Personal and Social Usage: The Origins of Active Customers and Ways to Keep Them Engaged

Technical Appendix [*]

Clarence Lee

Elie Ofek

Thomas Steenburgh

Version: January 18, 2017

Forthcoming at Management Science

---

[*] Clarence Lee is an assistant professor at the S.C. Johnson School of Management at Cornell University (clarence.lee@cornell.edu). Elie Ofek is the T.J. Dermot Dunphy Professor of Business Administration, Harvard Business School (eofek@hbs.edu). Thomas Steenburgh is the Paul M. Hammaker Professor of Business Administration, University of Virginia, Darden School of Business (SteenburghT@darden.virginia.edu).

1 Model Specification and Estimation Details

In this section, we describe our model specification in detail. Table 1 is a summary of the model parameters, corresponding variables, and interpretations. We describe the likelihood specification and choice of priors in the following sections.

1.1 Likelihood Specification

Under the standard HMM specification, an individual's usage probabilities are correlated within each path through the unobserved states r_{it} and v_{it}. Therefore, the joint likelihood must be summed over all possible paths an individual could take over all T time periods:

$$
\begin{aligned}
L_{iT} = {} & P(Y^p_{i1} = y^p_{i1}, ..., Y^p_{iT} = y^p_{iT}, Y^s_{i1} = y^s_{i1}, ..., Y^s_{iT} = y^s_{iT}) \\
= {} & \sum_{r_1=\{1,...,N\}} \cdots \sum_{r_T=\{1,...,N\}} \cdot \sum_{v_1=\{1,...,N\}} \cdots \sum_{v_T=\{1,...,N\}} [P(R_1 = r_1, V_1 = v_1) \\
& \times \prod_{t=2}^{T} P(R_{it} = r_t, V_{it} = v_t | R_{i,t-1} = r_{t-1}, V_{i,t-1} = v_{t-1}) \\
& \times \prod_{t=1}^{T} P(Y^p_{it} = y^p_{it}, Y^s_{it} = y^s_{it} | R_{it} = r_{it}, V_{it} = v_{it}) \Big] \\
= {} & \Pi \cdot A_{i1} \cdot Q_{i,1\to 2} \cdot A_{i2} \ldots \cdot Q_{i,T-1\to T} \cdot A_{iT} \cdot 1',
\end{aligned}
$$

where 1' is an N^2 x 1 vector of ones and A_{it} is a diagonal matrix of the state-dependent usage behaviors:

$$
A_{it} = \begin{bmatrix}
a_{it11} & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & a_{it12} & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & 0 & \ldots & 0 \\
0 & 0 & 0 & a_{it1N} & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & a_{it21} & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \ldots & a_{itNN}
\end{bmatrix}. \tag{1}
$$

1.2 Samplers and Prior Specification

We assumed standard non-informative priors for all of the parameters. Suppose random effect parameters $\Theta_i = \{\theta^p_{irv}, \theta^s_{irv}\}$ for individual $i \in \{1, ...I\}$, and population level parameters $\Psi = \{\beta^p_r, \beta^s_s, \lambda^c_c, \pi^p, \pi^s\}$ and $\Delta = \{\delta^p, \delta^s\}$.

For each of the following parameters, at each MCMC iteration k, we first sample a new parameter draw using a random walk Metropolis algorithm whereby we sample the candidate parameters using a multivariate normal. We chose the random walk Metropolis approach due to the calculation simplicity of the Metropolis criterion (no Hastings adjustment), and therefore facilitated computation speed at each iteration. For more complex models, an adjusted Metropolis algorithm with global adaptive scaling, such as one found in Andrieu and Thoms (2008), could be used to actively tune the covariance jump matrix to achieve convergence with smaller numbers of iterations.

Step 1: Sample Θ^*_i

Step 2: Generate Δ^* and Σ^*_Θ

Once the latest draws of Θ^*_i are obtained for all customers, we can use them to generate the latest draws of Δ^* and Σ^*_Θ. Δ^* and Σ^*_Θ can be generated as the result of a Bayesian multivariate normal regression, and could be estimated using the R package *bayesm*. Although the current specification is the same as in Netzer et al. (2008), if one were to incorporate a different set of covariates W^p_i and W^s_i on the parameters $\theta^p_i = \{\theta^p_{i11}, \theta^p_{i21}, \theta^p_{i22}, \theta^p_{i31}, \theta^p_{i32}\}$ and $\theta^s_i = \{\theta^s_{i11}, \theta^s_{i21}, \theta^s_{i22}, \theta^s_{i31}, \theta^s_{i32}\}$, the procedure in this step would become a multivariate Seemingly Unrelated Regression (SUR). More details on prior specification can be found in Rossi et al. (2005).

Step 3: Sample Ψ^*

1.3 Identification

We separate the discussion of identification of parameters into the three parts of an HMM: transition probability matrix, initial state distribution, and state-dependent distributions. For simplicity of exposition, we focus our discussion on one form of usage, referred to hereafter as simply "usage," with the implication that the argument could be applied to personal or social usage interchangeably. We expand this back to both forms of usages during our discussion of the covariance parameters λ^c.

Latent States and Transition Probabilities

We begin with the identification of the latent states, which is required before we can identify the parameters Θ_i, Δ, and Σ_Θ. For the identification of states, we use two sources of variation, one at the individual level and the other at the population level. The first variation is from individual-level customer usage across weeks. Given that a customer exhibits low, medium, and high usage regimes in time series usage data, the differences in average usage for each regime helps to identify the clusters of passive, semi-active, and active regimes in customer usage. It is helpful to think about consecutive periods of usage with similar magnitudes as a "class" in a latent class model, but in our case the customer is allowed to move from one class to another across time. Therefore, the larger the difference in the usage means of the regimes, the easier it is for the model to identify the boundaries of, and subsequently the transitions between, regimes. Given the number of transitions, we could identify the transition probabilities. The more switches from one regime to another, the higher the transition probability of moving from one state to another, and conversely the probability of a customer staying in one particular state. Given this, there is a baseline in transition probability for each customer.

Applying this argument to the population, the boundaries are normalized by the population level boundaries of "active," "semi-active," and "passive" regimes across all customers, and therefore, the second source of variation comes from the overall shifts in usage regimes across the entire population. Given that each customer has a baseline transition probability, the variation across transition probabilities across the entire population can help us pinpoint the location of a customer's baseline probability as compared to the population distribution. For instance, if a customer has a higher transition probability of moving from the passive to semi-active state than the population probability average, the individual will have a low threshold parameter of θ_{i11}. If the customer has a lower transition probability than the population average, θ_{i11} would be high, leading to a higher probability of staying in the passive state. The argument for identifying Δ and Σ_Θ is similar to that of a multivariate normal regression. To identify Δ, we need variations in the hierarchical design matrix *across* customers. For example, taking the binary variable "PR Professionals," we need enough customers who belong to this group (1), and enough customers not belonging to this group (0), to ensure close to equal quantities of 1's and 0's. Essentially, what the Δ parameter for this variable picks up is the average of Θ_i's for customers who belong to the "PR Professionals" group. This argument can be extended to all of the other variables in the hierarchical design matrix in a similar fashion to a multiple regression. The Σ_Θ matrix is identified from the covariation of the individual elements in the Θ_i vector. If the co-occurrences of two of the variables is very high across the population, then the covariance terms for Σ_Θ will be high as well.

Initial State Probabilities

To identify all elements of the vector Π, the initial state probabilities, we follow an argument similar to that for identification of the transition probabilities. Note that the initial state probabilities are estimated at the population level. Therefore, starting at the individual level, the usage level at the initial period in comparison with the rest of the usage level can give clues as to whether a regime shift has occurred between periods one and two. For example, suppose that for periods two through five the estimated active state probability is very high, and the initial state usage level is dramatically lower than the average of periods two to five. Then the customer is most likely to be in a passive state in the first period, and therefore have a high estimate of passive state probability. This argument can be applied to all individuals, thus obtaining the initial state probability across all individuals. The average of the initial state probabilities across all individuals will give us the population level initial state probabilities, thereby giving us Π. In our paper, predicting the initial state likelihood of new individuals is not our focus, therefore, we have chosen to estimate Π at the population level for parsimony. However, it is possible to do so given panel data. For more considerations of estimating initial state probabilities, please refer to MacDonald and Zucchini (1997). The posterior means and HPDs of the initial state probabilities for the full model, for both data sets, can be found in Tables 2 and 3.[1]

---

[1] To obtain the set of nine initial state probabilities $\{\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{31}, \pi_{32}, \pi_{33}\}$, simply multiply the probabilities according to the passive, semi-active, or active state counterparts (π_r^p,π_v^s) in Tables 2 and 3. For instance, π_{12} = P(Passive Personal)·P(Semi-Active Social)=(0.35)(0.25) =0.09, for the annotation data set. Note that due to the independece assumption we make between the two latent states, there are only a total of $(N-1)+(N-1)$ parameters to be estimated. Therefore, for the $N = 3$ model, we estimated the parameters π_1^p, π_2^p, π_1^s and π_2^s, and we can obtain π_3^p and π_3^s given that $\Sigma_{k=1}^3 \pi_k^p = \Sigma_{k=1}^3 \pi_k^s = 1$.

State-Dependent Distributions

Given our multivariate Poisson distribution assumption, the most important thing to identify are the three means of the bivariate Poisson distribution: λ^p, λ^s, and λ^c. Given that we can identify the latent states, we can find the weekly usages, *averaged across time* periods, when a particular customer is in a passive, semi-active, and active states. For example, if a customer has high personal usage when in an active state, the λ^p_{active} will be higher than a scenario in which the same customer has lower personal usage average when in the same active states. This same argument can be applied to λ^s with social usage.

The λ^c variable captures the covariance between weekly personal and social usage. For a customer in a given state, say the active state, if high personal usage weeks are associated with high social usage, then λ^c_{active} would be high as well. However, if personal usage and social usage do not move together, then λ^c would be very low or zero. Such an estimate of λ^c would simply indicate that personal and social usage are independent for the particular given state. This argument applies to the nine potential personal-social usage state configurations (personal state, social state) $\in$ {(passive, passive), (passive, semi-active), (passive, active), (semi-active, passive), (semi-active, semi-active), (semi-active, active), (active, passive), (active, semi-active), (active, active)}.

Once the λ_{state} can be identified, the argument for identifying the β's is similar to that for a Poisson regression. β_{active} for *Inbound-Sharing* would be the mean effect of receiving an inbound message on that particular week's usage, conditioned on the fact that the customers is already in an active state that week. In addition, this mean effect is averaged across all of the time periods in which the customer is identified to be in an active state.

1.4 Label Switching

As discussed in the sections from above, one of the main concerns with HMM estimation is the ability to identify each state, and given that HMMs can be thought of as a subclass of latent class models, it is also susceptible to the label-switching problem. Therefore, we place a restriction on $\tilde{\beta}_{0v} = \beta_{01} + \sum_{k=2}^{v} exp(\beta_{0k})$ in a similar fashion to Netzer et al. (2008) so that $\tilde{\beta}_{01} \leq \tilde{\beta}_{02} \leq \cdots \leq \tilde{\beta}_{0N}$. This ensures that the coefficients for the active state are identified from the higher usage activities than the passive state, which is identified from the lower usage activities.

Similarly, the thresholds are identified in this manner. Note that given a particular usage state j, say personal state thresholds θ^p_{ijr}, we have to ensure a similar order between destination states 1 and 2 so that state 1 refers to a lower personal usage state and state 2 refers to a higher personal usage state. In order to do so, define $\widetilde{\theta}^p_{ijv} = \theta^p_{ij1} + \sum_{k=2}^{v} \left\{ \exp(\theta^p_{ijk}) \right\}$. This ensures that the ordering of $\widetilde{\theta}^p_{ij1} \leq \widetilde{\theta}^p_{ij2} \leq \ldots \leq \widetilde{\theta}^p_{ijN-1}$. The same goes for the social state thresholds.

An additional challenge with the multivariate usage is how to prevent switching of the personal and social latent states. To address this, we find that using the $\tilde{\beta}$ and $\tilde{\theta}$ to specify the ordering in the state-dependent distributions and the transition probabilities is sufficient to ensure no label switching for personal or social usage states. In addition, we find that because the magnitudes of personal and social usage for the two data sets differ, this ensures the ordering of the usage states themselves. This is verified through the use of Monte Carlo simulations, in which we generate synthetic personal and social usage data in the neighborhood of our model estimates and were able to recover the true parameters through our estimation procedures.

1.5 Monte Carlo Simulations

In this section, we simulate data using the parameter estimates from both data sets to confirm that we could recover these parameter estimates in our model context. For simplicity, we conduct the simulation study to recover the baseline intercept and correlation parameters for the state-dependent distributions (β^p_{0r},β^s_{0v},λ^c_{rv}) and for all baseline threshold parameters for the transition probability matrix (θ^p_{jr},θ^s_{jv}).

In order to simulate data that mirrors our empirical conditions, we generate 100 customers for the average number of weeks in the observed data sets ($T = 24$ for the annotation data set, $T = 50$ for the cloud data set). We use the population average estimates of the threshold parameters θ^p_{ijr} and θ^s_{ijv} and the mean of the posteriors of β^p_{0r}, β^s_{0s}, λ^c_{rv} as true values to simulate data.

We display the mean and 95% HPD of the posterior distributions for each parameter in Table 4. Note that the true values for all parameters fall within the 95% HPD of the estimated posteriors. Thus, we were able to successfully recover the true posterior means of our data sets.

2 Alternative Model Specification

2.1 Univariate HMM

The univariate HMM we specified is a version of our full model with a univariate latent Markov chain. We assume that instead of taking on both personal and social latent states R_t and V_t, the univariate takes on a single latent state U_t. The Markov assumption is:

$$P(U_t|U_{t-1}, ..., U_1) = P(U_t|U_{t-1}).$$

Therefore, the log-likelihood can be specified as:

$$\begin{aligned} L_{iT} = & \; P(Y^p_{i1} = y^p_{i1}, ..., Y^p_{iT} = y^p_{iT}, Y^s_{i1} = y^s_{i1}, ..., Y^s_{iT} = y^s_{iT}) \\ = & \sum_{u_1=\{1,...,N\}} \cdots \sum_{u_T=\{1,...,N\}} [P(U_1 = u_1) \\ & \times \prod_{t=2}^{T} P(U_{it} = u_t|U_{i,t-1} = u_{t-1}) \times \prod_{t=1}^{T} P(Y^p_{it} = y^p_{it}, Y^s_{it} = y^s_{it}|U_{it} = u_t) \Bigg] . \end{aligned}$$

where the initial state probability vector is $\pi = [P(U_T = 1) \; ... \; P(U_T = N)]$, the transition probability matrix is specified as:

$$Q_{i,t-1\to t} = \begin{bmatrix} q_{it11} & q_{it12} & \cdots & q_{it1N} \\ q_{it21} & q_{it22} & \cdots & q_{it2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{itN1} & q_{itN2} & \cdots & q_{itNN} \end{bmatrix}, \qquad (2)$$

and q_{itjk}, Θ_i are defined in the manner described in Netzer et al. (2008). However, for the state-dependent distribution, we have:

$$a_{itu} = P(Y^p_{it}, Y^s_{it}| = u, \, \lambda^p_{iu}, \, \lambda^s_{iu}, \, \lambda^c_u) =$$

$$e^{-(\lambda^p_{iu}+\lambda^s_{iu}+\lambda^c_u)} \frac{(\lambda^p_{iu})^{Y^p_{it}}}{Y^p_{it}!} \frac{(\lambda^s_{iu})^{Y^s_{it}}}{Y^s_{it}!} \sum_{l=0}^{min(Y^p_{it},Y^s_{it})} \begin{pmatrix} Y^p_{it} \\ l \end{pmatrix} \begin{pmatrix} Y^s_{it} \\ l \end{pmatrix} l! \left(\frac{\lambda^c_u}{\lambda^p_{iu}\lambda^s_{iu}} \right)^l ,$$

where λ^p_{iu} and λ^s_{iu} are the means of the personal and social usages for customer i in state u, and λ^c_u is the parameter that captures the correlation between personal and social usage in state u. We can then incorporate covariates into the first two mean parameters in the following fashion:

$$\lambda_{iu}^{p} = exp(\beta_{u}^{p} \cdot X_i),\ \lambda_{iu}^{s} = exp(\beta_{u}^{s} \cdot X_i), \tag{3}$$

which leads to a state-dependent matrix A_{it}:

$$A_{it} = \begin{bmatrix} a_{it1} & 0 & \dots & 0 \\ 0 & a_{it2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & a_{itN} \end{bmatrix}. \tag{4}$$

Model with Equality Constraint on the State-Dependent Coefficients

An alternative model we estimated is one in which the number of states in Equation 2 is twice that of the N from the bivariate model from the main text. For example, in this alternative model there are a total of four states from the transition matrix of the bivariate model with $N = 2$, per Equation 3 from the paper:

$$Q_{i,t-1\to t} = \begin{bmatrix} q_{it1111} & q_{it1112} & q_{it1121} & q_{it1122} \\ q_{it1211} & q_{it1212} & q_{it1221} & q_{it1222} \\ q_{it2111} & q_{it2112} & q_{it2121} & q_{it2122} \\ q_{it2211} & q_{it2212} & q_{it2221} & q_{it2222} \end{bmatrix} \tag{5}$$

Recall that because the interpretation of the states above is such that each $q_{itjkj'k'}$ captures the product of the personal and social transition probabilities in a given time period, the corresponding state-dependent distributions a_{itrv}, replicated here from Equation 1 for $N = 2$, has the same λ_{ir}^{p}, β_{r}^{p},λ_{iv}^{s} , and β_{v}^{s} coefficients for states that have the same corresponding personal (r) or social (v) latent states. Specifically,

$$A_{it} = \begin{bmatrix} a_{it11} & 0 & 0 & 0 \\ 0 & a_{it12} & 0 & 0 \\ \vdots & \vdots & a_{it21} & 0 \\ 0 & 0 & 0 & a_{it22} \end{bmatrix}. \tag{6}$$

For example, quantities q_{it1111} and q_{it1121} would have the same social usage coefficients. This is because they are conceptually linked to quantities a_{it11} and a_{it21}, therefore leading to a bivariate

Poisson with parameters λ_{i1}^{s} and β_1^s, as both quantities involve the customer transitioning to a social passive state ($v = 1$), but having different personal usage coefficients. Similarly, quantities q_{it1111} and q_{it1112} would have the same λ_{i1}^{p} and β_1^p coefficients.

A similar, but not identical, way to capture this relationship using the univariate HMM would be to specify Equation 2 with $N = 4$ states, such as:

$$Q_{i,t-1\to t} = \begin{bmatrix} q_{it11} & q_{it12} & q_{it13} & q_{it14} \\ q_{it21} & q_{it22} & q_{it23} & q_{it24} \\ q_{it31} & q_{it32} & q_{it33} & q_{it34} \\ q_{it41} & q_{it42} & q_{it43} & q_{it44} \end{bmatrix} \tag{7}$$

If we were to define the states $(1, 2, 3, 4)$ to have the same interpretation as the bivariate model, such that $(1, 2, 3, 4) = ((r = 1, v = 1),(r = 1, v = 2),(r = 2, v = 1),\ (r = 2, v = 2))$, we could then use the definition of A_{it} from Equation 6. This would lead to the personal usage state-dependent coefficients being equal in states 1 and 2, and in states 3 and 4, and the social usage state-dependent coefficients being equal in states 1 and 3, and in states 2 and 4.[2]

In Tables 5 and 6, we compare the in- and out-of-sample fit of the bivariate HMM, univariate HMM, and univariate HMM with the aforementioned equality constraint on the state-dependent coefficients, at the two-state level. We find that although the equality constraint improves the fit of the univariate model ($N = 4$), compared to a univariate model without the equality constraint ($N = 2$), our proposed bivariate model still outperforms this model on all fit metrics in both in-sample and out-of-sample tests. In the second data set, the proposed bivariate model improvement over all univariate models is even more dramatic, as shown in Table 6. We see this pattern across all in- and out-of-sample fit metrics. In the next section, we explain the conceptual difference between the univariate and bivariate models, and outline substantive differences in the two models.

2.2 Differences between Multivariate and Univariate Specifications

Conceptual Difference

While the equality constraint on the state-dependent distributions from the previous section may allow the univariate model to employ the same A_{it} as the bivariate model, ultimately, the

---

[2] We thank anonymous reviewers for making this suggestion.

transition matrix $Q_{i,t-1\to t}$ is still different. Specifically, the bivariate model allows the personal and social usage processes to distinctly vary dynamically (per Equation 5 in the main text), thus allowing the model to capture heterogeneity in baseline personal and social transition probabilities via separate θ_i^p and θ_i^s parameters (per Equation 6 in the main text). For instance, the bivariate model would incorporate heterogeneity for the quantity q_{it1112} with the following product of two ordered logits:

$$q_{it1111} \quad = \quad \left(\frac{exp(\theta^p_{i11})}{1+exp(\theta^p_{i11})}\right)\cdot\left(\frac{exp(\theta^s_{i11})}{1+exp(\theta^s_{i11})}\right). \tag{8}$$

On the other hand, a univariate model, regardless of equality constraints, handles heterogeneity in baseline transition probabilities with one joint parameter θ_i. Thus, if we were to create a model with the corresponding comparison to q_{it1111}, even with an equality-constraint assumption, we would express the quantity q_{it11} as a single ordered logit:

$$q_{it11} \quad = \quad \left(\frac{exp(\theta_{i11})}{1+exp(\theta_{i11})}\right). \tag{9}$$

One can imagine that for data sets in which the latent usage states vary by usage process, the difference between Equations 8 and 9 would matter. Taking the bivariate model estimates, we use the filtering method from Netzer et al. (2008) to visualize the dynamic evolution of the personal and social usage state probabilities. We display the state probability trajectories for both data sets in Figures 5 and 6, and here we see that the state probabilities evolve differently from the latent usage process.

Note that although we did not include time-varying covariates in the transition probabilities in our model, this inclusion should accentuate the differences between Equations 8 and 9, potentially leading to a more drastic difference in the bivariate and univariate models.[3]

Next, we compare the differences in results between the univariate and bivariate HMM specification as proposed in the main body of the paper.

---

[3] We confirm this result in a separate model comparison with time-varying covariates included. Results are available upon request.

Differences in Estimation Results

In comparing the results between the bivariate and univariate HMM specifications, we find several notable differences. We now briefly compare our results from the main text with those from an alternative specification of the HMM, namely, a three-state model with univariate latent Markov chain. Recall that although a univariate model with the same (or more) states may be easier to estimate in terms of the data requirements, one will sacrifice flexibility and richness in modeling the latent behaviors as separate processes. Furthermore, the state-dependent effects may differ using such a model, and one must approach inference with caution knowing that the multivariate counterpart has superior in-sample and out-of-sample fit. For example, the three-state univariate model shows that Mass-Invite customers have higher social usage in the semi-active state, a result that is in contrast with that from the multivariate model. A manager would thus infer an inaccurate social usage comparison between WOM and Mass-Invite customers in this state. One potential reason why these two models may differ in the results they produce is that the univariate model restricts customers to be in a semi-active state *concurrently* for both personal and social processes, whereas in the multivariate model customers can *move flexibly* between the three personal latent states while being in the semi-active social state.

For the second data set, WOM customers exhibit less personal usage (in the semi-active and active states) than customers acquired through other routes in the univariate model, the opposite of what is uncovered in the multivariate model results. Furthermore, the univariate model masks the substitution effect of personal and social customer usage in the semi-active state induced by firm Tweets. We also find that the univariate model shows a positive impact of Blog Posts on social usage in the active state, whereas the effect is negative in the multivariate model. Given that the multivariate model performs better than the univariate model on all measures of in-sample and out-of-sample fit, this raises doubts regarding the face validity of the effects from the univariate analysis. The results for these models are available upon request.

Latent Class Models

The latent state model can be thought of a static version of an HMM, therefore, our formulation is nested within the univariate HMM. To be specific, we note two differences between the latent class model that we run and the univariate HMM. The first is that we assume the initial state probability to be individual specific, and we link the W_i covariates to the initial state probability instead. The second difference is that we assume the transition probability Q_{itjk} to be an identity matrix. This yields the reduced formulation of the following individual log-likelihood:

$$\begin{aligned} L_i = \; & P(Y^p_{i1} = y^p_{i1}, ..., Y^p_{iT} = y^p_{iT}, Y^s_{i1} = y^s_{i1}, ..., Y^s_{iT} = y^s_{iT}) \\ = & \sum_{u=\{1,...,K\}} P(U_i = u) \left[\prod_{t=1}^{T} P(Y^p_{it} = y^p_{it}, Y^s_{it} = y^s_{it} | U_i = u) \right]. \end{aligned}$$

Note that the latent state variable U_i is now time-invariant, and the parameterization for $P(U_i = u)$ and $P(Y^p_{it} = y^p_{it}, Y^s_{it} = y^s_{it} | U_i = u)$ follows the ordered Logit and bivariate Poisson as presented in the univariate HMM model.

3 Managerial Simulation

We now explain how we used our model estimates to calculate the three managerial simulations in Section 6.4 of the main text. The purpose of these three simulations are to (a) show how our proposed dynamic multivariate model can aid managers in decision making, and (b) demonstrate that a model that does not account for both dynamics and bivariate latent states can arrive at different, and potentially incorrect conclusions.

3.1 Potential Advertising Revenue

To calculate the potential advertising revenue, we first computed the posterior distributions of the state-dependent personal usages by applying parameter posterior estimates to Equation 8 from the main text and setting the binary variable Search to one and all other variables to zero. The expected state-dependent personal and social usages for all states by adoption route can be found in Table 4 of the main text. To account for the dynamic evolution of a typical customer journey, we assume that a customer begins in the active-active state, and calculate overall expected personal usage by simulating the expected weekly personal usage using each draw of the transition probability posterior distributions. Then, we simulate the evolution of personal usage for 52 weeks and multiply all periods by 0.6 cents. We calculate the average weekly revenue by averaging the 52-week revenue. Multiplying this weekly average by four yields a posterior distribution of monthly average advertising revenues. We then take the mean of this distribution, for each types of customers, to arrive at the monthly average advertising revenue figure of \$0.09 for Search-acquired customers and \$0.04 for WOM-acquired customers. Repeating this calculation for the estimates from the latent class model yields 7.68 and 2.78 annotations for Search-acquired and WOM-acquired customers, respectively. Thereby did we we arrive at the monthly revenue figures of \$0.18 and \$0.07.

To compute the break-even analysis for a search engine optimization campaign, we calculated the Search customer lifetime revenue figure by assuming a 10% annual churn and interest rate and using the posterior distribution of average monthly revenue (from above). The posterior mean of the customer lifetime revenue is calculated to be \$5.49 ($\text{Monthly Revenue} \cdot 12 \text{ months} \cdot \frac{0.90}{1+0.10-0.90}$), where the 0.9 and 0.1 reflect the 90% annual retention rate and 10% interest rate. Given this figure, and assuming that a potential campaign would cost \$500 a month, this would translate to a posteior mean break-even point of 1,094 Search customers, accounting for the lifetime revenue that each customer would bring in for an annual campaign ($\frac{\$500 \text{ per month} * 12 \text{ months}}{\text{Customer Lifetime Revenue}}$). Repeating this calculation for the posterior distributions from the latent class model yields a customer lifetime revenue of around \$10. Therefore, the breakeven point is now only at 617 customers under the latent class model estimates, a reasonable goal for an agency that has a track record of acquiring 600 customers.

3.2 Potential Premium Upgrade Revenue

To approximate the probability of an upgrade, we first use our estimates of the expected state-dependent usages and transition probabilities to calculate the expected number of files synced in a year in a similar fashion to the previous analysis. This leads to the 1,821 and 1,725 files synced figures as sums of personal and social usage for WOM-referred and non-WOM-referred customers, respectively. Therefore, we find that WOM-referred customers have a 6% higher usage propensity. Assuming that this usage percent difference translates to a 6% higher upgrade propensity, and using a 3% upgrade rate and \$10 per month upgrade revenue, we calculate that this 6% difference amounts to an annual difference of \$2,160 for 10,000 customers (6% upgrade difference * 3% upgrade rate * \$10 * 12 months * 10,000 customers = \$2,160).

We also calculate the cost of a referral program by approximating the cost of the additional storage expenses that would be incurred. Given that the firm grants a user 250MB for every WOM customer they successfully bring into the system, and using the median cost figure from Amazon's S3 storage service of \$0.02875/GB, we arrive at a cost of around 8 cents per referral per year. Thus, in a population of 10,000 WOM-acquired customers, the cost of this referral program would be \$863 a year (10,000 people * 0.25 GB * \$0.02875GB/month * 12 months). We also have to take into account that each joiner through this referral program would get 250MB as well, therefore the total cost would double, leading to a total of \$1,726. Under this calculation, the cost of the referral program would be offset by the additional revenue of \$2,160 alone.

For comparison, we also calculate the predicted personal and social usage from the three-state latent class model. To do so, we first calculate the predicted WOM and non-WOM customer expected usage levels, which turn out to be 1,360 and 1,814 for the year, respectively. This model would value the non-WOM customers higher than the WOM customers, and therefore it would not make sense to launch the referral program, as the cost of the program would far exceed the revenue gained.

3.3 The Value of Social Media Effort

To calculate the value of a Tweet, we want to capture the average effect of a Tweet over the course of a year during which a customer would transition through different states. We therefore simulate expected personal usage per week, over the course of a year, of the customer exposed to a single Tweet per week. This not only accounts for the customer's probability of transitioning from one state to another, but also incorporates the varying effect that Tweets have on a customer's personal usage in different states. We then average the calculated expected usage across the 52 weeks to arrive at a net increase of 0.41 files synced per week. Note that we make this calculation using the baseline group of customers, which are the non-WOM-referred customers. This 0.41 files synced translates to a yearly estimate of additional 21 files synced, which is a 1.2% increase compared to the 1,747 figure from the previous section ($\frac{0.41 \text{ files/week} * 52 \text{ weeks}}{1{,}747 \text{ files}} \approx 1.2\%$). Using the same logic as in the upgrade analysis in the previous section, this 1.2% difference in activity, when translated to upgrade propensity, equates to annual revenue of \$0.04 per customer (1.2% upgrade difference * 3% upgrade rate * \$10 per month * 12 months ≈ \$0.04). Note that in a yearly campaign, we allowed for a Tweet per week, therefore, if we were to divide the \$0.04 by 52 weeks, we would arrive at the 0.08 cents figure. We also calculated the cost of each entry to be \$50 ($\frac{\$100{,}000}{50 \text{ weeks}*40\text{hours}*1\text{entry/hour}}\approx\$50$).

As in the previous analysis, we can use the same set of calculations for the three-state latent class model. The latent class model would predict a 2.7% annual increase in usage, leading to a Tweet value of 0.19 cents ($\frac{2.7\% \text{ upgrade difference } *3\% \text{ upgrade rate } *\$10 \text{ per month } *12 \text{ months}}{52 \text{ weeks}}=\$0.0019$), more than twice of what the multivariate HMM would predict.

4 Checks for Dynamics

4.1 Run Test

As an additional test for dynamics, we conduct the Run Test from Frank (1962) to test for the existence of dynamics in both data sets. The purpose of the Run Test is to test if a customer's past usage choices have a persistent effect on the customer's current usage choice. In our application, we simplify the usage data (counts) into a binary choice of 1 if a customer has used the service in a given week and 0 if the customer did not use the service. A run is defined as a series of uninterrupted zeroes or ones. In this fashion, the Run Test tests whether the population distribution of customers' observed runs arises from pure random chance. If a distribution of customer usage behavior does not have dynamic effects, then the run distribution, after the proper transformations in the Run Test procedure, should follow a standard normal distribution. If dynamic effects exist for a majority of the customers, their expected number of runs should be low, therefore we will see an empirical run distribution that is more negatively skewed than the standard normal distribution. We compare the distribution of the personal and social usage runs for both data sets to the pdf of a standard normal distribution. We further present the QQNORM plots for both data sets in Figures 1 and 2. We see that for the first data set, all of the calculated deviates are below the 45% line; for the second data set, we see a majority of customers falling below the 45% line (73% of the deviates). Therefore, in conjunction with the in-sample and out-of-sample fit/prediction tests we performed, we conclude that there exists state-dependence for a non-trivial number of customers that we must account for in our two contexts.

4.2 Predicted Dynamic State Probabilities

In Figures 3 and 4, we display the predicted average semi-active and active state probability evolution across different adoption routes from our full model, calculated via the filtering method described in Netzer et al. (2008). Note that we observe that the probabilities vary dynamically across the different adoption routes, suggesting temporal patterns that model-free statistics would not be able to capture. For the Web annotation service, we see two noteworthy patterns. The first is that at 61% and 25% of customers from all routes are at least in a semi-active personal or social state in their first week of adoption, confirming the need to correctly estimate the initial state probability distribution (Figure 5). The second is that there are non-negligible fluctuations in the probability of being in the active state over time depending on the route of acquisition. For example, whereas WOM customers have a higher probability of being in the active state than customers from the other two acquisition routes in some weeks, there are other weeks in which customers from the other two routes are equally or more likely to be in the active states (e.g., weeks 3, 9, and 16 of Figure 3a). Similarly, dynamics are highly apparent when comparing Search and Mass-Invite acquired customers: there are weeks in which one overtakes the other and vice versa in terms of likelihood of being in an active state. This pattern could be seen in the semi-active states as well.

For the cloud-based file storage service, per Figure 4, we see an even more jagged pattern over time for likelihood of being in the active state: WOM customers overtake the baseline-acquired customers in certain weeks, and the reverse is true in other weeks.

We also see a similar difference when comparing the transition probabilities as broken down by personal and social latent states in Figures 5 and 6, signifying the need of a bivariate Markov chain specification.

4.3 Additional Data Checks

One potential concern with our data set is that we include weeks in the early days of both services. As a result, there may be adoption timing issues because early customers may have fewer opportunities to share and we might expect WOM as an acquisition route to emerge mainly later in the data. We conducted a number of analyses and tests to look into these matters. Primarily, we wanted to ensure that there are no significant differences in customers' social usage levels based on when they joined the service. We split the data into two parts: an “early period” covering those customers who adopted in the first half of the data, and a “later period” covering those who adopted in the second half. We examined whether the average first-half social usage is significantly lower than the average second-half social usage. A series of t-tests reveal that this is not the case for either data set. In addition, we examined how many people, on average, joined through WOM in the early vs. late periods. Once again, we find no significant differences across periods. We conclude that there do not appear to be major concerns with respect to social adoption and social usage issues based on timing. The results from both data sets are reported in Table 7.

References

Andrieu C, Thoms J (2008) A tutorial on adaptive mcmc. *Statistics and Computing* 18(4):343–373.

Frank RE (1962) Brand choice as a probability process. *The Journal of Business* 35(1):43–56.

MacDonald IL, Zucchini W (1997) *Hidden Markov and other models for discrete-valued time series*, volume 110 (CRC Press).

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

Rossi PE, Allenby GM, McCulloch RE (2005) *Bayesian statistics and marketing* (J. Wiley & Sons).

Table 1: Description and Interpretation of Parameters.

Parameter Set	Variables	Interpretation
	Initial State Dist.	
π^p, π^s	No Covariates	Initial State Probability
	State-Dependent Dist.	
β_r^p, β_s^s	Intercept	Baseline Usage
	WOM (Binary)	
	Mass-Invite (Binary)*	
	Search (Binary)*	
	Others Routes (Binary)**	
	Inbound Sharing (Personal Only)	Weekly Effect of Customer-to-Customer Communication
	Tweet	Weekly Effect of Twitter Communication
	Blog Post	Effect of Blog Communication (Within Week)
λ_{rs}^c	Intercept	Weekly Personal and Social Usage Covariance
	Transition Matrix	
θ_{irs}^p,θ_{irs}^s	Linked to δ^p and δ^s	Individual Level Baseline Transition Threshold
δ^p,δ^s	PR Professional (binary)*	Demographic Effects on Threshold Parameters
	Academics & Researchers (binary)*	
	Feedback When Joining*	
	Invited Others First Week	
	Search (Binary)*	
	Mass-Invite (Binary)*	
	WOM (Binary)	
	Other Routes (Binary)**	

*Denotes variables only in the Web-Annotation context. ** Denotes variables only in the Cloud-based file sharing context.

Table 2: Posterior of Initial Probabilities (Annotation Data Set)

	Passive	Semi-Active	Active
Personal (π^p)	35% (31%, 38%)	61% (57%, 64%)	4.8% (3.4%, 6.2%)
Social (π^s)	69% (64%, 75%)	25% (19%, 30%)	5.7% (2.2%, 9.5%)

Table 3: Posterior of Initial Probabilities (Cloud Data Set)

	Passive	Semi-Active	Active
Personal (π^p)	91% (89%, 92%)	6.7% (5.3%, 8.1%)	2.7% (1.8%, 3.6%)
Social (π^s)	88% (86%, 90%)	9.9% (8.3%, 12%)	2.3% (1.5%, 3.1%)

Table 4: Monte Carlo Recovery Simulation Estimates

Parameter	Annotation Data Set		Cloud Data Set	
	True Value	Estimated (95%HPD)	True Value	Estimated (95%HPD)
		State-Dependent Usage Parameters		
β_{01}^p	-3.4	-3.6 (-4, -3.2)	-0.20	-0.29 (-0.45, -0.13)
β_{02}^p	1.7	1.6 (1.4, 1.8)	1.60	1.6 (1.5, 1.7)
β_{03}^p	0.82	0.83 (0.75, 0.91)	0.68	0.68 (0.59, 0.76)
β_{01}^s	-7.9	-7.8 (-11, -4.2)	0.81	0.77 (0.7, 0.85)
β_{02}^s	-5.5	-5.3 (-10.8, -0.6)	1.25	1.24 (1.05, 1.44)
β_{03}^s	2.0	2.1 (0.9 , 3.2)	0.66	0.64 (0.55, 0.74)
λ_{11}^c	0.0003	0.0009 (0.00015, 0.0017)	0.08	0.12 (0.076, 0.17)
λ_{12}^c	0.03	0.03 (0.016, 0.04)	2.30	2.30 (0.98, 3.6)
λ_{13}^c	0.13	0.12 (0.01, 0.23)	3.20	3.27 (0.13, 6.47)
λ_{21}^c	0.03	0.027 (0.0014, 0.05)	1.43	1.5 (0.5, 2.6)
λ_{22}^c	1.4	1.5 (0.00064, 3.1)	13	13.4 (8.2, 18.8)
λ_{23}^c	0.02	0.02 (0.007, 0.025)	50	54 (44, 63)
λ_{31}^c	0.07	0.07 (0.02, 0.14)	1.2	1.35 (0.55, 2.2)
λ_{32}^c	2.9	3.4 (0.01, 6.7)	32	30 (17, 43)
λ_{33}^c	13.4	14 (11 , 17)	164	163 (136, 190)
		Transition Probability Baseline Parameters		
	True Value	Pop. Mean (95% HD)	True Value	Pop. Mean (95% HPD)
θ_{11}^p	4.4	4.5 (3.9, 5.2)	5.3	5.1 (3.5, 6.7)
θ_{21}^p	1.5	1.6 (0.57, 2.6)	1.9	1.9 (0.8, 3.1)
θ_{22}^p	0.93	0.98 (0.5, 1.5)	0.7	0.8 (0.52, 1.1)
θ_{31}^p	-0.03	-0.04 (-0.27, 0.20)	2.2	2.1 (0.9, 2.5)
θ_{32}^p	0.09	0.12 (-0.28, 0.51)	0.3	0.3 (-1.2, 1.7)
θ_{11}^s	5.9	6.1 (3.8, 8.4)	3.9	3.8 (3.1, 4.5)
θ_{21}^s	-1.2	-1.3 (-2.2, -0.4)	1.6	1.4 (0.3, 2.2)
θ_{22}^s	1.9	1.8 (0.8, 2.7)	0.6	0.5 (0.06, 0.70)
θ_{31}^s	1.7	1.6 (1.3, 1.9)	1.3	1.3 (0.9, 1.8)
θ_{32}^s	-2.1	-2.0 (-3.9, -0.11)	0.6	0.5 (-0.5, 1.5)

Table 5: Univariate versus Bivariate Model Comparison (Annotation Data Set)

	LMD	AICM	BICM	-DIC
Bivariate Markov Chain, HMM (N=2)	**-27,586**	**-55,531**	**-56,184**	**-55,413**
Univariate Markov Chain, HMM (N=2)	-28,291	-57,311	-58,485	-56,721
Univariate Markov Chain, HMM (N=4)†	-27,645	-56,812	-57,248	-55,503
Out-of-Sample				
Bivariate Markov Chain, HMM (N=2)	**-21,059**	**-42,514**	**-43,217**	**-79,078**
Univariate Markov Chain, HMM (N=2)	-21,633	-43,769	-44,561	-81,741
Univariate Markov Chain, HMM (N=4)†	-21,141	-42,837	-43,687	-79,427

*The more positive the number the better the fit. Best model in bold.

†With equality assumption on state-dependent distributions.

Table 6: Univariate versus Bivariate Model Comparison (Cloud Data Set)

	LMD	AICM	BICM	-DIC
Bivariate Markov Chain, HMM (N=2)	**-2,231,180**	**-4,464,299**	**-4,468,535**	**-4,463,558**
Univariate Markov Chain, HMM (N=2)	-3,012,548	-6,026,218	-6,028,808	-6,025,857
Univariate Markov Chain, HMM (N=4)†	-2,432,860	-4,865,573	-4,869,979	-4,864,962
Out-of-Sample				
Bivariate Markov Chain, HMM (N=2)	**-2,038,429**	**-4,078,824**	**-4,083,112**	**-7,740,220**
Univariate Markov Chain, HMM (N=2)	-2,743,816	-5,488,646	-5,490,995	-10,410,634
Univariate Markov Chain, HMM (N=4)†	-2,339,270	-4,679,374	-4,686,031	-8,880,445

*The more positive the number the better the fit. Best model in bold.

†With equality assumption on state-dependent distributions.

Figure 1: Run Test Quantiles for Annotation Data Set

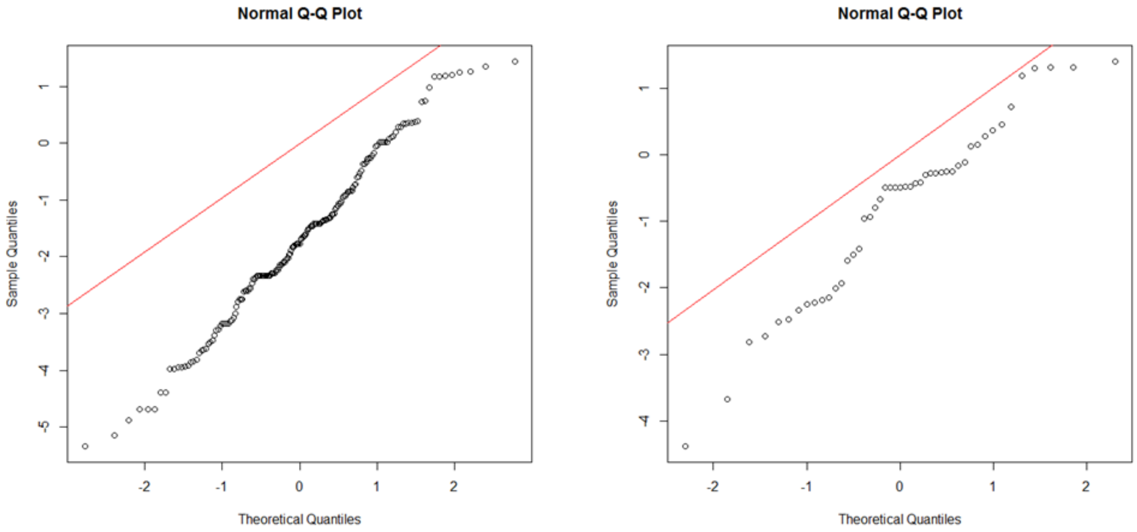


Figure 2: Run Test Quantiles for Cloud Data Set

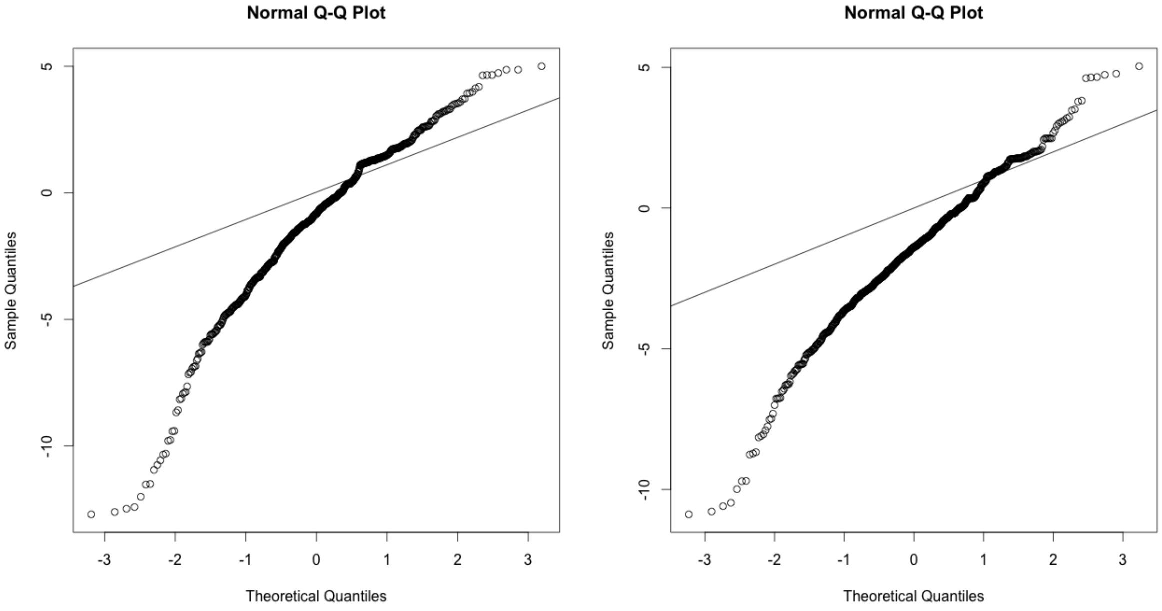
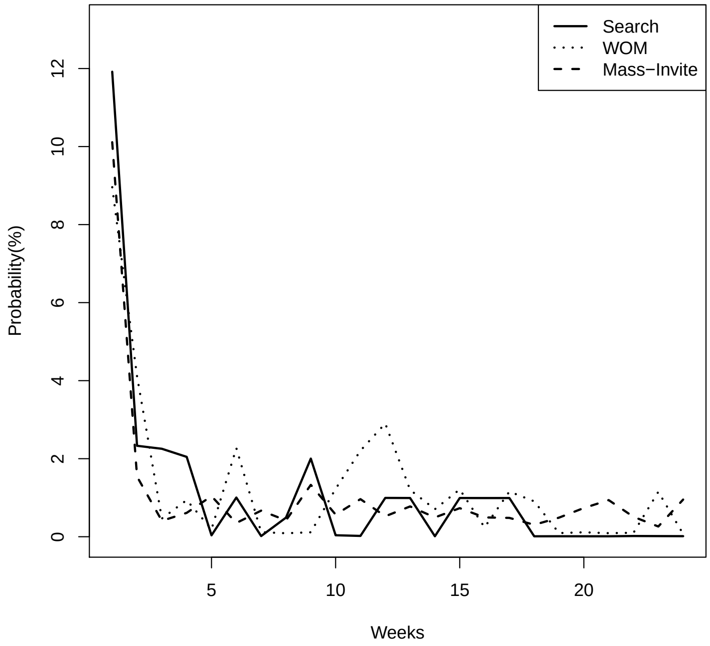


Figure 3: State Probabilities by Adoption Route (Annotation Data Set)

(a) Active State



(b) Semi-Active State

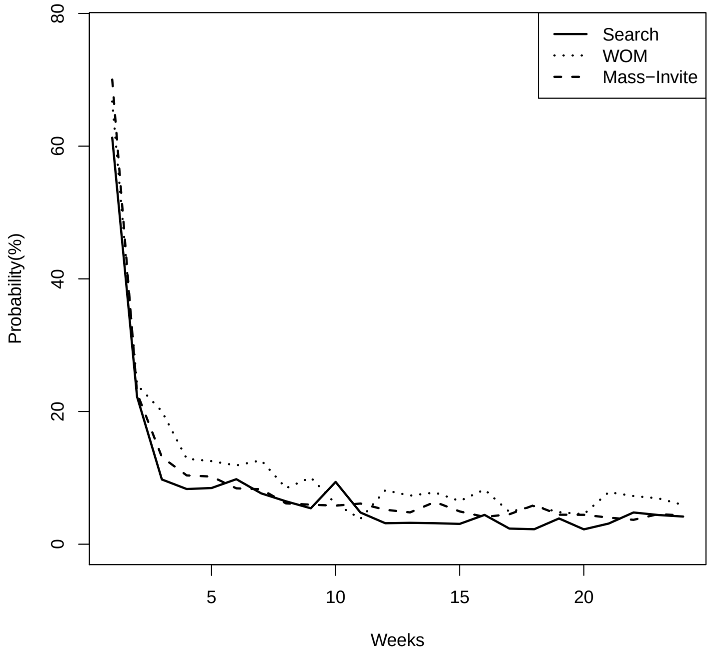
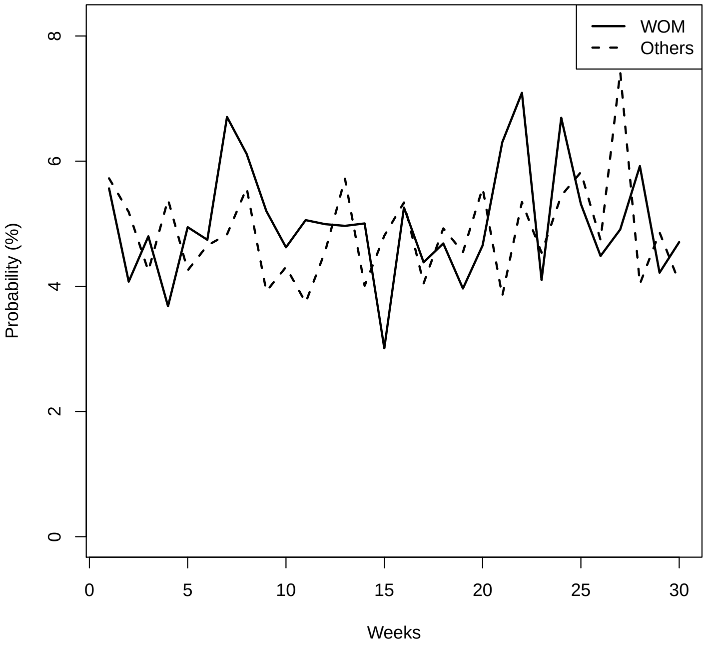


Figure 4: State Probabilities by Adoption Route (Cloud Data Set)

(a) Active State



(b) Semi-Active State

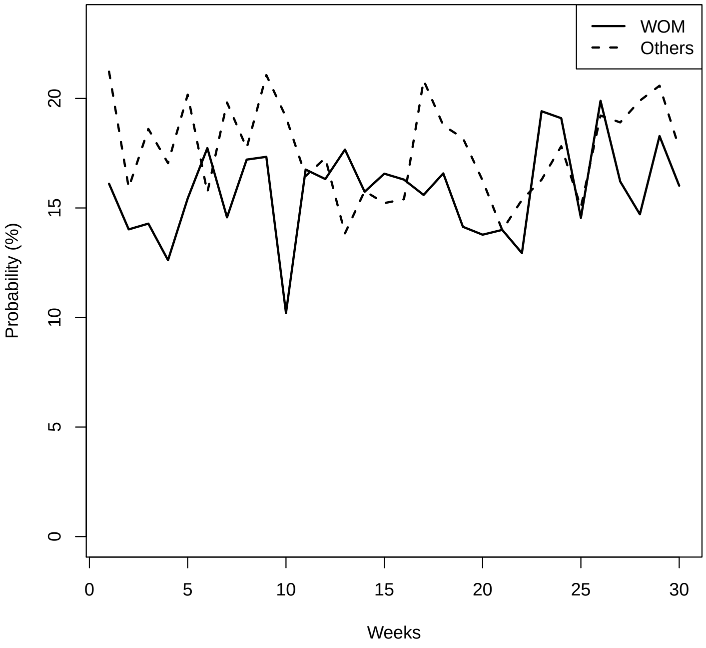
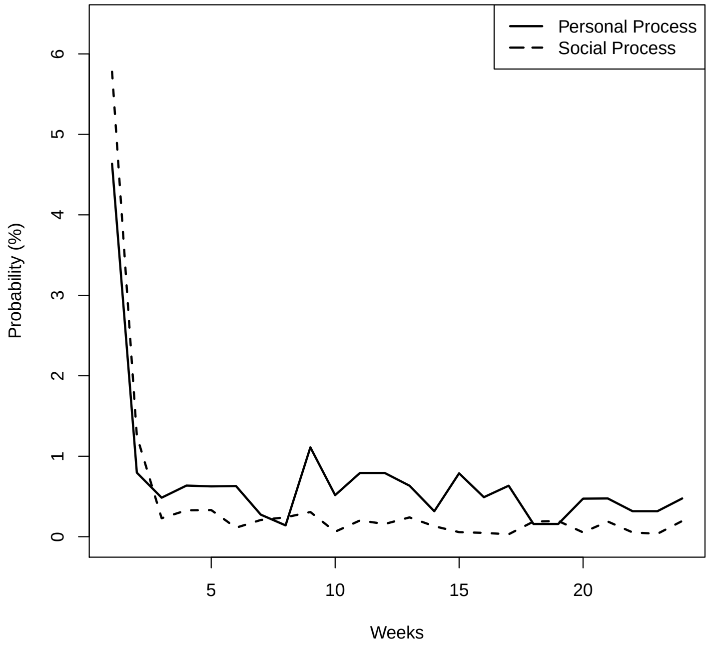


Figure 5: State Probabilities Comparison by Latent Process (Annotation Data Set)

(a) Active State



(b) Semi-Active State

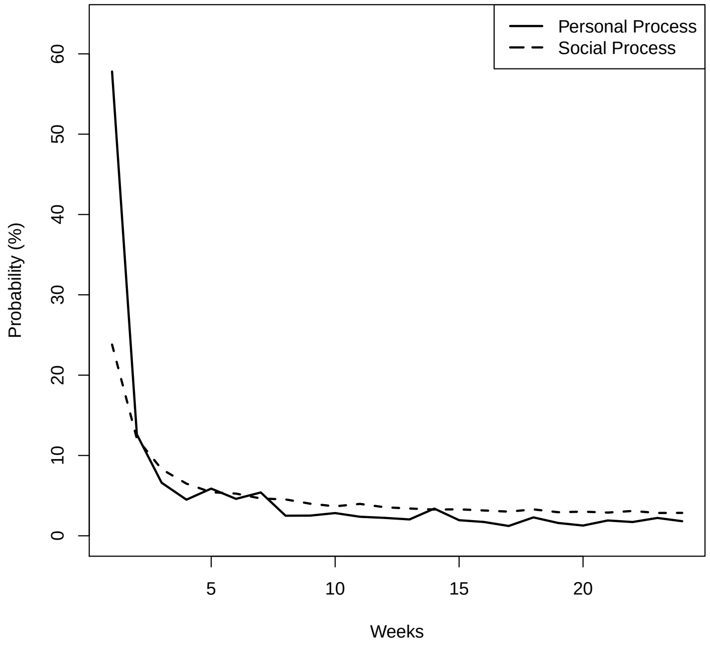
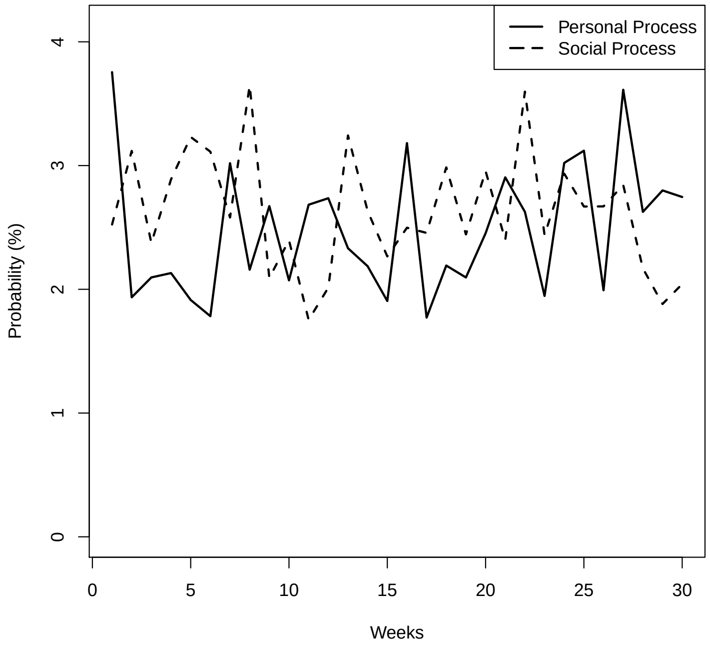


Figure 6: State Probabilities Comparison by Latent Process (Cloud Data Set)

(a) Active State



(b) Semi-Active State

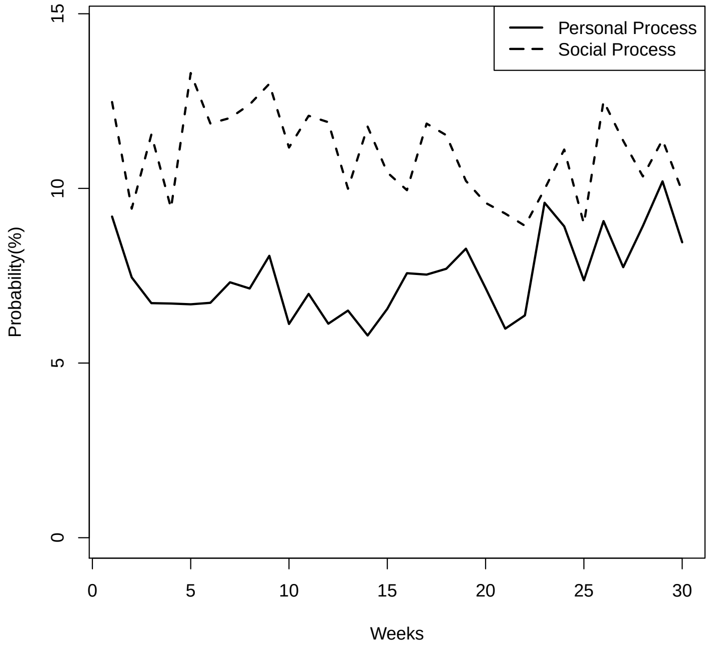


Table 7: Comparison of Customer Behavior by Time Period Joined

	First Half	Second Half	First Half	Second Half
	(Web Annotation	Data Set)	(Cloud Storage	Data Set)
Avg. Num. of Customers Joining Per Week from WOM	17.5 (56.17)	17.7 (10.71)	6.3 (13.14)	5.34 (11.06)
Avg. Personal Usage	0.927 (1.13)	0.880 (0.394)	110.84 (158.66)	129.49 (77.20)
Avg. Social Usage	0.0838 (0.126)	0.0247 (0.021)	59.18 (128.15)	79.36 (138.20)

*SD is denoted within parenthesis.