attendance of these other events by each segment. In contrast to the results obtained with the model-based segmentation, we do not find a monotonic relationship between the usage segments and the number of other events attended.

To summarize, we have shown that there is a positive relationship between commitment (inferred from changes in renewal and usage behavior) and the propensity to attend other events offered by the organization. We have also shown that this relationship cannot be explained by usage behavior alone. Therefore, the segmentation scheme suggested by our proposed model seems to discriminate these other behaviors more efficiently than a segmentation scheme based on usage behavior alone.

6. Discussion

In this paper we propose a model of usage and churn behaviors in which both behaviors reflect a dynamic latent variable. The model not only provides accurate multiperiod forecasts of both behaviors, a key input into any serious effort to quantity customer equity, but also offers several insights for managers operating in contractual business settings.

At the heart of our model is a hidden Markov process that characterizes the dynamics of a latent variable, which we label commitment. Churn is deterministically linked to this latent variable and usage is modeled as a state-dependent Poisson process that incorporates time-invariant cross-sectional heterogeneity. The model is flexible enough to be applied in situations where these two processes occur on different time scales, as is the case for most contractual businesses. We validate the model using data from an organization for which an annual membership is required to gain the right to buy its products and services.

Given the task of making multiperiod forecasts of customer behavior, the proposed model outperforms a set of benchmark models, providing more accurate predictions of both churn and usage behaviors. Such predictions lie at the heart of any attempt to quantify customer equity, a concept central to any firm's efforts to become more customer-centric. Forecasts of usage can have additional value, such as in settings where usage levels affect service quality, which in turn affects customer retention and usage (e.g., gym memberships, DVD rental services). It is important to note that the model only requires information readily available in the firms' database, which certainly facilitates its use among practitioners.

Our analysis did not include the effects of marketing actions because no such variables were available in our data set.[28] (Section 3.3 explains how measures of these actions, when present, could be incorporated in our model.) We do not feel that this takes anything away from our research; even though it effectively assumes a status quo behavior on the part of the firm, our proposed model generates good multiperiod forecasts of usage and renewal.

Looking beyond the value of these forecasts, the model also allows us to generate additional insights into customer behavior that are managerially useful. For example, we show how our model can be used to segment the customer base. Moreover, the longitudinal aspect of the hidden Markov model makes it possible to identify the commitment patterns that customers go through over the course of their relationship. As a consequence, we can detect the most common paths to death. For example, in our empirical investigation, we find that almost 60% of those customers that churn at the end of year 1 exhibit low levels of commitment a long time before the contract reaches its expiration date. In other words, the majority of customers were "dead" several periods before they actually cancelled their subscription.

The model presented here is suitable to bigger data sets than the one used in our empirical analysis, but we acknowledge that in some cases, the customer database might be extremely large, presenting an issue of scalability. One approach to dealing with such a situation would be to apply the model to a random sample of customers and then use the joint posterior distribution of the (hyper)parameters to make predictions about the remaining customers. One could obtain individual-level predictions by combining the population priors (obtained from the model) data on the behavior of the remaining customers using Bayes' rule.

[28] As such, it has more in common with the customer lifetime value-related research that emphasizes forecast accuracy (e.g., Fader et al. 2005, 2010) than the work that focuses on resource allocation (e.g., Venkatesan and Kumar 2004, Kumar et al. 2008).

The model makes several assumptions that could be viewed as limitations in some empirical settings. For example, our framework assumes that the renewal decision relies entirely on the current state of commitment. It could be made a function of commitment in both the current and previous periods. Alternatively, one could explicitly relate the renewal decision to customers' expectations about future usage. This approach would require the model to incorporate a forward-looking decision at each renewal opportunity. It is not obvious, a priori, whether and how modeling that behavior would improve the performance of the current model. However, we think that this is an interesting avenue for future research. Second, we have assumed that the number of latent states is common across all customers. A natural extension would be to allow for heterogeneity in K. Third, the way the two observed behaviors are related to the dynamic latent variable model imposes a "positive" relationship between usage and renewal. Although this relationship is very much consistent with previous research (e.g., Bolton and Lemon 1999, Reinartz and Kumar 2003), we acknowledge that there could be patterns of behavior not captured by the proposed model. One such example would be where a customer starts using the service very intensively over the remaining contract period, having decided not to renew her contract when it comes up for renewal. (We did not observe any such behavior in our empirical setting but acknowledge that there could be other settings where it might be present.) Accommodating such behavior would require modifications to our proposed modeling framework. Finally, we have not formally defined or measured the latent variable that drives usage and renewal (even though we have called it commitment). Although the goal of this work is to provide a tool to predict usage and renewal, it would be useful for the marketer to determine what this latent variable actually represents and also investigate what causes it to change over time. To address this issue, customers' attitudes could be measured periodically and linked to the latent variable (using a factor-analytic measurement model). We hope that this research opens up new avenues for understanding the dynamics of customer behavior in contractual settings.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2013.0786.

Acknowledgments

The authors thank Asim Ansari, Kamel Jedidi, Oded Netzer, Catarina Sismeiro, Naufel Vilcassim, and the editor and reviewing team for their insightful comments.

References

Berry MJA, Linoff GS (2004) *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (Wiley Publishing, Indianapolis).

Bhattacharya CB (1998) When customers are members: Customer retention in paid membership context. *J. Marketing* 26(1):31–44.

Blattberg RC, Getz G, Thomas JS (2001) *Customer Equity: Building and Managing Relationships as Valuable Assets* (Harvard Business School Press, Boston).

Blattberg RC, Kim B-D, Neslin SA (2008) *Database Marketing. Analyzing and Managing Customers* (Springer, New York).

Bolton RN (1998) A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* 17(1):45–65.

Bolton RN, Lemon KN (1999) A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *J. Marketing Res.* 36(2):171–186.

Bonfrer A, Knox G, Eliashberg J, Chiang J (2010) A first-passage time model for predicting inactivity in a contractual setting. Working paper, Australian National University, Canberra, ACT. http://ssrn.com/abstract=997810.

Borle S, Singh S S, Jain D C (2008) Customer lifetime value measurement. *Management Sci.* 54(1):100–112.

Chintagunta PK (1993) Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* 12(Spring):184–204.

Cook RJ, Lawless JF (2007) *The Statistical Analysis of Recurrent Events* (Springer, New York).

Danaher PJ (2002) Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Sci.* 21(2):119–138.

Diggle P, Kenward MG (1994) Informative drop-out in longitudinal data analysis. *Appl. Statist.* 43(1):49–93.

Essegaier S, Gupta S, Zhang ZJ (2002) Pricing access services. *Marketing Sci.* 21(2):139–159.

Fader P (2012) *Customer Centricity*, 2nd ed. (Wharton Digital Press, Philadelphia).

Fader PS, Hardie BGS (2010) Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Sci.* 29(1):85–93.

Fader PS, Hardie BGS, Huang C-Y (2004) A dynamic changepoint model for new product sales forecasting. *Marketing Sci.* 23(1):50–65.

Fader PS, Hardie BGS, Lee KL (2005) RFM and CLV: Using isovalue curves for customer base analysis. *J. Marketing Res.* 42(November):415–430.

Fader PS, Hardie BGS, Shang J (2010) Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 29(6):1086–1108.

Garbarino E, Johnson MS (1999) The different roles of satisfaction, trust, and commitment in customer relationships. *J. Marketing* 63(2):70–87.

Gruen TW, Summers JO, Acito F (2000) Relationship marketing activities, commitment, and membership behaviors in professional associations. *J. Marketing* 64(3):34–49.

Hanemann WM (1984) Discrete/continuous models of consumer demand. *Econometrica* 52(3):541–561.

Hashemi R, Janqmin-Dagga H, Commenges D (2003) A latent process model for joint modeling of events and marker. *Lifetime Data Anal.* 9(4):331–343.

Hausman J, Wise DA (1979) Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica* 47(2):455–473.

Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4):465–480.

Krishnamurthi L, Raj SP (1988) A model of brand choice and purchase quantity price sensitivities. *Marketing Sci.* 7(1): 1–20.

Kumar V, Venkatesan R, Bohling T, Beckmann D (2008) The power of CLV: Managing customer lifetime value at IBM. *Marketing Sci.* 27(4):585–599.

Larivière B, Van den Poel D (2005) Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems Appl.* 29(2):472–484.

Lemmens A, Croux C (2006) Bagging and boosting classification trees to predict churn. *Marketing Res.* 43(2):276–286.

Lemon KN, White TB, Winer RS (2002) Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *J. Marketing* 66(1):1–14.

Lu J (2002) Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS®. *SAS User Group Internat. (SUGI27) Online Proc.* (SAS Institute, Cary, NC), Paper 114.

Moe WW, Fader PS (2004a) Capturing evolving visit behavior in clickstream data. *J. Interactive Marketing* 18(1):5–19.

Moe WW, Fader PS (2004b) Dynamic conversion behavior at e-commerce sites. *Management Sci.* 50(3):326–335.

Montoya R, Netzer O, Jedidi K (2010) A dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Sci.* 29(5):909–924.

Morgan RYH, Hunt SD (1994) The commitment-trust theory of relationship marketing. *J. Marketing* 58(3):20–38.

Mozer MC, Wolniewicz R, Grimes DB, Johnson E, Kaushansky H (2000) Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Trans. Neural Networks* 11(3):690–696.

Naik PA, Mantrala MK, Sawyer AG (1998) Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* 17(3):214–235.

Narayanan S, Chintagunta PK, Miravete EJ (2007) The role of self-selection and usage uncertainty in the demand for local telephone service. *Quant. Marketing Econom.* 5(1):1–34.

Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2):185–204.

Parr Rud O (2001) *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management* (John Wiley & Sons, New York).

Reinartz W, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1):77–99.

Risselada H, Verhoef PC, Bijmolt THA (2010) Staying power of churn prediction models. *J. Interactive Marketing* 24(3):198–208.

Rust RT, Zahorik AJ (1993) Customer satisfaction, customer retention, and market share. *J. Retailing* 69(2):193–215.

Rust RT, Zeithaml VA, Lemon KN (2001) *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy* (Free Press, New York).

Sabavala DJ, Morrison DG (1981) A nonstationary model of binary choice applied to media exposure. *Management Sci.* 27(6): 637–657.

Schweidel DA, Bradlow ET, Fader PS (2008a) Modeling the evolution of customers' service portfolios. Working paper, Emory University, Atlanta. http://ssrn.com/abstract=985639.

Schweidel DA, Fader PS, Bradlow ET (2008b) Understanding service retention within and across cohorts using limited information. *J. Marketing* 72(1):82–94.

Schweidel DA, Bradlow ET, Fader PS (2011) Portfolio dynamics for customers of a multi-service provider. *Management Sci.* 57(3):471–486.

Scott SL, James GM, Sugar CA (2005) Hidden Markov models for longitudinal comparisons. *J. Amer. Statist. Assoc.* 100(470): 359–369.

Sriram S, Chintagunta PK, Manchanda P (2012) The effects of service quality on usage and termination of a video on demand service. Working paper, University of Michigan, Ann Arbor.

Venkatesan R, Kumar V (2004) A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing* 68(4):106–125.

Verhoef PC (2003) Understanding the effect of customer relationship management efforts on customer retention and customer share development. *J. Marketing* 67(4):30–45.

Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

Wübben M, Wangenheim F (2008) Instant customer base analysis: Managerial heuristics often get it right. *J. Marketing* 72(3):82–93.

Xie JX, Song M, Sirbu M, Wang Q (1997) Kalman filter estimation of new product diffusion models. *J. Marketing Res.* 34(3):378–393.

Xu J, Zeger SL (2001) Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl. Statist.* 50(3):375–387.