A Joint Model of Usage and Churn in Contractual Settings

Web Appendix

Eva Ascarza

Bruce G. S. Hardie

Appendix A: MCMC Procedure for the Proposed Model

The model is estimated using a hierarchical 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 $\boldsymbol{\Omega}$ 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 $\boldsymbol{\Pi} = \{\boldsymbol{\Pi}_i\}_{i=1,\ldots,I}$, and the set of augmented paths of commitment states $\boldsymbol{s} = \{\tilde{s}_i\}_{i=1,\ldots,I}$. We write the full joint posterior distribution as

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

where $f(\boldsymbol{s}|\boldsymbol{q}, \boldsymbol{\Pi})$ refers to the distribution of the latent states, assumed to follow a hidden Markov process with renewal restrictions. The term $f(\boldsymbol{\Pi}|\boldsymbol{A})$ corresponds to the prior (or mixing) distribution for the individual transition probabilities. Each row j of the matrix $\boldsymbol{\Pi}_i$ is assumed to follow a Dirichlet distribution with parameter vector $[\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jK}]$; we let $\boldsymbol{A}$ denote 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_is, where β_i is assumed to follow a lognormal distribution with mean 0 and standard deviation σ_β.

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-Gamma prior with parameter $R = 0.05$ and degrees of freedom $df = 2$. Noting that $q_1 = 0$, we use a Dirichlet prior with a $1 \times (K-1)$ parameter vector of ones for the remaining elements of $\boldsymbol{q}$.

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. For mathematical convenience we reparameterize $\alpha_{jk} = e^{\rho_{jk}} \ \forall j, k \in \{1, \ldots, K\}$ and estimate $\boldsymbol{\rho} = [\rho_{11}, \ldots, \rho_{1K}, \ldots, \rho_{K1}, \ldots, \rho_{KK}]$. We assume $\Phi = \{\boldsymbol{\gamma}, \boldsymbol{\rho}\}$ follows a multivariate normal distribution with parameters $\mu_\Phi = [3 \times \mathbb{1}_K, 4 \times \mathbb{1}_{K^2}]$ and $\text{diag}(\Sigma_\Phi) = [\mathbb{1}_K, (1/2) \times \mathbb{1}_{K^2}]$, where $\mathbb{1}_n$ is a $1 \times n$ vector of ones. (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] $f(\sigma_\beta|\boldsymbol{\beta}, R, df) \sim$ inv-Gamma$(\sum_{i=1}^{I}((\ln\beta_i)^2 + (df/R), df + I)$.
- [Gibbs] $f([q_2, \ldots, q_K]|\boldsymbol{s}) \sim$ Dirichlet$(1 + n_{02}, \ldots, 1 + n_{0K})$, where $n_{0k} = \sum_{i=1}^{I} \mathbf{1}(s_{i1} = k)$.
- [Metropolis-Hastings] $f(\Phi|\mu_\Phi, \Sigma_\Phi, \boldsymbol{s}, \text{data}) \propto \exp\left(-.5(\Phi - \mu_\Phi)'\Sigma_\Phi^{-1}(\Phi - \mu_\Phi)\right) f(\text{data}|\boldsymbol{\beta}, \Phi, \boldsymbol{s})$, where
$$f(\text{data}|\boldsymbol{\beta}, \Phi, \boldsymbol{s}) = \prod_{i=1}^{I} f(\text{data}|\beta_i, \Phi, \tilde{s}_i)$$
and $f(\text{data}|\beta_i, \Phi, \tilde{s}_i) = \mathscr{L}_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data})$ with the $\boldsymbol{\theta} \to \boldsymbol{\Phi}$ mapping discussed above. We use a Gaussian random-walk Metropolis-Hasting algorithm to draw from this distribution; in particular, we follow the procedure proposed by Atchadé (2006) and adapt the tuning parameters in each iteration to get an acceptance rate of approximately 20%.
- For each individual i,
  - [Gibbs] For the jth row of $\boldsymbol{\Pi}_i$, $f(\boldsymbol{\pi}_{ij}|\Phi, \boldsymbol{s}) \sim$ Dirichlet$(\alpha_{j1} + n_{ij1}, \ldots, \alpha_{jK} + n_{ijK})$, where $n_{ijk} = \sum_{t=1}^{T_i-1} \mathbf{1}(s_{it} = j \text{ and } s_{it+1} = k)$, where $\mathbf{1}(\cdot)$ is the indicator function that equals 1 if the condition is met, 0 otherwise.
  - [Metropolis-Hastings] $f(\beta_i|\sigma_\beta, \Phi, \tilde{s}_i, \text{data}) \propto \exp\left(\frac{\beta_i^2}{2\sigma_\beta}\right) f(\text{data}|\beta_i, \Phi, \tilde{s}_i)$.
  We use a Gaussian random-walk Metropolis-Hasting algorithm to draw from this distribution; in particular, we follow the procedure proposed by Atchadé (2006) and adapt the tuning parameters in each iteration to get an acceptance rate of approximately 20%.
  - [Gibbs] We draw from the distribution of the hidden states using the direct Gibbs sampler approach proposed by Scott (2002) (eq(8) p.340):
$$P(S_{i1} = k|\boldsymbol{q}, \dot{s}_{i(1)}, \text{data}) \propto q_k P(S_{i2} = s_{i2}|S_{i1} = k)\mathbf{1}(\tilde{s}_{i(1,k)} \in \Upsilon_i)$$
$$P(S_{it} = k|\boldsymbol{\Pi_i}, \dot{s}_{i(t)}, \text{data}) \propto P(S_{it} = k|S_{it-1} = s_{it-1}) \times P(S_{it+1} = s_{it+1}|S_{it} = k)\mathbf{1}(\tilde{s}_{i(t,k)} \in \Upsilon_i),$$

where $\dot{s}_{i(t)} = [s_{i1}, ..., s_{it-1}, s_{it+1}, ..., s_{iT_i}]$ and $\tilde{s}_{i(t,k)} = [s_{i1}, ..., s_{it-1}, k, s_{it+1}, ..., s_{iT_i}]$, and Υ_i is the set of possible paths through the commitment states given T_i periods. When $t = T_i$, $P(S_{it+1} = s_{it+1}|S_{it} = k) = 1$.

In the empirical analysis reported in the paper, 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. Convergence was assessed by visual inspection and confirmed using the Geweke (1992) convergence diagnostic.

Appendix B: Exploring the Model Identification with Simulations

In this appendix we present the simulation analyses that were performed to confirm the identification of the proposed model specification. We simulate and estimate multiple versions of the full model (i.e., the model with unobserved heterogeneity in both usage and transition dynamics), varying the number of states (K), the initial probabilities ($\boldsymbol{q}$), and the heterogeneity in transition probabilities ($\boldsymbol{A}$).

We use three sets of parameter vectors in this analysis:

Set 1: Equal initial state probabilities

	Number of states		
	K = 2	K = 3	K = 4
$\boldsymbol{q}$	[0 1]	[0 .5 .5]	[0 .333 .333 .333]
$\boldsymbol{A}$	$\begin{bmatrix} 20 & 5 \\ 5 & 20 \end{bmatrix}$	$\begin{bmatrix} 20 & 5 & 1 \\ 5 & 20 & 5 \\ 1 & 5 & 20 \end{bmatrix}$	$\begin{bmatrix} 20 & 5 & 1 & 0.1 \\ 5 & 20 & 5 & 1 \\ 1 & 5 & 20 & 5 \\ 0.1 & 1 & 5 & 20 \end{bmatrix}$
$\boldsymbol{\theta}$	[0.1 2]	[0.1 2 5]	[0.1 2 5 10]
σ_β	0.1	0.1	0.1

Set 2: Unequal initial state probabilities

	Number of states	
	K = 3	K = 4
$\boldsymbol{q}$	[0 .2 .8]	[0 .1 .3 .6]
$\boldsymbol{A}$	$\begin{bmatrix} 20 & 5 & 1 \\ 5 & 20 & 5 \\ 1 & 5 & 20 \end{bmatrix}$	$\begin{bmatrix} 20 & 5 & 1 & 0.1 \\ 5 & 20 & 5 & 1 \\ 1 & 5 & 20 & 5 \\ 0.1 & 1 & 5 & 20 \end{bmatrix}$
$\boldsymbol{\theta}$	[0.1 2 5]	[0.1 2 5 10]
σ_β	0.1	0.1

Set 3: Unequal initial state probabilities with more heterogeneous transition probabilities

	Number of states	
	K = 3	K = 4
$\boldsymbol{q}$	[0 .2 .8]	[0 .1 .3 .6]
$\boldsymbol{A}$	$\begin{bmatrix} 10 & 2.5 & 0.5 \\ 2.5 & 10 & 2.5 \\ 0.5 & 2.5 & 10 \end{bmatrix}$	$\begin{bmatrix} 10 & 2.5 & 0.5 & 0.05 \\ 2.5 & 10 & 2.5 & 0.5 \\ 0.5 & 2.5 & 10 & 2.5 \\ 0.05 & 0.5 & 2.5 & 10 \end{bmatrix}$
$\boldsymbol{\theta}$	[0.1 2 5]	[0.1 2 5 10]
σ_β	0.1	0.1

We simulate customer behavior assuming the data generating process of our proposed model (as presented in Section 3.1) and fit the model to these simulated datasets using the model estimation procedure described in Web Appendix A. As reported in Tables B1 to B7, the 95% central posterior intervals (CPIs) include the simulated values for all but three of the 124 simulated parameters in the seven cases considered in the simulation. (We do not report the elements of $\boldsymbol{\theta}$ and $\boldsymbol{A}$; rather we report their reparameterizations (i.e., $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$).) We therefore conclude that the proposed model is identified.

Parameter	Simulated	Posterior mean	95% CPI
ρ_{11}	3.00	3.86	[2.51 5.12]
ρ_{12}	1.61	2.37	[1.06 3.64]
ρ_{21}	1.61	1.58	[1.33 1.93]
ρ_{22}	3.00	2.91	[2.63 3.31]
γ_1	−2.30	−2.32	[−2.46 −2.22]
γ_2	0.64	0.63	[0.61 0.65]
σ_β	0.10	0.11	[0.08 0.13]

Table B1: Simulated and estimated values of Set 1 parameters ($K = 2$).

Parameter	Simulated	Posterior mean	95% CPI
q_1	0.50	0.46	[0.42 0.51]
q_2	0.50	0.54	[0.49 0.58]
ρ_{11}	3.00	3.09	[2.88 3.37]
ρ_{12}	1.61	1.68	[1.51 1.87]
ρ_{13}	0.00	0.10	[−0.25 0.42]
ρ_{21}	1.61	1.62	[1.54 1.70]
ρ_{22}	3.00	2.97	[2.87 3.05]
ρ_{23}	1.61	1.62	[1.51 1.77]
ρ_{31}	0.00	−0.13	[−0.31 0.03]
ρ_{32}	1.61	1.40	[1.15 1.67]
ρ_{33}	3.00	2.76	[2.53 3.03]
γ_1	−2.30	−2.43	[−2.54 −2.29]
γ_2	0.64	0.62	[0.60 0.65]
γ_3	1.10	1.10	[1.07 1.13]
σ_β	0.10	0.10	[0.09 0.13]

Table B2: Simulated and estimated values of Set 1 parameters ($K = 3$).

Parameter	Simulated	Posterior mean	95% CPI	
q_1	0.33	0.34	[0.30	0.38]
q_2	0.33	0.31	[0.27	0.36]
q_3	0.33	0.35	[0.31	0.39]
ρ_{11}	3.00	2.94	[2.84	3.04]
ρ_{12}	1.61	1.58	[1.47	1.69]
ρ_{13}	0.00	0.00	[−0.14	0.21]
ρ_{14}	−2.30	−2.38	[−2.45	−2.31]
ρ_{21}	1.61	1.57	[1.48	1.65]
ρ_{22}	3.00	2.96	[2.89	3.04]
ρ_{23}	1.61	1.53	[1.42	1.63]
ρ_{24}	0.00	−0.04	[−0.12	0.05]
ρ_{31}	0.00	−0.04	[−0.16	0.08]
ρ_{32}	1.61	1.61	[1.48	1.78]
ρ_{33}	3.00	2.87	[2.76	2.98]
ρ_{34}	1.61	1.53	[1.39	1.66]
ρ_{41}	−2.30	−2.24	[−2.40	−2.08]
ρ_{42}	0.00	0.01	[−0.05	0.09]
ρ_{43}	1.61	1.54	[1.43	1.64]
ρ_{44}	3.00	2.98	[2.93	3.04]
γ_1	−2.30	−2.20	[−2.34	−2.05]
γ_2	0.64	0.67	[0.64	0.71]
γ_3	1.10	1.09	[1.03	1.14]
γ_4	1.61	1.58	[1.55	1.61]
σ_β	0.10	0.11	[0.10	0.13]

Table B3: Simulated and estimated values of Set 1 parameters ($K = 4$).

Parameter	Simulated	Posterior mean	95% CPI	
q_1	0.20	0.19	[0.16	0.23]
q_2	0.80	0.81	[0.77	0.84]
ρ_{11}	3.00	2.84	[2.63	3.01]
ρ_{12}	1.61	1.41	[1.15	1.62]
ρ_{13}	0.00	0.08	[−0.15	0.31]
ρ_{21}	1.61	1.66	[1.24	1.99]
ρ_{22}	3.00	3.14	[2.66	3.52]
ρ_{23}	1.61	1.63	[1.16	2.02]
ρ_{31}	0.00	0.00	[−0.28	0.26]
ρ_{32}	1.61	1.62	[1.37	1.82]
ρ_{33}	3.00	3.03	[2.73	3.28]
γ_1	−2.30	−2.23	[−2.37	−2.11]
γ_2	0.64	0.63	[0.59	0.66]
γ_3	1.10	1.09	[1.06	1.12]
σ_β	0.10	0.11	[0.09	0.13]

Table B4: Simulated and estimated values of Set 2 parameters ($K = 3$).

Parameter	Simulated	Posterior mean	95% CPI	
q_1	0.10	0.09	[0.06	0.11]
q_2	0.30	0.29	[0.25	0.34]
q_3	0.60	0.62	[0.58	0.66]
ρ_{11}	3.00	3.05	[2.96	3.16]
ρ_{12}	1.61	1.53	[1.41	1.63]
ρ_{13}	0.00	0.19	[0.05	0.36]
ρ_{14}	−2.30	−2.27	[−2.55	−2.01]
ρ_{21}	1.61	1.49	[1.32	1.71]
ρ_{22}	3.00	2.89	[2.68	3.06]
ρ_{23}	1.61	1.53	[1.37	1.66]
ρ_{24}	0.00	0.13	[−0.22	0.49]
ρ_{31}	0.00	0.17	[0.00	0.34]
ρ_{32}	1.61	1.49	[1.34	1.65]
ρ_{33}	3.00	2.98	[2.83	3.12]
ρ_{34}	1.61	1.63	[1.42	1.85]
ρ_{41}	−2.30	−2.51	[−2.69	−2.30]
ρ_{42}	0.00	−0.07	[−0.25	0.07]
ρ_{43}	1.61	1.46	[1.33	1.63]
ρ_{44}	3.00	2.91	[2.83	3.02]
γ_1	−2.30	−2.28	[−2.42	−2.12]
γ_2	0.64	0.63	[0.60	0.67]
γ_3	1.10	1.07	[1.03	1.10]
γ_4	1.61	1.62	[1.60	1.64]
σ_β	0.10	0.10	[0.08	0.11]

Table B5: Simulated and estimated values of Set 2 parameters ($K = 4$).

Parameter	Simulated	Posterior mean	95% CPI	
q_1	0.20	0.19	[0.16	0.23]
q_2	0.80	0.81	[0.77	0.84]
ρ_{11}	2.30	2.33	[1.85	2.90]
ρ_{12}	0.92	1.15	[0.62	1.72]
ρ_{13}	−0.69	−0.67	[−1.05	−0.16]
ρ_{21}	0.92	1.12	[0.88	1.36]
ρ_{22}	2.30	2.48	[2.16	2.79]
ρ_{23}	0.92	1.03	[0.68	1.31]
ρ_{31}	−0.69	−0.70	[−0.86	−0.46]
ρ_{32}	0.92	0.88	[0.64	1.10]
ρ_{33}	2.30	2.33	[2.09	2.59]
γ_1	−2.30	−2.32	[−2.53	−2.18]
γ_2	0.64	0.66	[0.62	0.69]
γ_3	1.10	1.08	[1.05	1.11]
σ_β	0.10	0.10	[0.08	0.12]

Table B6: Simulated and estimated values of Set 3 parameters ($K = 3$).

Parameter	Simulated	Posterior mean	95% CPI	
q_1	0.10	0.10	[0.07	0.12]
q_2	0.30	0.28	[0.24	0.32]
q_3	0.60	0.62	[0.59	0.66]
ρ_{11}	2.30	2.40	[2.25	2.59]
ρ_{12}	0.92	1.10	[0.74	1.34]
ρ_{13}	−0.69	−0.85	[−1.22	−0.40]
ρ_{14}	−3.00	−3.10	[−3.25	−2.94]
ρ_{21}	0.92	0.93	[0.73	1.13]
ρ_{22}	2.30	2.38	[2.23	2.51]
ρ_{23}	0.92	0.86	[0.70	0.99]
ρ_{24}	−0.69	−0.68	[−0.91	−0.50]
ρ_{31}	−0.69	−0.59	[−0.80	−0.37]
ρ_{32}	0.92	0.96	[0.86	1.08]
ρ_{33}	2.30	2.37	[2.25	2.53]
ρ_{34}	0.92	0.98	[0.89	1.10]
ρ_{41}	−3.00	−2.90	[−3.13	−2.69]
ρ_{42}	−0.69	−0.70	[−0.95	−0.49]
ρ_{43}	0.92	0.97	[0.71	1.19]
ρ_{44}	2.30	2.29	[1.97	2.50]
γ_1	−2.30	−2.23	[−2.35	−2.09]
γ_2	0.64	0.62	[0.59	0.65]
γ_3	1.10	1.09	[1.06	1.12]
γ_4	1.61	1.61	[1.59	1.63]
σ_β	0.10	0.10	[0.09	0.12]

Table B7: Simulated and estimated values of Set 3 parameters ($K = 4$).

Appendix C: Model with Seasonal Dummies and Time Trend

In this appendix we present the results for alternative model specifications that allow for seasonality and a time trend in the usage process.

Model with seasonal dummies: We first estimate a model in which we allow for seasonality in usage behavior. Recalling the discussion in Sections 3.1 and 3.3, we replace (5) with

$$\lambda_{it} \,|\, [S_{it} = k] = \theta_k \beta_i \exp(\delta_1 d_{1t} + \delta_2 d_{2t} + \delta_3 d_{3t}) \,, \tag{C1}$$

where $d_{1t} = 1$ if t corresponds to the first quarter of the year, 0 otherwise, $d_{2t} = 1$ if t corresponds to the second quarter of the year, 0 otherwise, etc.

Table C1 reports the posterior means and 95% central posterior intervals (CPIs) for the parameters of the usage model under the three-state specification (cf. Table 3), Table C2 reports the posterior estimate of $\boldsymbol{q}$ (cf. Table 4), and Table C3 reports the average and 95% interval of the individual posterior means of the transition probabilities (cf. Table 5).

	Parameter	Posterior mean	95% CPI
Usage	θ_1	0.21	[0.19 0.23]
Propensity	θ_2	0.22	[0.20 0.24]
	θ_3	1.19	[1.11 1.27]
Heterogeneity	σ_β	0.91	[0.85 0.98]
Quarterly dummies	$\exp(\delta_1)$	0.84	[0.80 0.90]
	$\exp(\delta_2)$	0.88	[0.84 0.93]
	$\exp(\delta_3)$	1.04	[0.98 1.10]

Table C1: Usage parameters for the model with seasonality in the usage process.

Parameter	Posterior mean	95% CPI
q_1	0.00	--
q_2	0.41	[0.33 0.48]
q_3	0.59	[0.52 0.67]

Table C2: Initial-state parameters for the model with seasonality in the usage process.

Table C4 compares the accuracy of the usage forecasts from the specification with seasonality in the usage process with those of the proposed model for period 12 (cf. Table 7) and periods

From state	To state 1	To state 2	To state 3
1	0.663	0.332	0.004
	[0.659 0.668]	[0.328 0.336]	[0.004 0.005]
2	0.299	0.436	0.266
	[0.127 0.580]	[0.276 0.735]	[0.111 0.549]
3	0.084	0.206	0.711
	[0.017 0.173]	[0.050 0.309]	[0.555 0.933]

Table C3: Mean transition probabilities and the 95% interval of individual posterior means for the model with seasonality in the usage process.

14–16 (cf. Table 9). The inclusion of seasonality effects in the usage process does not lead to any improvement in the accuracy of the usage forecasts.

Table C5 compares the accuracy of the renewal forecasts associated with these two model specifications (cf. Table 8). The results are mixed. The specification with seasonality in the usage process is slightly more accurate in terms of predicting total churn, but has a lower hit rate.

	Aggregate (% error)	Disaggregate (χ^2)	Individual (MSE)
Period 12:			
Proposed model	−7.2	6.5	1.4
Seasonality	−13.5	17.8	1.5
Periods 14–16:			
Proposed model	2.4	16.0	3.1
Seasonality	7.0	16.2	3.4

Table C4: Assessing the accuracy of usage forecasts.

	Period 13			Period 17		
	Renewal Rate	% error	Hit Rate	Renewal Rate	% error	Hit Rate
Proposed model	88%	2.7	78%	91%	0.5	68%
Seasonality	87%	1.8	77%	90%	−0.4	67%
Actual	86%	--	--	91%	--	--

Table C5: Assessing the predictions of period 13 and 17 renewal.

Taken together, we conclude that, in this particular case, there is no substantive benefit associated with an alternative specification that allows for seasonality in the usage process.

Model with seasonal dummies and time trend: We extend the seasonality in usage model by including a parameter to capture any possible trend in usage behavior. This sees us replacing (C1) with

$$\lambda_{it} \,|\, [S_{it} = k] = \theta_k \beta_i \exp(\delta_1 d_{1t} + \delta_2 d_{2t} + \delta_3 d_{3t} + \delta_4 t)\,, \tag{C2}$$

where δ_4 is a parameter that captures any time trend.

We find that while the additional trend parameter is positive (posterior mean: 0.005), it does not have any significant impact on usage behavior (95% CPI: $[-0.002, 0.013]$). The rest of the parameters are consistent with the previous results.

Appendix D: Alternative Model Specifications

The proposed model assumes that, conditional on the underlying state, the usage behavior of interest is characterized by the Poisson distribution. In some settings, usage per period is a discrete quantity with an upper bound and may be better characterized by the binomial distribution. In other settings, the usage behavior of interest is a non-negative continuous quantity and should be characterized by distributions such as the gamma or lognormal. We now consider how the model specification can be changed for these alternative settings.

D1 Binomial Specification for the Usage Model

For each customer i we have a total of T_i usage observation periods. Let m_t denote the number of transaction opportunities (e.g., number of days) in usage observation period t, y_{it} be customer i's observed usage in period t, 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-specific time-invariant parameter β_i and the commitment state at every period:

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

We impose the restrictions that $0 < \theta_k < 1$ for all k, and that the θ_ks increase with the level of commitment (i.e., $0 < \theta_1 < \theta_2 < \ldots < \theta_K < 1$). The usage propensity parameter β_i is assumed to follow a lognormal distribution with mean 0 and standard deviation σ_β. 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 transaction probability 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.) This specification guarantees that the transaction probability is increasing with the level of commitment.

Recalling that $\tilde{S}_i = [S_{i1}, S_{i2}, \ldots, S_{i\,T_i}]$ denotes the (unobserved) sequence of states to which customer i belongs during her entire lifetime, with realization $\tilde{s}_i = [s_{i1}, s_{i2}, \ldots, s_{i\,T_i}]$, the customer's usage likelihood function is

$$L_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data}) = \prod_{t=1}^{T_i} P(Y_{it} = y_{it} \,|\, m_t, S_{it} = s_{it}, \boldsymbol{\theta}, \beta_i)$$
$$= \prod_{t=1}^{T_i} \binom{m_t}{y_{it}} (\theta_{s_{it}}^{\beta_i})^{y_{it}} (1 - \theta_{s_{it}}^{\beta_i})^{m_t - y_{it}}, \tag{D2}$$

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

D2 Continuous Usage Process

As previously noted, the gamma and lognormal distributions are natural candidates for accommodating 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:

$$L_i^{\text{usage}}(\boldsymbol{\theta}, \beta_i \,|\, \tilde{S}_i = \tilde{s}_i, \text{data}) = \prod_{t=1}^{T_i} f(y_{it} \,|\, S_{it} = s_{it}, \boldsymbol{\theta}, \beta_i)\,, \tag{D3}$$

where $f(y_{it} \,|\, S_{it} = s_{it}, \boldsymbol{\theta}, \beta_i)$ is the gamma or lognormal pdf and there exists some function $h(\theta_{s_{it}}, \beta_i)$ that maps the individual-specific time-invariant parameter β_i and the commitment state at every period s_{it} to the parameters of the chosen distribution (i.e., the equivalent of (5) and (D1)). In cases where we have individuals with zero-valued observations in several periods, a mixture model combined with the gamma or lognormal distribution could be used to accommodate the non-positive observations (Yoo 2004).

Appendix E: Further Validation Analysis

We further assess the validity of the proposed model by looking at the relationship between the observed behaviors and the states to which customers are assigned.

- Standing at the end of time t, we create three groups of customers: (1) those whose usage in both the current and last period was below their individual average (computed across periods 1 to $t-2$), (2) those whose usage in the current period was below their average (but not in the period before that), and (3) the rest of customers, for whom usage in the current period was at or above their average.

- We then compute, for each group, the probability of being assigned to each state. So as to emphasize the distinction between states 1 and 2, we also compute the ratio between the probability of being assigned to state 1 and the probability of being assigned to state 2 for each customer group; this captures the relative probability of churning.

- Finally, we relate this information to observed churn behavior and compute, for each group, the proportion of customers who actually churned, and within the churners, the proportion who churned from state 1, state 2, and state 3.

The following two tables report these results for the case of $t = 8$. (We also considered the case of $t = 4$ and obtained similar results.)

We see from Table E1 that the probability of being assigned to state 1 is highest when the customer's usage in the last two periods is below their individual average, and it decreases monotonically as customers show higher levels of usage in recent periods. We note that the ratio of the probability of belonging to state 1 to that of belonging to state 2 is much higher — almost double — when customers have exhibited lower than average levels of usage for two periods in a row.

We observe in Table E2 that the churn rate is highest for those customers whose usage in the previous two periods is below their individual average. Looking across the last three columns, we observe how individuals assigned to state 1 have much higher churn rates than those customers assigned to state 2; this difference is especially pronounced for those customers in the "below average in periods $t-1$ and t" group.

Observed usage is below average	% Assigned to state 1	% Assigned to state 2	% Assigned to state 3	% state 1 / % state 2
in periods $t-1$ and t	24	48	28	0.51
in just period t	11	40	49	0.27
for neither period	8	25	67	0.34

Table E1: The relationship between state membership in period t and relative usage.

Observed usage is below average	Observed churn	% Churning from state 1	% Churning from state 2	% Churning from state 3
in periods $t-1$ and t	25%	72	24	4
in just period t	11%	62	31	7
for neither period t	11%	55	21	24

Table E2: The relationship between churn in period $t+1$ and relative usage.

These results provide evidence of the validity of the latent states inferred by the proposed model.

Appendix F: 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 (to date) 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 (to date). (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 select the specification that provides the most accurate in-sample hit-rates. The associated parameter estimates are given in Table F1. 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 $\times$ Fsum	-0.071	0.015
Frequency $\times$ Monetary value	-0.002	0.000
LL	-327.4	

Table F1: Parameter estimates for the cross-sectional logistic regression model.

Given the nature of the usage data, we use a Poisson regression model with a normal random effect to account for the observed overdispersion in the data. We select those individuals that were still members at the end of our calibration period, using the number of transactions in the last period ($t = 11$) as the dependent variable and the RFM measures as predictors (Table F2). 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 F2: Parameter estimates for the cross-sectional Poisson regression model.

Noting that our dataset has multiple observations per individual, not just the information for the most recent period, we can extend the cross-sectional models and estimate longitudinal models using (where available) more than one observation per customer. 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 F3 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 the variance of the random effect is not significant.)

Similarly, we estimate a random-effects (panel) Poisson regression model using transaction behavior from all preceding periods — see Table F4. 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 significant.

	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 value	−0.001	0.000
Frequency × Monetary value	−0.003	0.000
LL	−775.6	

Table F3: 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 F4: Parameter estimates for the panel Poisson regression model.

Appendix G: Estimating the Bivariate Model

As discussed in Section 4.3 of the paper, another way to model our data is to use a Tobit-type model. Given that customers need to be "under contract" in order to use the service, we can relate usage observations to renewal behavior as in a Type II Tobit model and therefore correct for a possible selectivity bias. Such an approach would assume the existence of two latent variables — one driving renewal decisions, the other usage — instead of the single latent variable our proposed model assumes.

This approach can be seen as an extension of the traditional Type II Tobit model (Wooldridge 2002, p. 562), and is similar to the model used by Reinartz et al. (2005) to model customer profitability while correcting for acquisition, and the extensions of the Tobit models presented in Blattberg et al. (2008, pp. 391–392) to model censored data with selection effects. The two main differences between our setting and theirs is that our selection variable (renewal) occurs every n periods instead of just once (acquisition or adoption), and that our variable of interest is not continuous (number of transactions). As a consequence, we cannot make use of existing statistical routines, but we can adapt the likelihood function to accommodate these two changes.

In order to account for nonstationarity in the usage and renewal decisions, we also incorporate the effects of past usage in both equations. We add linear and quadratic terms for the effect of lagged usage so as to capture potential nonlinear effects. More formally, the model is specified as follows.

Usage behavior: While under contract, a customer's usage behavior is observed every period. We assume that the number of transactions for individual i in period t follows a Poisson distribution with parameter λ_{it}, which is determined by an individual-level parameter, the (non-linear) effect of past usage (y_{it-1}), and an unobserved random shock:

$$\lambda_{it} = \exp(\mu_i + \delta_1 y_{it-1} + \delta_2 y_{it-1}^2 + \epsilon_{it}) \,, \text{ for } t = 1, 2, 3, \ldots \,, \tag{F1}$$

where μ_i is normally distributed across the population with parameters $(\tilde{\mu}, \sigma_\mu)$.

Renewal behavior: At the end of each contract period, the customer makes the decision of whether or not to renew her membership. She renews with probability p_{it}, which is specified as

$$p_{it} = \frac{e^{\omega+\delta_3 y_{it-1}+\delta_4 y_{it-1}^2+\nu_{it}}}{1+e^{\omega+\delta_3 y_{it-1}+\delta_4 y_{it-1}^2+\nu_{it}}}\,, \text{ for } t = n, 2n, 3n, \ldots \tag{F2}$$

That is, renewal behavior is determined by the (non-linear) effect of past usage and an unobserved random shock.[1]

In order to capture the potential relationship between usage and renewal decisions (hence correcting for any selection effect), we allow the two random shocks to be correlated in the following manner:

$$\begin{bmatrix} \epsilon_{it} \\ \nu_{it} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\nu \\ \rho\sigma_\epsilon\sigma_\nu & \sigma_\nu^2 \end{bmatrix} \right),$$

where σ_ϵ is set to 1 to ensure identification.

We estimated the model in a Bayesian manner using the freely available WinBUGS software. Uninformative (vague) priors were used for all parameters in the model. We ran the simulation for 600,000 iterations. The first 500,000 iterations were used as a "burn-in" period, and the last 100,000 iterations were used to estimate the conditional posterior distributions. We examined the convergence of the parameters by visual inspection. The Geweke convergence diagnostic also confirmed that the parameters had converged. The posterior means and 95% CPIs are reported in Table F1.

We note that there is no significant effect of past usage on current usage (as captured by δ_1 and δ_2). However, the relationship between past usage and renewal behavior is significant and non-linear. This later result should come as no surprise as it has been well documented in the CRM literature (e.g., Blattberg et al. 2008). We note that this relationship exists above and beyond common temporary shocks affecting usage and renewal decisions.

[1] ω does not have subscript i because we are unable to identify unobserved heterogeneity in this parameter.

Parameter	Posterior mean	95% CPI	
$\tilde{\mu}$	1.294	[1.154	1.435]
σ_μ	1.052	[0.993	1.113]
δ_1	0.004	[−0.013	0.021]
δ_2	0.000	[−0.000	0.001]
ω	−0.735	[−0.808	−0.662]
δ_3	0.387	[0.279	0.502]
δ_4	−0.008	[−0.012	−0.005]
σ_ν	0.482	[0.442	0.521]
ρ	0.631	[0.360	0.855]

Table F1: Parameter estimates for the model with two latent variables.

In the spirit of Borle et al. (2008), we also considered a more complex model in which linear and quadratic effects of time (i.e., t and t^2) were added to (F1) and linear and quadratic effects of cumulative renewal occasions (i.e., t/n and $(t/n)^2$) were added to (F2). None of these additional parameters were significant.

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.

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

Borle, S., S. Singh, and D. Jain. 2008. Customer Lifetime Value Measurement. *Management Sci.* **54**(1) 100–112.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (eds), Oxford University Press, Oxford, 169–193.

Reinartz, W., J.S. Thomas, V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *J. Marketing* **69**(1) 63–79.

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

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, Cambridge, MA.

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.