A JOINT MODEL OF USAGE AND CHURN IN CONTRACTUAL SETTINGS

Eva Ascarza
Bruce G. S. Hardie[†]

July 2011

[†] Eva Ascarza is an Assistant Professor of Marketing, Columbia Business School (email: `ascarza@columbia.edu`; web: `www.evaascarza.com`). Bruce G. S. Hardie is Professor of Marketing, London Business School (email: `bhardie@london.edu`; web: `www.brucehardie.com`). An earlier version of this paper was titled "Modeling Churn and Usage Behavior in Contractual Settings." The second author acknowledges the support of the London Business School Centre for Marketing.

Abstract

A Joint Model of Usage and Churn in Contractual Settings

The ability to retain existing customers is a major concern for most businesses. However retention is not the only quantity of interest; there are other behaviors that influence the value of a customer. In most contractual situations, the exact revenue that will be generated by each customer is uncertain at the beginning of the contract period; rather, customer revenue is determined by how much of the associated products or services each individual uses. While a number of researchers have explored the problem of modeling churn in a contractual setting, there is surprisingly limited research on the modeling of usage while under contract.

We propose a dynamic latent variable model in which usage and renewal are modeled simultaneously by assuming that both behaviors are driven by 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 parameters are estimated using hierarchical Bayesian methods.

The proposed model is validated using data from an organization in which an annual subscription/membership is required to gain the right to buy its products or services. We compare the model's forecasts with those of existing methods widely used in practice and show how the proposed model outperforms these benchmark models on several important dimensions. We also demonstrate how the model provides additional insights into the behavior of the customer base that are of interest to managers.

Keywords: Contractual settings, access services, hidden Markov models, RFM, Bayesian estimation, latent variable models.

1 Introduction

The ability to retain existing customers is a major concern for most businesses, especially in mature industries where customer acquisition is very costly and the competitive environment is rather severe (Blattberg et al. 2001, Rust et al. 2001). However retention is not the only dimension of interest in the customer relationship; there are other behaviors that also influence the value of a customer. This is definitely the case in contractual settings where we observe customer *usage* while "under contract." Examples include so-called access services (Essegaier et al. 2002) such as mobile phone services (where we observe the number of calls made, the number of texts sent, etc.), gym memberships (where we observe visits to the gym, the number of extra classes attended, etc.), and "friends" schemes for performing arts organizations (where we observe the number of performances attended, the number of educational events attended, etc.). In such business settings, predictions of future usage are an important input into any analysis of customer value.[1]

Whereas the problem of predicting retention has received much attention in the literature, there is limited research on the modeling of usage in contractual settings (Blattberg et al. 2008). Moreover, none of the existing methods attempt to predict these two behaviors together, thus failing to fully capture the interdependencies between these two processes.

One characteristic of access-type contractual settings is that usage and churn/renewal are, by definition, interconnected processes. Customers need to renew their contracts/memberships/subscriptions in order to have access to the required service. Furthermore, in most such settings, usage and renewal do not occur in the same time frame. For example, let us consider a gym offering monthly memberships and summarizing attendance on a weekly basis, as illustrated in Figure 1. While usage is observed on a weekly basis, renewal only happens at the end of each month (i.e., every four periods). As a consequence, modeling these two processes independently, or alternatively ignoring intra-month usage, would be a waste of useful information that can enrich the model predictions. The complication is that these two processes are operating on two different time "clocks," the unit of time for renewal being a multiple of the unit of time for usage.

---

[1] We acknowledge that in some situations customer revenue is known in advance and is thus independent of usage. This is the case with flat-fee or "all inclusive" contracts, where a customer's revenue is fixed. Even in such cases, customer usage may be of interest to the firm because it affects service quality. For example, consider a broadband provider offering flat-fee contracts: if the company does not manage to predict usage accurately, it could face capacity problems when many customers connect at the same time, thus reducing the quality of their connections.

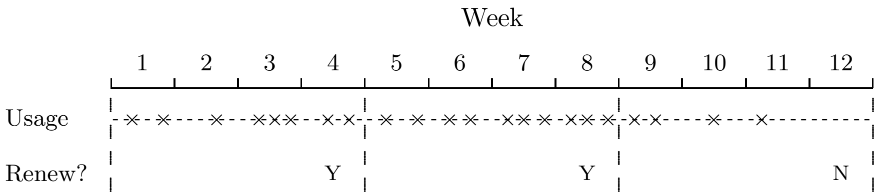


Figure 1: Illustration of usage and renewal behavior occurring on different "clocks."

The objective of this paper is to develop a combined model of usage and renewal that allows these two behaviors to occur on different "clocks." We propose a dynamic latent variable model in which usage and renewal are driven by the same (individual-level) underlying process. This unified framework allows us to capture the interdependencies between usage and renewal, hence providing more accurate predictions for both behaviors. The proposed model only requires information that can be extracted easily from the firm's database, which facilitates its use by both academics and practitioners.

The proposed model is validated using data from an organization in which an annual subscription/membership is required to gain the right to buy its products or services. We compare the model's forecasts with those of existing methods widely used in practice and show how the proposed model outperforms these benchmark models on several important dimensions.

Besides its methodological contribution, the model also provides some useful managerial insights. First, the retention and usage forecasts are vital inputs to any analysis of customer profitability. Hence, accurate forecasts of such behaviors are of interest to any manager working in contractual settings. Second, the proposed model provides information that could be used to target marketing efforts individually and dynamically. Third, because heterogeneity in usage behavior is accommodated, the model allows us to detect heavy users who might be at risk of churning, as well as light users whose expected residual value may be relatively high.

In the next section we introduce the proposed modeling framework and review the relevant literature. Section 3 formalizes the assumptions and builds the integrated model for usage and renewal behavior. The empirical analysis is presented in Section 4. We conclude with a summary of the methodological and practical contributions of this research, as well as a discussion of directions for future research (Section 5).

2 Proposed Modeling Framework

Our objective is to develop a joint model of usage and renewal that can be used to forecast these two behaviors in contractual settings. Such a model must accommodate several aspects that are common across these settings. First, one of the variables of interest is binary (renewal is always a "yes" or "no" decision) whereas the other is not (e.g., number of transactions is a count variable). Second, the renewal process is absorbing. That is, once a customer churns she cannot use the service in any future period (unless she takes out a new contract and is therefore treated as a new customer). Third, the model should allow the usage and renewal processes to occur on different time scales. And finally, the model must require only information that can be extracted easily from the firm's database. (This last point is not unique to contractual settings but is a realistic self-imposed requirement in order to make it easier for the model to be used in practice.)

At first glance, it appears that a discrete/continuous model of consumer demand (e.g., Chintagunta 1993, Hanemann 1984, Krishnamurthi and Raj 1988) 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. While they have been extended to accommodate dropout (e.g., Narayanan et al. 2007), we cannot make use of these models to address our research objective since they do not accommodate the two different time scales. Moreover, the assumptions of these descriptive models (e.g., based on a utility maximizing framework with stable preferences) means that they are more suited to answering "what if" questions rather than forecasting future behavior, which is our objective.

Another possible modeling solution is the set of methods used to model longitudinal panel with dropout (e.g., Diggle and Kenwark 1994). Danaher (2002) uses this approach to investigate price sensitivity in telecommunication markets, extending Hausman and Wise's (1979) sample selection model to correct for attrition bias. These models focus on controlling for dropout-induced bias, rather than predicting dropout, and do not naturally lend themselves to the two "clock" nature of the problem we are addressing. A number of biostatisticians have developed methodologies to model the evolution of a disease marker and the onset of a disease (e.g., Hashemi et al. 2003, Henderson et al. 2000, Xu and Zeger 2001) and it is tempting to draw parallels with the phenomenon we wish to model. Once again, the two "clock" nature of the problem we are addressing means these models cannot be used without radical modifications.

To motivate our proposed approach to the modeling of these behaviors, we start by reviewing the relevant literature on modeling renewal or usage alone. Regression-type models (e.g., logistic regression) and classification-tree methods are widely used by those interested in predicting customer churn. The binary dependent variable is renewal or churn, and other customer characteristics (e.g., past behavior, demographics, attitudes) are used as explanatory variables. This approach is used in financial services and telecommunications settings, to name but a few (e.g., Bolton et al. 2000, Larivière and Van den Poel 2005, Lemon et al. 2002, Mozer et al. 2000, Neslin et al. 2006, Parr-Rud 2001, Rust and Zahorik 1993, Verhoef 2003), where it has been shown that a high level of usage positively affects retention. Blattberg et al. (2008) acknowledge the usefulness of behavior-based variables (such as recency, frequency and monetary value) when modeling customer retention. Likewise, Berry and Linoff (2004) identify declining usage as a significant predictor of churn: customers who use the service less and less over time are more likely to leave it than customers whose usage is stable.

Little attention has been paid to modeling usage in contractual settings. Bolton and Lemon (1999) use a Tobit model to estimate usage of television entertainment and cellular phone services. In a similar manner, Bolton et al. (2000) measure the effect of a loyalty program on future usage of a credit card using a Tobit model.

Modeling either renewal or future usage as a function of past usage raises a fundamental problem in many forecasting settings. Such a modeling exercise sees the transaction database split into two consecutive periods, with data from the second period used to create the dependent variable of the model (e.g., renew (Y/N) when modeling churn, number of transactions or total spend when modeling usage), while data from the first period are used to create the predictor variables. In many settings, period 1 behavior is summarized in terms of each customers "RFM" characteristics: *recency* (time of most recent purchase), *frequency* (number of past purchases), and *monetary value* (average purchase amount per transaction). Having calibrated the regression-type model, we can predict period 3 behavior using the observed period 2 data as the predictor variables. However, it is difficult to use these models to forecast buyer behavior for period 4 as we are unable to specify the values of the RFM predictor variables in period 3 for each customer. (See Fader et al. (2005)

for a discussion of this and related problems with such a modeling approach.)

We overcome this limitation by proposing a joint model of usage and renewal in which both behaviors are driven by a common latent variable (Figure 2) that evolves over time in a stochastic manner.[2] We use data on past usage and renewal behavior to learn about the latent variable's pattern of evolution. Given an expectation of this latent variable for future periods, we can easily predict usage and renewal behavior for those periods.

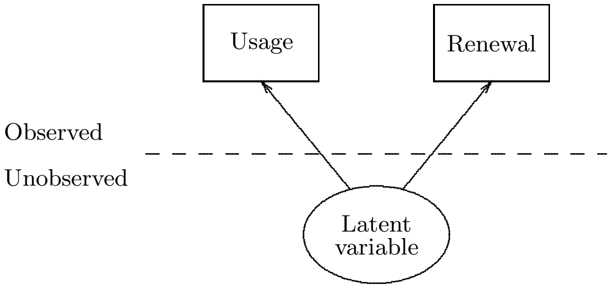


Figure 2: Conceptual framework underpinning the proposed model.

The idea of usage and renewal being driven by a common latent variable is supported by survey-based research in the marketing literature that has shown that contract renewal and service usage are driven by some underlying attitudinal construct. For example, Rust and Zahorik (1993) propose a dynamic framework for 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 in a telecommunications settings. Bolton and Lemon (1999) find a significant relationship between satisfaction and usage when investigating the usage of television entertainment and cellular communications services. 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 and seek to model it and its effects on manifest variables; however we do not to formally define the variable, a position consistent with many models from the statistics, biostatistics, econometric, and psychometric literatures.[3]

---

[2] Note that such a framework in which the two observed variables are driven by a common latent variable suggests that the observed correlation between usage and renewal, which lies at the heart of most churn models, is in fact a spurious correlation.

[3] 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. However we do not use the term "utility" to refer to the latent variable as we are not using a formal utility maximization framework when developing

For the sake of linguistic simplicity — and given the nature of our empirical setting — we will call it "commitment" from now on.[4]

To elaborate on the assumption that each customer's usage and renewal behavior reflect a common latent variable that evolves over time, let us reconsider the gym example with monthly contracts that give members unlimited use of the facilities. Visits to the gym are summarized on a weekly basis, as illustrated in Figure 1. In this example, the latent variable could, for example, represent "commitment to the gym." This commitment evolves over time (Figure 3) and is *reflected* in the number of times a customer goes to the gym each week (usage level). Moreover, her renewal decision also reflects the value of the latent variable; that is, at the end of each month (i.e., every four periods) a customer renews her membership if her commitment in that period is above a certain threshold. Notice that the fact that an individual is active at a particular point in time implies that the value of her latent variable was above some renewal threshold in all preceding *renewal* periods. For example, with reference to Figure 3, the fact that this person is a member in the third month (periods 9–12) means that she renewed in periods 4 and 8. This tells us that value of the latent variable had to be above the renewal threshold in usage periods 4 and 8. However, it does not tell us anything about the level of the latent variable in periods 1, 2, 3, 5, . . . ; this has to be inferred from her usage behavior. We now consider a formal statistical model for this problem.

3 Model Specification

Let t denote the usage time unit (periods) and i denote each customer ($i = 1, \ldots, I$). For each customer i we have a total of T_i usage observations. Let n denote the number of usage periods associated with each contract period (e.g., if the usage unit of time considered is a quarter and the contract is annual, then $n = 4$).

The model comprises three processes, all occurring at the individual level:

1. the underlying "commitment" process that evolves over time,
2. the usage process that is observed every period, and

---

our model.

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

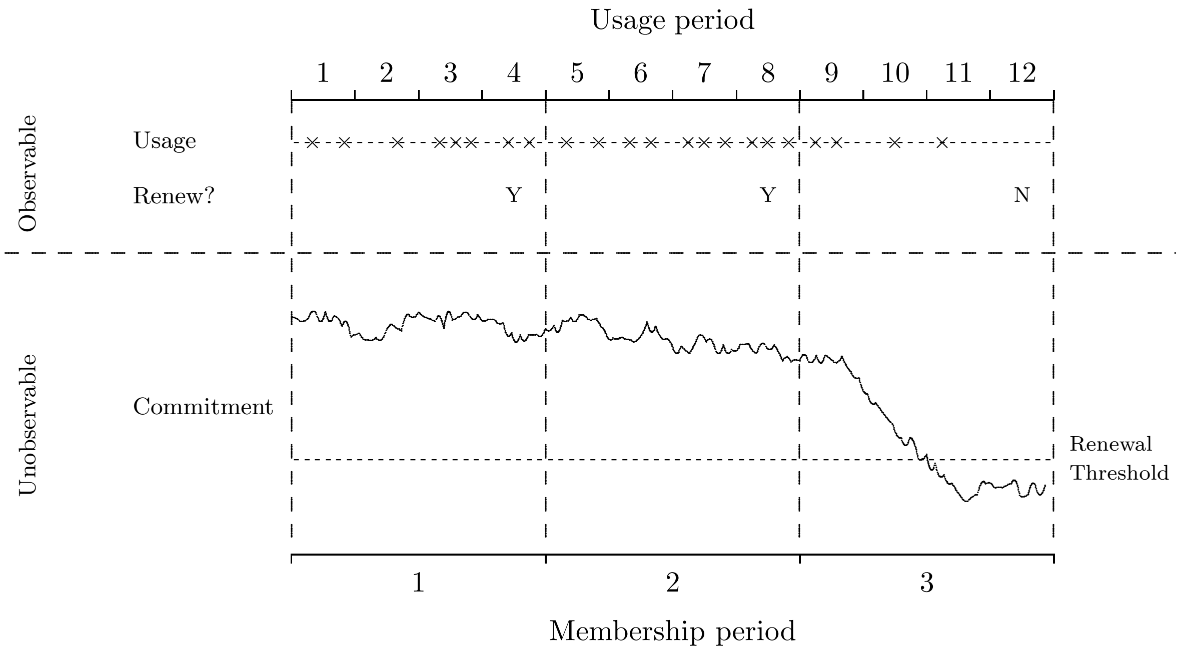


Figure 3: Illustrating the underlying logic of the proposed model.

3. the renewal process that is observed only every n periods and takes the value 1 if a person renews, 0 otherwise.

3.1 The 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. We allow this individual-level latent variable to change over time in a stochastic manner.

In Figure 3, the latent variable is presented as a continuous variable evolving in continuous time. However, we choose to model it as a discrete-time (hidden) Markov process. We 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 individual could occupy at any point in time. 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,

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

We allow individuals to move among the latent states at different rates.[5] We incorporate heterogeneity in dynamics with a Dirichlet mixing distribution for each row in the transition matrix, assuming independence across the rows of the transition matrix $\mathbf{\Pi}_i$:

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

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

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

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 $\boldsymbol{q} = [q_1, q_2, \ldots, q_K]$, where

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

Hidden Markov models (HMMs) were introduced in the marketing literature by Poulsen (1990) as a flexible framework for modeling brand choice behavior. Since then they have been applied in the marketing literature to model a wide range of behaviors (e.g., Montgomery et al. 2004, Montoya et al. 2010, Moon et al. 2007, Netzer et al. 2008, Schweidel et al. 2011, Smith et al. 2006).

Netzer et al. (2008) use a hidden Markov model to capture customer relationship dynamics. The approach taken in the current study is similar to theirs in the sense that we also link transaction behavior to underlying customer relationship strength, in our case labelled "commitment." However, our model specification differs from their approach in several ways. First, they are working in a "noncontractual" setting (where attrition is unobserved) and thus map the latent states to just one observable behavior (i.e., transactions). In our setting customer churn is observed, and therefore this information is used to define and identify the latent states. Second, they assume a transition process where the probability of switching among states is a function of the interactions between

---

[5] This results in 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).

the firm and the customer, and thus changes over time. Such interactions did not occur in our empirical setting during the period under study. As a consequence, we model the transition process in an time-homogeneous manner (i.e., time invariant) while allowing for heterogeneity in transition probabilities across individuals. Third, while Netzer et al. (2008) only allow for transitions between adjacent states, we estimate the full transition matrix. This is of particular interest in the current paper as it allows for a more flexible churn process. Finally, in this paper we allow the link between the underlying state and the observed usage process to be heterogeneous across individuals. This specification allows us to detect heavy users who might belong to low "commitment" states, as well as light-users whose underlying level of "commitment" may be high. The last two points will become clearer once we specify the mapping between the latent variable and the two observable behaviors of interest, usage and renewal.

3.2 The 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). As such, our model should allow two customers with the same underlying pattern of commitment to have different usage patterns.

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

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

In other words, the usage process is determined by a state-dependent parameter θ_k whose value depends on the underlying level of commitment (which varies over time) and an individual-specific parameter β_i that remains constant over time. The parameter β_i captures heterogeneity in usage across the population, allowing two customers with the same commitment levels to have quite different transactions patterns. 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 individual level parameter β_i is assumed to follow a lognormal distribution with parameters $\mu = 0$ and $\sigma = \sigma_\beta$.

The vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]$ of state-specific parameters allows the customer's mean usage levels to change over time as her underlying level of commitment changes. We impose the restriction that $\theta_k > 0 \; \forall \; k$ and is increasing with the level of commitment (i.e., $0 < \theta_1 < \theta_2 < \ldots < \theta_K$). Note that for each individual i, (i) the expected level of usage is increasing with her commitment level, and (ii) we can still observe non-zero usage in the lowest commitment state since $\theta_k \beta_i > 0$.

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 she belongs during her entire lifetime, with realization $\tilde{s}_i = [s_{i1}, s_{i2}, \ldots, s_{iT_i}]$, where s_{it} takes on the value $k = 1, \ldots, K$. The customer's usage likelihood function is

$$\begin{aligned}\mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data}) &= \prod_{t=1}^{T_i} P(Y_{it} = y_{it} | S_{it} = k, \boldsymbol{\theta}, \beta_i) \\ &= \prod_{t=1}^{T_i} \frac{e^{-\beta_i \theta_k} \left(\beta_i \theta_k\right)^{y_{it}}}{y_{it}!}. \qquad (6)\end{aligned}$$

3.3 The 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 for the following n periods based on her current level of commitment. We assume that a customer does not renew if her commitment state is 1 (the lowest commitment level); otherwise she renews. Thus, if a customer is active in a given period t, her commitment state in all preceding *renewal* periods ($n, 2n, \ldots \leq t$) had to be greater than 1; otherwise she would not have renewed her contract and been active at time t. However, an active customer could have been in state 1 in any preceding non-renewal period (i.e., $t \neq n, 2n, \ldots$).

For example, let us consider a gym membership that is renewed monthly and where we observe individual visits on a weekly basis. While usage is observed at every week, renewal/churn can only happen at week 4 (end of first month), week 8 (end of second month), etc. Therefore, the fact that an individual is active in a particular month implies that her commitment level *at the end* of all preceding months (i.e., weeks 4, 8, ...) was greater than 1. Moreover, given that in period 1 all customers have freely decided to take out a service contract, we restrict the commitment state to be different from 1 in the first period (i.e., we restrict $q_1 = 0$ in (4)).

Table 1 shows examples of sequences of commitment states that, based on our assumptions regarding the renewal process, can or cannot occur in our setting:

$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	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	✓

Table 1: Illustrative feasible and infeasible commitment-state sequences.

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 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 month 3 (week 9), her commitment at the end of months 1 and 2 (weeks 4 and 8) had to be greater that 1. However, there are no restrictions about her commitment 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 since renewal behavior is deterministic conditional on the commitment state. However, this hidden state itself evolves in a stochastic and heterogeneous manner. As a result, renewal behavior is modeled probabilistically and this particular model specification allows customers to churn at different rates.

3.4 Bringing It All Together

Now that the renewal process has been specified, we need to combine it with the submodel for usage 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:

$$\mathscr{L}_i(\mathbf{\Pi}_i, \boldsymbol{q}, \boldsymbol{\theta}, \beta_i \,|\, \text{data}) = \sum_{\tilde{s}_i \in \Upsilon} \mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data})\, f(\tilde{s}_i | \mathbf{\Pi}_i, \boldsymbol{q}), \tag{7}$$

where Υ denotes all possible commitment state paths customer i might have during her lifetime, $\mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data})$ is given in (6), and $f(\tilde{s}_i|\boldsymbol{A}, \boldsymbol{q})$ is the probability of path $\tilde{s}_i$.

If there were no restrictions due to the renewal process, the space Υ 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 contains $(K-1)^{\lfloor (T_i-1)/n \rfloor + 1} K^{T_i - \lfloor T_i/n \rfloor - 1}$ possible paths; otherwise Υ 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 random nature of β_i, the overall likelihood function is:

$$\begin{aligned} &\mathscr{L}(\boldsymbol{A}, \boldsymbol{q}, \boldsymbol{\theta}, \sigma_\beta \,|\, \text{data}) \\ &\qquad = \prod_{i=1}^{I} \int_0^\infty \int_{\omega(\boldsymbol{\pi}_{i1})} \cdots \int_{\omega(\boldsymbol{\pi}_{iK})} L_i(\mathbf{\Pi}_i, \boldsymbol{q}, \boldsymbol{\theta}, \beta_i \,|\, \text{data}) f(\mathbf{\Pi}_i \,|\, \boldsymbol{A}) f(\beta_i \,|\, \sigma_\beta) d\mathbf{\Pi}_i d\beta_i \,, \end{aligned} \tag{8}$$

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

To summarize, we have proposed a dynamic latent variable model that uses a hidden Markov structure combined with a heterogeneous Poisson process to model bivariate data where the two processes occur on different time scales. The hidden Markov process captures dynamics at the individual level as well as renewal behavior, while the Poisson process links these underlying dynamics with usage behavior allowing for unobserved cross-sectional heterogeneity. The resulting model has $K^2 + (K-1) + K + 1$ population parameters, which are the elements of $\boldsymbol{A}$, $\boldsymbol{q}$, $\boldsymbol{\theta}$ and σ_β, respectively. We estimate these model parameters using a hierarchical Bayes 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 (due to 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, $\mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data})$. (See Appendix A for further details.)

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 or services, as in the case of some "warehouse clubs" and priority-booking schemes for cultural organizations. The annual membership of this scheme also provides subscribers with several non-pecuniary benefits, including newsletters and invitations to special events.

We focus on the cohort of individuals who took out their initial subscription during the first quarter of 2002, and analyze their buying and renewal behavior for the following four years. Of the 1,173 members of this cohort, 884 renewed in year 2 (26.6% churn), 738 renewed in year 3 (16.5% churn), 634 renewed at least three times (14.1% churn) and 575 were still active after the four years of observation. Expressing these data in terms of periods (as we defined t in section 3), we have a total of 17 periods. We observe usage in periods 1 to 16, and renewal decisions in periods 5, 9, 13, and 17.[6]

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.) For those customers who did not renew their subscription during our observation period, we observe that their usage rate generally decreased before their renewal occasion. For example, for 70% of churners, transaction levels in the last two periods of their relationship with the firm were below their individual averages.

4.2 Model Estimation and Results

We split the four years of data into a calibration period (periods 1 to 11) and a validation period (periods 12 to 17). We can therefore examine model performance in the validation period in three ways. First, we will examine how well the model predicts usage for the remaining period under the

---

[6] In developing the logic of our model, we discussed a contract period of $n = 4$ with renewal occurring at 4, 8, 12, etc. 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) for the case of quarterly periods. In this empirical setting, customers are acquired throughout the first period, which means the first renewal occurs sometime in the fifth period.

current contract (i.e., forecast usage before the membership expires). Second, we will examine the accuracy of the model's predictions of renewal in all future *renewal* periods (i.e., at the end of the current contract and at the end of following contract). Third, we will examine the accuracy of the model's predictions of usage *conditional* on renewal (i.e., forecast usage beyond the first renewal opportunity in the validation period).

We first need to determine the optimal number of states in the hidden Markov chain. We estimate the model varying the number of (hidden) states from 2 to 4, 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 at the individual-period level.

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 to 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 1.06.

# States	Log marginal density	DIC	Individual MSE
2	−16,262	80,546	1.41
3	−14,966	73,517	1.06
4	−15,312	75,065	1.12

Table 2: Choosing the number of states.

Table 3 presents the posterior means and 95% central posterior intervals (CPIs) for the parameters of the usage model under the three-state specification. The first set of parameters (θ_is) correspond to the usage parameters common across all customers in a particular commitment state, while the σ_β parameter 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 4. The average of the state-specific Poisson means is 0.30 for state 1 and 0.32 for state 2. The important difference between these two states is with regards to renewal behavior. The interpretation of each state is determined by the transaction propensity and the renewal behavior. Hence, while those individuals in state 1 will on average make ever-so-slightly fewer transactions than those in state 2, they will

	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]

Table 3: Parameters of the usage model with three states

churn if they remain in that state during a renewal period. On the contrary, individuals in state 2 will renew their membership even though they may also make fewer purchases. For individuals in state 3, the highest commitment level, the average number of transactions per period is 1.80, more than seven transactions in a year.

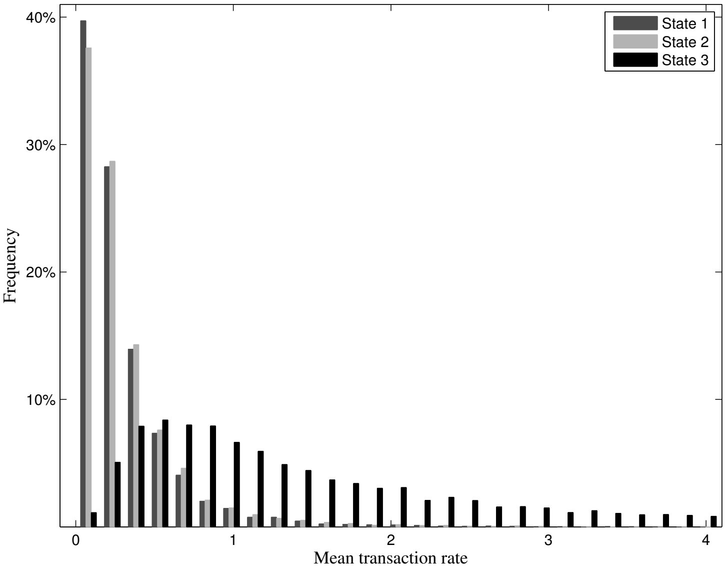


Figure 4: Distributions of the state-specific usage process Poisson means.

Dynamics in the latent variable are captured by the hidden Markov chain. The top part of Table 4 shows the posterior estimate of $\boldsymbol{q}$, 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 joined the membership scheme.

The bottom part of Table 4 shows the posterior estimates of the (Dirichlet) parameters that determine the transition matrix. For an easier interpretation, we report the implied transition

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 46.8]
α_{12}	24.20	[16.39 31.3]
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 4: Parameters of the commitment process with three states.

probabilities. Table 5 shows the population mean and the 95% CPI of the transition probabilities. For example, the third row should be read as follows: for an average individual in state 3, the probability of remaining in state 3 is 0.72, the probability of switching to state 2 in the next period is 0.23, while the 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% CPI, the probability of switching from state 3 to state 2 ranges from 0.06 to 0.34.

	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]

Table 5: Mean transition probabilities, with the 95% CPIs reported in brackets.

Care must be taken when interpreting the state 1 transition probabilities ($\pi_{11} = 0.60$, $\pi_{12} = 0.38$, and $\pi_{13} = 0.02$). At first glance, the fact that $\pi_{11} > 0$ — which implies that s_1 is not an absorbing state (i.e., $P(S_{it} = 1 | S_{i,t-1} = 1) \neq 1$) — appears to be inconsistent with the assumed renewal process (i.e., do not renew if in state 1). What must be noted is that we only assume that customers churn if they are in the lowest commitment state during a renewal period. But not all periods are renewal occasions. Therefore, even though all members in state 1 will not renew if they are in a renewal period, it is possible to find individuals who were in state 1 at a particular time and changed their commitment state before the renewal occasion occurred. In other words, the above state 1 transition probabilities are for non-renewal periods only.

To get a tangible sense of how the model fits the data, we compare the actual and predicted levels of usage for the calibration period. We find that the three-state specification of the proposed model gives an excellent fit when predicting total usage: the MAPE for the total number of transactions per period is 7.44%. But being able to track aggregate levels of usage is not enough; we would expect the model to capture cross-sectional differences. 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 data lie within the 95% CPI of the posterior distributions.

	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%	11.05	11.07	[10.25	12.00]
95%	33.08	33.10	[31.00	35.00]
Max	452.10	453.04	[396.00	513.00]

Table 6: Comparing the actual in-sample distribution of transactions with the model predictions.

4.3 Forecasting Performance

Having fitted the model to the calibration period data, we will examine the performance of the model in the hold-out validation period. In addition to comparing its performance relative to a set of benchmark models, we also consider two restricted versions of the proposed model: a homogeneous usage model, in which all members of each commitment state have the same expected purchase behavior, and a homogeneous transition model, where the transition probabilities from one state to the other are the same for all individuals.[7]

We first forecast usage behavior in period 12 for all members that were active at the end of our calibration period. Then, conditional on each individual's underlying state in period 13, we predict 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. This time-split structure allows us to analyze separately *usage* forecast accuracy (comparing 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).

Usage Process

In order to assess the quality of the usage predictions, we compare the forecasts from our dynamic latent variable model (both the full specification and the two restricted versions) with those generated using two RFM-based random-effects Poisson regression models — see Appendix B for details — and two heuristics based on the work of Wübben and Wangenheim (2008). Heuristic A, which we call "periodic usage," assumes that each individual repeats the same pattern every year. Heuristic B, which we call "status quo," assumes that all customers will make as many transactions as their current average.[8]

To assess the validity of the usage predictions, we compare the models' forecasts in period 12 with the actual data. (Period 12 is the first period of the validation period and the only pre-renewal period.) The predictive performance is compared at the aggregate level, looking at the percentage error in the predicted total number of transactions, at the "distribution" level, looking at the histogram of the number of transactions, and at the individual level, looking at the MSE computed across individuals. For the "distribution" level accuracy, we compute, based on the model predictions, how many customers are expected to have zero transactions, one transaction, two transactions, etc. and compare these values with the actual data. We assess the similarity the

---

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

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

distributions of actual and predicted number of transactions using by χ^2 statistic.[9]

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-level, 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.

	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
Propsed 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: Assessing the period 12 predictive performance of the usage models.

So as to better understand the meaning of a lower χ^2 (i.e., a better disaggregate fit) we compare in Figure 5 the histograms of the number of transactions as predicted by the best four methods on the basis of the disaggregate predictions (heuristic A, cross-sectional RFM, homogeneous transition, and full specification) with the histogram of actual period 12 transactions. The dominance of the full specification is very clear from this plot.

Thus, even though the full model provides a slightly higher % error at the aggregate level than the cross-sectional regression model and one of the constrained specifications, these histograms and the individual-level MSE show that it predicts usage more accurately than any of the other methods.

---

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

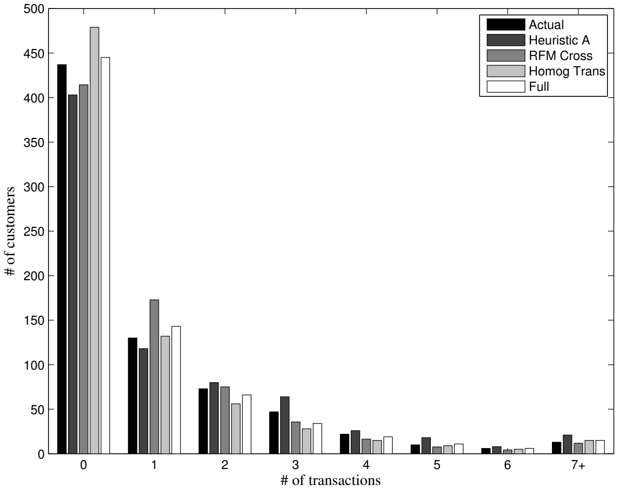


Figure 5: Comparing the predicted and actual distributions of the number of transactions in period 12.

Renewal Process

In order to assess the quality of the churn predictions, we compare the predicted renewal rates of the dynamic latent variable model (both the full and two restricted versions) with those generated using two RFM-based logistic regression models — see Appendix B for details — and two heuristics. The first heuristic (C), which we call "no usage," says that churn occurs if there is no usage activity during the last two periods (Wübben and Wangenheim 2008). The second heuristic (D), which we call "lower usage," says 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. (This is in the spirit of Berry and Linoff's (2004) discussion of how changes in usage can be a leading indicator of churn.)

The predictive performance for all churn models is presented in Table 8. First, we compare actual versus predicted renewal rates in period 13. As shown in the table, the dynamic latent variable models 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 estimates for future churn. The two logistic regression models overestimate future renewal (and therefore the size of the customer base) by more than ten percentage points. 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 since 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 retention rates). As a consequence, the high figures (for hit rate) associated with the RFM models are the consequence of classifying stayers correctly (at the cost of failing to predict churners).

	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	52%
Panel	95%	10.9	83%	68%	−24.7	60%
Proposed model						
Homogeneous usage	87%	1.0	77%	90%	0.4	67%
Homogeneous transition	81%	−6.1	73%	81%	−15.9	60%
Full specification	88%	2.7	78%	91%	3.2	68%
Actual	86%	--	--	91%	--	--

Table 8: Assessing the period 13 and 17 predictive performance of the renewal models.

Second, we compare the accuracy of the methods when predicting renewal behavior at the end of our validation period (i.e., period 17). One problem we encounter with the benchmark methods/models is that none of them can be used, by themselves, to forecast customer renewal behavior in future periods (e.g., period 17) since they are based on measures of usage behavior which are unobserved for future periods. For example, in order to use the logistic regression model to predict period 17 churn, we would need to know how many transactions each customer made up-to-and-including period 16. However, we can use the usage benchmark models recursively to simulate individual transactions, then update the RFM characteristics, and finally predict renewal behavior in period 17.[10] The period 17 predictions for the homogeneous usage specification and the full model specification are exceptionally accurate at the aggregate level (% error= 0.4 and 3.2 respectively), with hit rates of 67% and 69%. We note that the predicted renewal rates associated with the homogeneous transition specification are the same in periods 13 and 17; this is a natural consequence of the model specification.

---

[10] Notice that the accuracy of those estimates will not be a measure of the logistic regression alone, but also of the usage models presented in Table 7.

Given the performance of the homogeneous usage specification, it is tempting to think that the full specification is not needed. However, it must be remembered that renewal is just one of the behaviors of interest, and we recall from Table 7 that this homogeneous specification performs poorly when considering usage in period 12.

Renewal and Usage Processes

Finally, we consider the overall forecasting accuracy of the models by examining usage behavior in periods 14 to 16, which in turn depends on predicted renewal behavior in period 13. For the RFM-based regression models, we use a combination of the churn and usage models to make such predictions. We first predict renewal behavior in period 13 using the RFM logistic regression models (both the cross-sectional and panel specifications). 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.

We look at actual versus predicted usage levels in periods 14–16, examining the accuracy of the predictions at the aggregate, distribution, and individual level (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 second-best model is the homogeneous usage specification, with an MAPE of 4.8%, with the homogeneous transition specification coming a distant third. The combination of the RFM models results in very poor estimates of future behavior at the aggregate level. When we consider the disaggregate (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 prediction accuracy, we see that the full model specification is clearly superior.

4.4 Implications for the Firm

We now move away from the technical features of the model to discuss several aspects in which the proposed model can help managers in contractual businesses make better decisions.

	Aggregate (MAPE)	Disaggregate (Avg. χ^2)	Individual (MSE)
RFM			
Cross-sectional	26.4	291.8	18.5
Panel	68.2	48.2	18.4
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 9: Assessing the accuracy of usage predictions for periods 14–16.

Customer Valuation

We have shown that our dynamic latent variable model accurately predicts usage and renewal behavior, outperforming existing methods on both dimensions. This improvement in accuracy has important implications for firms operating in contractual businesses. First of all, a model that can forecast short and medium-term cash flows accurately is of great use for budget planning, setting acquisition strategies, and so on. In turn, the correctness of any strategic decision that depends on expected cash flows relies entirely on the precision of the model predictions. Similarly, since retention and usage forecasts are vital inputs to any analysis of customer profitability, the validity of any customer valuation exercise relies of the accuracy of the underlying model. As such, inaccurate models of customer behavior should be rejected (or used with extreme caution).

To provide a better sense of how the model accuracy translates into economic terms, we use the validation period to assess the exact monetary error incurred by the methods presented in the previous section. Given the models' predictions of usage and renewal behavior, we translate those behaviors into cash flows (in USD) generated by each customer during the six periods comprising the validation period.[11] We discount these cash flows to the start of the validation period and compare the model-based predictions to the actual numbers.[12]

The revenue generated by these individuals during the 2.5 years of validation was $1.85 million

---

[11] 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.

[12] We apply a discount rate of 5% per quarter, which translates to an annual rate of approximately 19%. We replicate the analysis using (quarterly) discounts rates ranging from 2% to 7% and obtain qualitatively similar results.

(Table 10). Predictions based on the RFM models are inaccurate both at the aggregate and individual level. The homogeneous usage specification and the full model predict \$1.82 and \$1.80 million respectively, which corresponds to percentage errors of only 2% and 3%. While the aggregate predictions are slightly better when using the homogeneous usage specification, individual-level valuations are more accurate under the full model. (These figures are consistent with Tables 7 to 9.)

	Total Revenue (\$000)	Aggregate % Error	Individual MSE (× 1e6)
RFM cross-sectional	2,386	26	216.5
RFM panel	4,548	136	490.0
Homogeneous usage	1,825	−2	6.1
Homogeneous transition	1,631	−12	3.1
Full	1,797	−3	2.5
Actual	1,855		

Table 10: Comparing the present value of actual and predicted validation-period revenue.

Basis for Segmentation

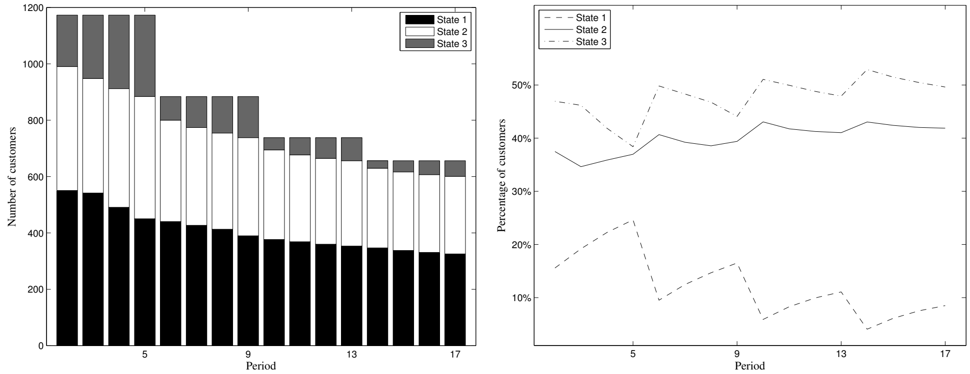
In addition to predicting future usage and renewal behavior, this model also provides several insights that can help the marketer better understand her customer base. In particular, this model allows us to segment customers dynamically on the basis of their underlying commitment levels. As a consequence, the model not only enables us to detect at-risk customers (i.e., potential churners), but also to identify highly committed customers. This is in the marketer's interests if, for instance, the organization wants to target its marketing activities.

The hidden Markov specification allows us to dynamically segment the customer base given usage and renewal behavior.[13] The hidden Markov model can provide insights into the underlying behavioral dynamics of the customer base. To do so, we need to compute the number of customers in each commitment state and track this information over time. Recovering the state membership is straightforward when using data augmentation techniques to estimate the hidden states. In each period we can easily compute the number of customers in each segment and look at how the segment sizes evolve over time. Figure 6a shows the size of each hidden state for all periods; the

---

[13] Hereafter we will use the terms segment and state interchangeably.

numbers for periods 12–17 are forecasts. We observe that the size of state 1 (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 also decreases after the renewal periods.



(a) # customers in each segment (b) % customers in each segment

Figure 6: Examining the dynamics of segment size.

A different way to look at the same pattern is by plotting the percentage of active customers belonging to each segment overtime (Figure 6b). 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., while 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.

Recovering State Membership

The model not only recovers the size of each segment over time, but also allows us to make inferences about the individual dynamics. We can easily analyze the individual-level transitions by recovering the distribution of the state membership for each customer over her lifetime. Looking at the posterior probability of belonging to each state, we can then analyze the commitment dynamics at the individual level. Recovering the underlying states over time would allow the firm to identify those customers who are likely to have changed (decreased or increased) their commitment state recently. This useful piece of information would help the marketer differentially target the customers. For example, in our setting, the organization would be interested in knowing, before the membership expiry 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, we analyze 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). Figure 7 shows how the distribution of state membership varies during the second year of their subscription and Table 11 shows the observed transaction patterns for these three customers during their first two years of membership (i.e., periods 1–8).

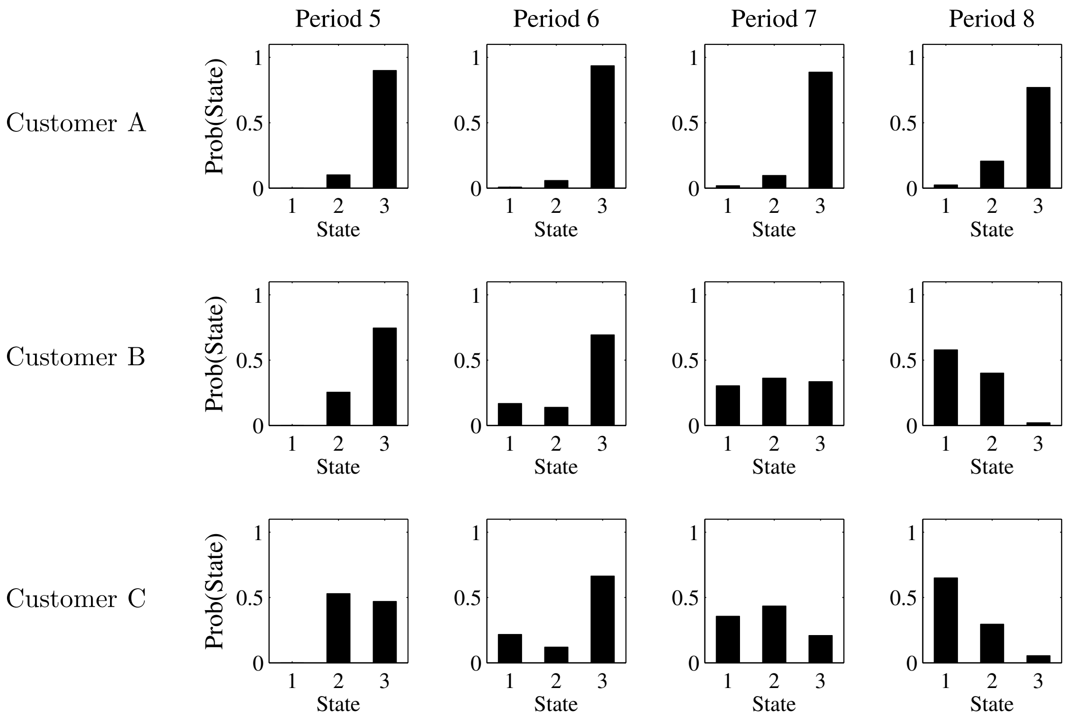


Figure 7: Segment membership dynamics for three customers.

Let us start by looking at customers A and B. While both customers have exactly the same usage behavior during year 2 (Table 11), their inferred commitment patterns are radically different (Figure 7). Customer A is highly committed in all periods, whereas customer B's commitment drops after period 6. (We observe how her probability of belonging to state 1 increases notably from periods 5 to 8.) This is reflected in the differences in their usage and renewal behavior over

	Year 1				Year 2			
	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

Table 11: Actual usage behavior for three customers.

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's A past behavior, one period of no purchases does not necessarily indicate a high risk of churn.

Turning to customers B and C, we note that customer B's purchasing in periods 1–5 is higher than that of customer C. This is reflected in the inferred commitment level for period 5, with customer B having a high probability of being in the highest commitment state and customer C having an almost equal probability of being in states 2 or 3. 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 as two purchases is not out-of-the-ordinary in light of her transactions in year 1. 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.

The focal firm did not perform any targeted marketing activities during the period under study, hence the present analysis only serves as an illustration of how the proposed model works and could feed into marketing decision making. Nevertheless, being able to recover the individual state membership over time is potentially a very useful piece of information for the firm.

4.5 Demonstrating Concurrent Validity

We have developed a joint model for usage and renewal behavior in contractual settings. The model focuses of these two behaviors for two reasons: (i) they are the two key drivers of customer profitability, and (ii) this information is generally available in a company's database. At the heart of the model is a dynamic latent variable that we have called commitment. While we have examined the predictive validity of the model, we have said little about the underlying process that drives both behaviors. In order to provide a comprehensive discussion of what the latent variable actually represents, we would need additional information about members' attitudes, which would be costly to obtain. Alternatively, we could provide some evidence that is consistent with our contention that the latent variable (as modeled in a discretized form by the states of the hidden Markov model) captures some notion of commitment. The basic idea, drawing on the notion of concurrent validity, is to assess the relationship between the latent variable and measures of other behaviors we would expect of "committed" individuals — see Figure 8.

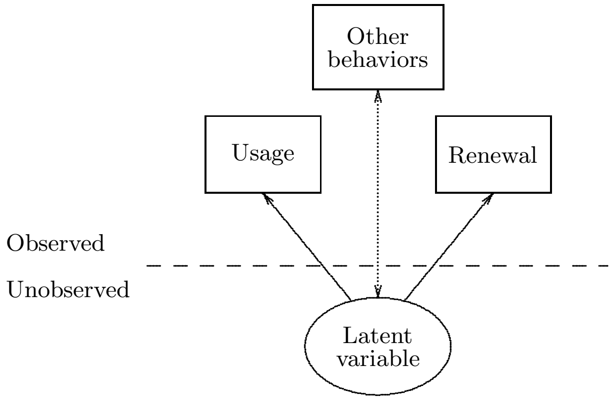


Figure 8: Exploring the meaning the latent variable.

In addition to access privileges, membership of the organization from which we obtained the data gives the option to attend special events, some of which are educational in nature. The rate of attendance of these events is low compared to the transaction rates; on average, a member attends to 0.41 special events a year (0.10 per transaction period), whereas the average number of transactions is 3.8. On average, we would expect that highly committed members are more likely to attend of such events than those whose commitment level is low. Hence, we could use information about attendance of these events as a proxy of commitment and thus to provide some validity to our suggested interpretation of the latent variable.

We obtained information on event attendance for 2004 and extracted the records for those members belonging to the cohort analyzed in this paper.[14] Given that period 12 corresponds to

---

[14] Collecting information on the attendance of these events is not straightforward as they were organized by different

the end of year 2004, we select the model predictions about state membership in period 12 and compute, for each commitment level, the average number of events attended. In other words, we segment the set of active customers into three groups — those who are predicted to belong to state 1, to state 2, and to state 3 — and then summarize attendance of special events for each group (Table 12).

State	# customers	Average # events
1	62	0.53
2	300	0.57
3	376	0.76

Table 12: Information on the attendance of other events by state membership in period 12.

We observe that the average number of special events attended is higher for higher commitment states, as is the percentage of members attending at least one event.[15] We view this as evidence that the latent variable used to segment the customers captures (to some extent) the notion of commitment.

An alternative explanation of this result could be that the attendance of these events does not reflect commitment (or any other label we could assign to the latent variable that drives behavior in a particular setting), but might be simply another measure of usage. In turn, given that in our model specification high levels of commitment are positively related with high transaction levels, finding a monotone relationship between underlying commitment and the attendance of other events might be just a measure of individual usage heterogeneity. To examine whether this is the case, we perform the same analysis while segmenting the sample on the basis of usage behavior. We find that usage behavior itself does not explain differences in attendance of other type of events. (See Appendix C for details of this analysis.)

We recognize that this analysis is not free of limitations. We only consider one year and one transaction period rather than performing a longitudinal analysis. If possible, one should get period-by-period information about the attendance of special events, and match this information

---

parts of the organization and there was no shared database.

[15] We also examine the percentage of members that attended at least one event, and find that this is higher for higher commitment states.

with the state membership estimates provided by the model. Nevertheless, while this analysis could be strengthened with additional data, we have shown that those customers identified by the model as being highly committed attend special events more often than those who are assigned to lower states of commitment.

4.6 Generalizing the Model

There are certain aspects of this model that may appear to have been dictated by the specific empirical setting under study. We now briefly discuss the generalizability of the model.

The proposed model assumes that, conditional on the underlying state, usage behavior follows a Poisson process. 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. For example, going back to the gym example, if one wants to model the number of days a member attends in a particular week, the Poisson may not be the most appropriate distribution since there is an upper bound of seven days. Similarly, consider the case of an orchestra wanting to predict the number of tickets that will be sold to their patrons. First, the number of performances attended is bounded by the total number of performances offered by the orchestra. Second, the number of performances offered should also be taken into consideration when predicting customers' future attendances; there will be periods with higher demand simply because more performances are on offer and so the model should accommodate this information. In these cases the Poisson distribution should be substituted by binomial distribution in which the upper bound (e.g., number of days in a week, total number of performances offered) is the number of trials.[16]

There are also situations where usage is not discrete. These are cases in which usage refers to time (e.g., minutes used in wireless contracts), expenditure (e.g., total amount spent), or other quantities (e.g., MB downloaded in an Internet data plan). The proposed model is easily applied in such settings, provided the distributional assumption of the usage process is modified. We could

---

[16] We note that the binomial distribution can be approximated by a Poisson distribution when there are a large number of trials.

simply replace the Poisson or binomial with, say, a gamma or lognormal distribution.

Furthermore, we may wish to incorporate time-invariant covariates (e.g., customer demographics) and/or time-varying covariates (e.g., marketing activities) into the model. This additional information might be included in different ways. If we expect the covariates to explain cross-sectional variation in mean usage levels or to have an short-term effect on usage behavior (e.g., a promotional campaign lasting for a very short period) we could simply make (5) a function of the covariates using, say, an exponential link function. If, on the contrary, the observed covariates could potentially affect customer "commitment" or, in other words, could have a longer-term effect on customer behavior, then we should incorporate covariates in the transition probabilities. (Technical details about these alternative specifications are provided in Appendix D.)

In a similar manner, one could also include information on competitors' actions (e.g., price reductions, new products introduction) when modeling both transition and usage 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 challenge faced by the analyst is to gain access to the relevant competitive information.

5 Discussion

We have proposed a dynamic latent variable model in which usage and renewal are modeled simultaneously by assuming that both behaviors are driven by the same (individual-level) underlying process that evolves over time. Dynamics in the underlying latent variable (which we label "commitment") are captured using a hidden Markov model.

With regards to predicting customer behavior, the proposed model outperforms a set of benchmark models. It not only gives more accurate predictions in the short-term, but also forecasts both usage and renewal further into the future. Given the usage component of the model, it is especially valuable in those contractual settings where future usage is not known in advance but is of importance, either because it directly affects revenue (e.g., credit cards, wireless contracts, performing arts organizations), or because it affects service quality, which in turn affects customer retention and usage (e.g., gym memberships, DVD rental services).

Besides its methodological contribution, the model can provide useful managerial insights. First, the retention and usage forecasts are vital inputs to any analysis of customer value. Second, the proposed model provides information that can be used by the marketer to target retention efforts individually and dynamically. Third, the model not only detects at-risk customers (potential churners) but also identifies highly committed customers. Fourth, because heterogeneity in usage behavior is allowed within each state, the model allows us to detect heavy users who might be at risk, and also light-users whose underlying level of commitment may be high. (This information cannot be extracted from a model-free segmentation scheme based on observed transaction data alone.)

We recognize that this analysis is not free of limitations. 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 methodological tool to predict usage and renewal, it would be very useful for the marketer to determine what this latent variable actually represents and also investigate what makes it 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). Incorporating this information might be costly. We hope that this research opens new avenues to understand the dynamics of customer behavior in contractual settings.

Appendix A: Model Estimation

The model is estimated using a Bayesian framework. We obtain estimates of all model parameters by drawing from the marginal posterior distributions, and use a data augmentation approach to deal with the latent states S_{it}.

Let $\mathbf{\Omega}$ denote all the model parameters, including the population parameters $\boldsymbol{A}$, $\boldsymbol{\theta}$, $\boldsymbol{q}$, and σ_β, the individual-level parameters $\boldsymbol{\beta} = \{\beta_i\}_{i=1,\ldots,I}$ and $\mathbf{\Pi} = \{\mathbf{\Pi}_i\}_{i=1,\ldots,I}$, and the set of augmented paths of commitment states $\boldsymbol{s} = \{\tilde{s}_i\}_{i=1,\ldots,I}$. The full joint posterior distribution can be written as:

$$f(\mathbf{\Omega}|\text{data}) \propto \left\{ \prod_{i=1}^{I} \mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data}) \right\} f(\boldsymbol{s}|\boldsymbol{q}, \mathbf{\Pi}) f(\mathbf{\Pi}|\boldsymbol{A}) f(\boldsymbol{\beta}|\sigma_\beta) f(\sigma_\beta) f(\boldsymbol{q}) f(\boldsymbol{A}) f(\boldsymbol{\theta})$$

where $f(\boldsymbol{s}|\boldsymbol{q}, \mathbf{\Pi})$ refers to the distribution of the latent states, assumed to follow a hidden Markov process with renewal restrictions in periods $t = 1, n+1, 2n+1, \ldots$ The term $f(\mathbf{\Pi}|\boldsymbol{A})$ corresponds to the prior (or mixing) distribution for the individual transition probabilities. Each row j of the matrix $\mathbf{\Pi}_i$ is assumed to follow a Dirichlet distribution with parameter vector $[\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jK}]$; we let $\boldsymbol{A}$ denotes the matrix whose jth row is the vector $[\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jK}]$. The term $f(\boldsymbol{\beta}|\sigma_\beta)$ denotes the prior (or mixing) distribution for the β_i, which is assumed to follow a lognormal distribution with parameters $\mu = 0$ and $\sigma = \sigma_\beta$.

The terms $f(\sigma_\beta)$, $f(\boldsymbol{q})$, $f(\boldsymbol{A})$, and $f(\boldsymbol{\theta})$ denote the (hyper)priors for the population parameters. Uninformative (vague) priors are used for all parameters. We assume σ_β has an inverse-Wishart prior with parameter $R = 0.05$ and $df_\sigma = 2$. For the vector $\boldsymbol{q}$ we use a Dirichlet prior with parameter vector $\boldsymbol{\lambda}_{\boldsymbol{q}} = [1, 1, \ldots, 1]$.

We reparameterize $\alpha_{jk} = e^{\rho_{jk}} \; \forall j, k \in (1, K)$ and estimate $\boldsymbol{\rho} = [\rho_{11}, \ldots, \rho_{1K}, \ldots, \rho_{K1}, \ldots, \rho_{KK}]$. We need to ensure that $0 < \theta_1 < \theta_2 < \ldots < \theta_K$. We therefore reparameterize $\theta_1 = e^{\gamma_1}$ and $\theta_k = \theta_{k-1} + e^{\gamma_k} \; \forall\, k > 1$ and estimate $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_K]$ instead. We assume $\Phi = \{\boldsymbol{\gamma}, \boldsymbol{\rho}\}$ follows a multivariate normal distribution with parameters $\mu_\Phi = [4 \times \mathbb{1}_{K^2}, 3 \times \mathbb{1}_K]$ and $\text{diag}(\Sigma_\Phi) = [(1/2) \times \mathbb{I}_{K^2 \times K^2}, 1 \times \mathbb{I}_{K^2}]$, where $\mathbb{1}_n$ is a $1 \times n$ vector of ones, and $\mathbb{I}_n$ is the $n \times n$ identity matrix. (The values of μ_Φ and Σ_Φ were chosen to ensure uninformative priors in the transformed space.)

We draw recursively from the following posterior distributions:

- [Gibbs] For the jth row of $\mathbf{\Pi}_i$, $f(\boldsymbol{\pi}_{ij}|\Phi, \boldsymbol{s}) \sim \text{Dirichlet}(\alpha_{j1} + n_{ij1}, \ldots, \alpha_{jK} + n_{ijK})$, where

$n_{ijk} = \sum_{t=1}^{T_i-1}(s_{it} = j \text{ and } s_{it+1} = k).$

- [Metropolis-Hastings] $f(\beta_i|\sigma_\beta, \boldsymbol{s}, \text{data}) \propto \exp\left(\frac{\beta_i^2}{2\sigma_\beta}\right) P(\text{data}|\beta_i, \Phi, \boldsymbol{s}).$
- [Gibbs] $f(\sigma_\beta|\boldsymbol{\beta}, R, df_\sigma) \sim \text{inv-Wishart}(\sum_{i=1}^{I}(\beta_i^2 + (df_\sigma/R), df_\sigma + I).$
- [Gibbs] $f(\boldsymbol{q}|\lambda_q, \boldsymbol{s}) \sim \text{Dirichlet}(1 + n_{01}, \ldots, 1 + n_{0K})$, where $n_{0k} = \sum_{i=1}^{I}(s_{i1} = k).$
- [Metropolis-Hastings] $f(\Phi|\mu_\Phi, \Sigma_\Phi, \boldsymbol{s}, \text{data}) \propto \exp\left(.5(\Phi - \mu_\Phi)'\Sigma_\Phi^{-1}(\Phi - \mu_\Phi)\right) P(\text{data}|\beta_i, \Phi).$
- For each individual i, we draw from the distribution of the hidden states using the direct Gibbs sampler approach proposed by Scott (2002):

  $$P(S_{i1} = k|\boldsymbol{Q}, \dot{s}_{i(1)}, \text{data}) \propto q_k P(S_{i2} = s_{i2}|S_{i1} = k)\mathbf{1}_{\boldsymbol{\Upsilon}}(\tilde{s}_{i(1,k)})$$

  $$P(S_{it} = k|\boldsymbol{\Pi_i}, \dot{s}_{i(t)}, \text{data}) \propto P(S_{it} = k|S_{it-1} = s_{it-1})P(S_{it+1} = s_{it+1}|S_{it} = k)\mathbf{1}_{\boldsymbol{\Upsilon}}(\tilde{s}_{i(t,k)}),$$

  where $\dot{s}_{i(t)} = [s_{i1}, ..., s_{it-1}, s_{it+1}, ..., s_{iT_i}]$, $\tilde{s}_{i(t,k)} = [s_{i1}, ..., s_{it-1}, k, s_{it+1}, ..., s_{iT_i}]$, and $\mathbf{1}_{\boldsymbol{\Upsilon}}(z)$ is the indicator function with value 1 if $z \in \Upsilon$, 0 otherwise.

Since there is no closed-form expression for the posterior distributions of $\boldsymbol{\beta}$ and Φ, we use a Gaussian random-walk Metropolis-Hasting algorithm to draw from these distributions. Following the Metropolis-Hasting procedure proposed by Atchadé (2006), for each iteration, l, we draw a proposal vector of parameters $\eta^{(l)}$ (either for β_i or Φ):

$$\eta^{(l)} \sim \text{Normal}(\eta^{(l-1)}, \sigma^{(l-1)}\Delta^{(l-1)})$$

and then accept the vector using the Metropolis-Hastings acceptance ratio. The tuning parameters $\sigma^{(l)}$ and $\Delta^{(l)}$ are adapted in each iteration to get an acceptance rate of approximately 20%.

We ran the simulation for 500,000 iterations. The first 450,000 iterations were used as a "burn-in" period, and the last 50,000 iterations were used to estimate the conditional posterior distributions.

Appendix B: Estimating the RFM-based Benchmark Models

Within both academic and practitioner circles, there is a tradition of building regression-type models for predicting churn and, to a lesser extent, usage (or related quantities). In this appendix, we describe the specification of the benchmark regression models used in our analyses.

As previously noted, the regressions model the behavior of interest as a function of the customer's past behavior, frequently summarized in terms of her RFM characteristics. We operationalize these RFM characteristics in the following manner. *Recency* is defined as the number of periods since the last usage transaction. *Frequency* is defined as the total number of usage transactions in the previous four periods. We also compute another measure of frequency, *Fsum*, which is the total number of transactions per customer over the entire period of interest. *Monetary value* is the average expenditure per transaction, where the average is computed over the previous four periods. We also compute *Msum*, the customer's total spend over the entire period of interest. (In exploring possible model specifications, we also consider logarithmic transformations of these variables, as well as interactions between the RFM measures.)

Perhaps the most common approach to developing a churn model is to use a cross-sectional logistic regression with the last renewal observation as the dependent variable and RFM measures as covariates. In developing such a benchmark model, we selected the specification that provided the most accurate in-sample hit-rates. The associated parameter estimates are given in Table B1. We note that the recency variable is not a significant predictor by itself, although its interaction with frequency is significant.

	Coef.	Std. Err.
Intercept	0.746	0.294
Recency	0.016	0.076
Fsum	0.058	0.017
Msum	0.002	0.000
Recency × Fsum	−0.071	0.015
Frequency × Monetary value	−0.002	0.000
LL	−327.4	

Table B1: Parameter estimates for the cross-sectional logistic regression model.

Given the nature of the usage data, we used a count model to develop our regression-based benchmark model. We used a Poisson regression model with a normal random effect to account for the observed over-dispersion in our data. We selected those individuals that were still members at the end of our calibration period. We used the number of transactions in the last period (11) as the dependent variable and the RFM measures as predictors. The parameter estimates are presented in Table B2. We note that the frequency variable is not significant, although its interaction with recency is significant and positive. In other words, this model suggests that the extent to which recency is correlated with future purchasing depends on the past purchasing rate of each individual.

	Coef.	95% CPI	
Random effect			
μ	0.003	[−0.562	0.952]
σ^2	0.644	[0.509	0.820]
Recency	−0.404	[−0.470	−0.348]
Frequency	−0.020	[−0.056	0.022]
Monetary value	0.001	[−0.001	0.002]
Recency × Frequency	0.068	[0.028	0.103]
Log marginal density	−783.8		

Table B2: Parameter estimates for the cross-sectional Poisson regression model.

Noting that the longitudinal nature of our dataset gives us multiple observations per individual, and not just the information for the most recent period, we can extend the cross-sectional models and estimate longitudinal models using (when available) more than one observation per customer.

Using these panel data, we estimate a logistic regression model using observed renewal behavior for all the periods, not just the most recent one; this gives us several observations for those customers that have renewed at least once. We allow for unobserved heterogeneity in renewal behavior using a normal random effect. Table B3 shows the parameter estimates for the (longitudinal) random-effects churn model. The sign and magnitude of all covariates are consistent with the results obtained in the cross-sectional specification. (Note that variance of the random effect is not significant.)

Similarly, we estimate a random-effects Poisson regression model using transaction behavior from all preceding periods — see Table B4. The results are consistent with those obtained in the cross-sectional model, with the only exception that the frequency variable now is significant by itself and the interaction of recency with monetary value is significantly positive.

	Coef.	Std. Err.
Random effect		
μ	-0.084	0.176
σ^2	0.000	0.152
Recency	0.091	0.049
Frequency	0.059	0.030
Msum	0.003	0.000
Recency × Frequency	-0.099	0.027
Recency × Monetary	-0.001	0.000
Frequency × Monetary	-0.003	0.000
LL	-775.6	

Table B3: Parameter estimates for the panel logistic regression model.

	Coef.	95% CPI	
Random effect			
μ	-0.085	[-1.250	1.694]
σ^2	0.998	[0.863	1.149]
Recency	-0.211	[-0.233	-0.192]
Frequency	-0.033	[-0.042	-0.025]
Monetary value	-0.004	[-0.006	-0.003]
Recency × Frequency	0.039	[0.030	0.047]
Recency × Monetary value	0.001	[0.000	0.001]
Log marginal density	$-8{,}085.9$		

Table B4: Parameter estimates for the panel Poisson regression model.

Appendix C: Additional Concurrent Validity Analysis

We have shown how our model provides a classification of customers that is positively related to the attendance of other events. To rule out the possibility of such a correlation being explained purely by differences in usage propensity, we perform an additional analysis in which we segment the customer base on the basis of usage alone (instead of (predicted) latent commitment).

We first split all active customers into three (similarly sized) groups, depending on the number of transactions during their lifetime. We then summarize the information about attendance of other events by each segment — see Table C1. In contrast to the results obtained in Section 4.5, we do not find a monotonic relationship between the usage segments and the number of other events attended.

State	# customers	Average # events
1	273	0.71
2	226	0.46
3	239	0.79

Table C1: Attendance of other events.

The analysis is repeated using historical data from 2004 only, and we also consider usage-based segments with sizes similar to those obtained in Section 4.5. All results are robust to these changes.

Appendix D: Alternative Model Specifications

Binomial Specification for the Usage Model

Let m_t denote the number of transaction opportunities (e.g., number of days in a particular period of time, number of performances offered) and p_{it} the probability of a transaction occurring at any given transaction opportunity for customer i in period t. As with the Poisson specification, the transaction probability depends on the individual's specific time-invariant parameter β_i and the commitment state at every period:

$$p_{it} \,|\, [S_{it} = k] = \theta_k^{\beta_i} \,. \tag{D1}$$

This specification also guarantees that the transaction probability is increasing with the level of commitment. The usage propensity parameter β_i is assumed to follow a lognormal distribution with parameters $\mu = 0$ and $\sigma = \sigma_\beta$. The inclusion of β_i as an exponent (as opposed to a multiplier) ensures that the transaction probabilities remain bounded between zero and one. (As this transformation is not linear in β_i, the average probability of transaction across all customers belonging to state k is not equal to θ_k; this quantity is found by taking the expectation of $\theta_k^{\beta_i}$ over the distribution of β_i.) We impose the restrictions that $0 < \theta_k < 1$ for all k and that they increase with the level of commitment (i.e., $0 < \theta_1 < \theta_2 < \ldots < \theta_K < 1$).

It follows that the customer's usage likelihood function is

$$\begin{aligned} L_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data}) &= \prod_{t=1}^{T_i} P(Y_{it} = y_{it} | S_{it} = k, \boldsymbol{\theta}, \beta_i, m_t) \\ &= \prod_{t=1}^{T_i} \binom{m_t}{y_{it}} \big(\theta_k^{\beta_i}\big)^{y_{it}} \big(1 - \theta_k^{\beta_i}\big)^{m_t - y_{it}} \,. \end{aligned} \tag{D2}$$

Continuous Usage Process

As previously noted, the gamma and lognormal distributions would be natural candidates to accommodate a continuous usage process. We propose these distributions because (i) they ensure that usage is never negative, and (ii) cross-sectional heterogeneity in average usage can easily be accommodated by linking their parameters to the individual-level parameter β_i. We would use the following usage likelihood function in place of (6):

$$L_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \mid S_{it} = k, \text{data}) = \prod_{t=1}^{T_i} f(y_{it} \mid S_{it} = k, h(\boldsymbol{\theta}, \beta_i)) \tag{D3}$$

where $f(y_{it} \mid S_{it} = k, h(\boldsymbol{\theta}, \alpha_i))$ is the gamma or lognormal pdf and $h(\boldsymbol{\theta}, \beta_i)$ is the function that maps the usage rate parameters $(\boldsymbol{\theta}, \beta_i)$ to the parameters of the chosen distribution (i.e., the equivalent of (5)). In cases where we have individuals with zero-valued observations in several periods, then a mixture model combined with the gamma or lognormal distribution could be used to accommodate the non-positive observations (Yoo 2004).

Incorporating covariates into the model

In many situations the firm will have reliable individual-level data such as customer demographic information and details of interactions between the firm and the customer (e.g., direct mail and e-mail communications). One way to incorporate the effects of these data would be to make the usage rate in equation (5) a function of the available covariates:

$$\lambda_{it}|[S_{it} = k] = \theta_k \beta_i \exp(\boldsymbol{\delta}_1 \boldsymbol{x}_i + \boldsymbol{\delta}_2 \boldsymbol{z}_{it}), \tag{D4}$$

where $\boldsymbol{\delta}_1$ is the sensitivity of the usage process with respect to individual time-invariant covariates, denoted by $\boldsymbol{x}_i$, and $\boldsymbol{\delta}_2$ captures the effect of the time-varying covariates (e.g., marketing activities at time t), denoted by $\boldsymbol{z}_{it}$.

Moreover, this information could be used to explain heterogeneity in the unobserved state transition process, thus capturing the effect of such covariates on customer "commitment." In this case, a multinomial logit (or probit) model could be used in place of the Dirichlet distribution since the process of incorporating covariates is less cumbersome. Alternatively, we could follow the example of Netzer et al. (2008) and Montoya et al. (2010) and use an ordered logit (or probit) model to incorporate the such covariate effects. (Notice that in all of these cases, the posterior draws for the transition probabilities would require a Metropolis-Hastings step as there is no conjugate prior for these variables.)

References

Atchadé, Y.F. 2006. An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodology and Comput. Appl. Probab.* **8**(2) 235-254.

Berry, M.J.A., G.S. Linoff. 2004. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* John Wiley & Sons, Indianapolis, IN.

Blattberg, R.C., G. Getz, J.S. Thomas. 2001. *Customer Equity: Building and Managing Relationships as Valuable Assets.* Harvard Business School Press, Boston, MA.

Blattberg, R.C., B-D. Kim, S.A. Neslin. 2008. *Database Marketing. Analyzing and Managing Customers.* Springer, New York, NY.

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

Bolton, R.N., P.K. Kannan, M.D. Bramlett. 2000. Implications of loyalty program membership and service experiences for customer retention and value. *J. Academy Marketing Sci.* **28**(Winter) 95–108.

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

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

Danaher, P.J. 2002. Optimal pricing of new subscription services: Analysis of a market experiment. *Marketing Sci.* **21**(Spring) 119–138.

Diggle, P., M.G. Kenward. 1994. Informative drop-out in longitudinal data analysis. *Appl. Stat.* **43**(1) 49–93.

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

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

Fader, P.S., B.G.S. Hardie, K.L. Lee. 2005. RFM and CLV: Using iso-value curves for customer base analysis. *J. Marketing Res.* **42**(November) 415–430.

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

Gruen, T.W., J.O. Summers, F. Acito. 2000. Relationship marketing activities, commitment, and membership behaviors in professional associations. *J. Marketing* **64**(July) 34–49.

Hanemann, W.M. 1984. Discrete/continuous models of consumer demand. *Econometrica* **52**(3) 541–561.

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

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

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

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

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

Lemon, K.N., T.B. White, R.S. Winer. 2002. Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *J. Marketing* **66**(January) 1–14.

Montgomery, A.L., S. Li, K. Srinivasan, J.C. Liechty. 2004. Predicting online purchase conversion using web path analysis. *Marketing Sci.* **23**(4) 579–595.

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

Moon, S., W.A. Kamakura, J. Ledolter. 2007. Estimating promotion response when competitive promotions are unobservable. *J. Marketing Res.* **44**(3) 503–515.

Morgan, R.Y.H., S. Hunt. 1994. The commitment-trust theory of relationship marketing. *J. Marketing* **58**(July) 20–38.

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

Narayanan, S., P.K. Chintagunta, E.J. Miravete. 2007. The role of self selection and usage uncertainty in the demand for local telephone service. *Quantitative Marketing and Econom.* **5**(1) 1–34.

Neslin, S., S. Gupta, W. Kamakura, J. Lu, C.H. Mason. 2006. Defection detection: Measuring and understanding the predictive accuracy of customer churn models," *J. Marketing Res.*, **46**(2) 204–211.

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

Parr Rud, O. 2001. *Data Mining Cookbook.* John Wiley & Sons, New York, NY.

Poulsen, C.S. 1990. Mixed Markov and latent Markov modelling applied to brand choice behaviour. *Internat. J. Res. in Marketing* **7**(1) 5–19.

Rust, R.T., A.J. Zahorik. 1993. Customer satisfaction, customer retention, and market share. *J. Retailing* **69**(2) 193–215.

Rust, R.T., V.A. Zeithaml, K.N. Lemon. 2001. *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy.* The Free Press, New York, NY.

Schweidel, D.A., E.T. Bradlow, P.S. Fader. 2011. Portfolio dynamics for customers of a multiservice provider. *Management Sci.* **57**(3) 471–486.

Scott, S. 2002. Bayesian methods for hidden Markov models. *J. Amer. Statist. Assoc.* **97**(457) 337–351.

Smith, A., P.A. Naik, C.L. Tsai. 2006. Markov-switching model selection using Kullback-Leibler divergence. *J. Econometrics* **134**(2) 553–577.

Verhoef, P.C. 2003. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *J. Marketing* **67**(October) 30–45.

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

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

Yoo, S. 2004. A note on an approximation on the mobile communications expenditures distribution function using a mixture model. *J. Appl. Statist.* **31**(August) 747–752.