Heterogeneity in HMM:

Allowing for heterogeneity in the number of states

Nicolas Padilla[1]

Ricardo Montoya[2]

Oded Netzer[1]

[1] Columbia Business School, New York, NY

[2] University of Chile, Santiago, Chile

June 27th, 2019

Abstract

Hidden Markov Models (HMMs) have been widely used in marketing to study dynamics in customer behavior. HMMs have been successfully applied to model how customers transition among a set latent states such as attention levels, web search behavior, customer's relationships, and purchase intent. While most HMMs in marketing allow for heterogeneity in the model's parameters, these models assume that the number of latent states is common across customers. In this work, we analyze the potential bias of making such an assumption, assess to what extent heterogeneity in the model's parameters can mitigate the impact of such bias, and provide a mixture of HMMs model that relaxes this assumption.

Using a comprehensive Monte Carlo simulation exercise and secondary data from an online role playing game, we demonstrate that ignoring heterogeneity in the number of states could lead to model identification problems and to erroneous interpretations of customer dynamics. In particular, we show that: (1) even when only a small proportion of customers have a larger number of latent states (and most customers transition among fewer states), the best fitting model would be an expensive HMM in terms of number of states; (2) even when heterogeneity is accounted for in the HMM parameters, the inference from analyzing the population estimates, a common practice in the literature, can be biased; (3) even the individual-level estimates of customers with the correct number of states can be biased. We propose a mixture of HMMs with different number of states to account for heterogeneity in the number of states which captures well the behavior at both individual and population level.

Keywords: Hidden Markov Models (HMM); Heterogeneity; Hierarchical Bayesian Hamiltonian Monte Carlo.

1 Introduction

Hidden Markov Models (HMMs) have been used in various domains to model how a sequence of observations is governed by transitions among a set of latent states. These include domains such as speech or word recognition (Rabiner, 1989), image recognition (Yamato et al., 1992), economics (Hamilton, 1989, 2008), finance (Mamon and Elliott, 2007), genetics (Eddy, 1998), and earth studies (Hughes and Guttorp, 1994). Over the last decade the number of applications of HMMs in marketing has grown substantially. HMMs in marketing have been primarily used to model how customers (and sometimes firms) transition among a set of latent states over time. In the context of customers' behavior, the latent states could be attention states (Liechty et al., 2003; Wedel et al., 2008), the relationship between the customer and the firm (Netzer et al., 2008; Ascarza and Hardie, 2013; Romero et al., 2013) purchase propensity (Schwartz et al., 2014), channel migration (Mark et al., 2013), internet browsing behavior (Montgomery et al., 2004; Stüttgen et al., 2012), consumers' choice among a portfolio of products (Paas et al., 2007; Schweidel et al., 2011), purchase cycles states (Park and Gupta, 2011), latent behavioral learning strategies (Ansari et al., 2012) and households lifecycle stages (Du and Kamakura, 2006). HMMs have been also used to capture how marketing actions could affect the transition among states (Netzer et al., 2008; Montoya et al., 2010; Kumar et al., 2011; Zhang et al., 2014). In some marketing applications the unit of analysis does not involve consumers. For example, Ebbes et al. (2010) looked at how firms' (banks') competitive landscape changed over time. Moon et al. (2007) used a HMM to uncover firms' latent competitive promotions. Lemmens et al. (2012) looked at evolving segments of countries. Montoya and Gonzalez (in press) looked at product sales for multiple SKUs to identify on-shelf out-of-stocks.

Thus, while HMMs reflect a fairly recent advance in the marketing literature, applications of HMMs in marketing are numerous and are rapidly growing. According to Netzer et al. (2017), there are over 30 published papers leveraging HMMs in the marketing literature. See Appendix A, adapted from Netzer et al. (2017), for a selected list of HMMs in marketing.

One of the main differences between the applications of HMMs in marketing versus other disciplines is that in most applications in fields other than marketing, HMMs are used to study a single unit of analysis. That is, the data comprise one (often long) time series that is used to infer the state of the system at any point in time (e.g., GNP of the US from 1951 to 1984 to estimate the latent state of recession from Hamilton, 1989). In marketing, in contrast, HMMs are commonly estimated across multiple time series generated by heterogeneous agents (e.g. customers or firms).[1] Accounting for heterogeneity across customers has both statistical and

[1] To simplify exposition, we will use hereafter customers to refer to the agent generating the sequence of observations. All analyzes and discussions should follow through to other agents such as firms.

substantive reasons. Statistically, accounting for heterogeneity in the parameters across customers in dynamic models is crucial to properly disentangle heterogeneity from dynamics (Heckman, 1981). Failure to properly account for heterogeneity may lead to biased estimates and misleading conclusions. Substantively, marketers are particularly interested in understanding and exploiting heterogeneity across consumers (Fader, 2012).

Ideally, one would estimate a separate HMM per customer but due to limited number of observations per customer, researchers typically pool the information across customers and use random effect or latent class methods to account for heterogeneity. Indeed, of the 32 marketing papers that employ an HMM reported in Table 15, 20 accounted for unobserved heterogeneity in this form or another. Several approaches have been proposed to capture heterogeneity in the HMM parameters across time series (customers) such as Hierarchical Bayesian modeling (Netzer et al., 2008; Schwartz et al., 2014; Scott, 2002) and latent class (Schweidel and Knox, 2013). However, while marketing studies have accounted for heterogeneity in the HMM parameters, all these studies assume that the number of latent states is common across customers and use model selection criteria to determine the number of states for the entire customer base. That is, while customers can differ in how they transition among states or even in the way they behave given a state, marketing, and other fields, typically assume that all agents transition among the same number of states. We argue and demonstrate that the common practice in marketing of estimating one model with the same number of states for all customers, while flexibly accounting for heterogeneity in the model parameters, could lead to an identification problem and misleading insights.

Thus, the main objective of this research is to examine potential biases and identification problems introduced by assuming that customers have the same number of states and propose an approach to relax this assumption allowing different customers to have different number of HMM states.

To illustrate the identification problem, imagine a set of customers who are transitioning among low, medium and high levels of relationship with a firm, and another set of customers that only transition between low and medium levels of relationship. Estimating a common model for both groups with three HMM states can lead to a parameter identification problem for those customers who transition between only two states even when parameter heterogeneity is allowed. To see this intuitively, customers who only transition between two states but are represented by a model with three states could rationalize such third state in different ways. For instance, they could have a third state that they never visit. In that case the parameters representing the behavior in the third state could take any value, leading to an identification problem. Alternatively, these customers can have a third state that mimics the behavior of the first (low relationship) state and freely transition between the first and third, nearly identical, states, or have a third state that mimics the behavior of the second (medium relationship) state and freely transition between the second and third, nearly identical, states. These multiple model representations of the observed data constitute an identification problem. We further develop this intuitive example and elaborate on the identification problem in Section 2.

We first explore, using a simulation exercise, how severe the problem of ignoring heterogeneity in the number of states is, while accounting for heterogeneity in the HMM parameters. We confirm that ignoring heterogeneity in the number of states, when such heterogeneity exists, can lead to an identification problem, particularly for the customers who transition among fewer states. Furthermore, even when only 5% of the customers have a larger number of states, we find that various model selection criteria would recommend more expensive models with more states. Therefore, we may select the model that yields an identification problem and is the "wrong" model for the vast majority of the customers. Interestingly, we find biased estimates even for customers with the same number of states as the implied by the estimated HMM. This bias is caused by the identification problem of the parameters of the customers with fewer states, which shifts the estimates of the customers with the correct number of states through the population shrinkage. In addition, we find and demonstrate that accounting for heterogeneity in the HMM parameters but reporting insights from population level estimates, as is commonly done in the marketing literature (e.g., Netzer et al., 2008; Montoya et al., 2010; Kumar et al., 2011; Schweidel et al., 2011; Ansari et al., 2012), could lead to misleading insights about the dynamics in customer behavior.

We present a solution to this problem by proposing a mixture of HMMs (MHMM) that accounts for heterogeneity in the number of states. The model is able to identify different segments of customers with different number of states and captures well the behavior at both the individual and population levels.

We examine the limitations of estimating a HMM with the same number of states for all consumers and the performance of the MHMM using an empirical application to users engagement in a role playing online game. We find that the traditional HMM can lead to misleading inference confusing heterogeneity in the number of states for dynamics. On the other hand, the MHMM demonstrates not only superior predictive ability, but also richer insights with respect to the players' dynamics.

The rest of the paper is organized as follows. In Section 2, we outline a typical HMM and illustrate using a numerical example the identification problem that can arise when estimating a HMM with the same number of states for all customers. Section 3 describes several simulation experiments that show the conditions under which the assumption of homogeneity in the number of states can lead to identification problems. In Section 4 we propose a model that allows for different number of states across customers. In Section 5, we use data from an online role-playing game to examine the identification problem and illustrate the value obtained from accounting for heterogeneity in the number of states. We conclude in Section 6 with the discussion of the main contributions of this paper and directions for future research.

2 Heterogeneity in the number of states of HMM

Before further illustrating the implications of the assumption of common HMM states across customers, it is instructive to formally outline a typical HMM specification that we use in the remaining of this paper.

2.1 Model specification

We assume a customer i ($i = 1, \dots, I$), exhibits a particular behavior (Y_{it}, e.g., purchase) at each time period ($t = 1, \dots, T$). The customer behavior at each time period is governed by the customer's state in that period ($Z_{it} = s, \quad s = 1, \dots, S$).[2]

2.1.1 HMM components

An HMM can be defined by three components:

1. The initial state distribution: The probability that customer i is in state s at period 1 is $P(Z_{i1} = s) = \pi_{is}$.
2. The transitions: Customer i transitions from state s at time t to state s' at time $t+1$ with probability $P(Z_{it+1} = s' | Z_{it} = s) = q_{itss'}$.
3. The state dependent behavior: The probability of observing behavior $Y_{it} = y_{it}$ for customer i at time t given that she is in state s at time t is $P(Y_{it} = y_{it} | Z_{it} = s) = m_{it|s}(y_{it})$.

This state dependent behavior can be represented by any discrete or continuous probability distribution function. However, for the purpose of the simulation and empirical application presented in this paper, we assume that customer i makes N_{it} purchase decisions in each time period t, and chooses a focal product in y_{it} of them. Accordingly, throughout the paper we model the observed behavior y_{it} using a Binomial distribution with known total number of trials N_{it} (e.g., Montoya et al., 2010). The likelihood of purchasing at each occasion is affected by the customer's state membership at time period t, Z_{it}. Thus, we can write the state-dependent behavior as:

$$P(Y_{it} = y_{it} | Z_{it} = s) = m_{it|s}(y_{it}) = \binom{N_{it}}{y_{it}} p_{is}^{y_{it}} \cdot (1 - p_{is})^{N_{it} - y_{it}} \tag{1}$$

[2]Given that each customer may transition among different number of states, we describe the model for a customer with a general number of states S

where p_{is} is the probability of purchasing given that the customer i is at state s. We parametrize p_{is} to be increasing in the state labels ($p_{i1} \leqslant p_{i2} \leqslant \ldots \leqslant p_{iS}$) to avoid the label switching problem (See Appendix B.1).

In addition, we define the transition matrix Q_{it} that contains the transition probabilities between time $t-1$ and time t as:

$$Q_{it} = \begin{bmatrix} q_{it11} & \cdots & q_{it11} \\ \vdots & \ddots & \vdots \\ q_{itS1} & \cdots & q_{itSS} \end{bmatrix} \tag{2}$$

We parametrize each row of the transition matrix using the softmax function (See Appendix B.2).

2.2 Likelihood

Defining the diagonal matrix $M_{it} = \text{diag}\begin{pmatrix} m_{it|1} & \ldots & m_{it|S} \end{pmatrix}$, we integrate out the hidden states (Zucchini and MacDonald, 2009) to obtain,

$$L_i = p(Y_i|\theta_i, u) = \pi M_{i1} \left(\prod_{t=2}^{T_i} Q_{it} M_{it} \right) \mathbf{1}', \tag{3}$$

where $\mathbf{1}$ is a vector of ones of dimension S. The likelihood across consumers is given by: $L = \prod_{i=1}^{I} L_i$.

2.3 Numerical example

Similar to the stylized example briefly described in the introduction let assume a set of customers whose purchase behavior follows an HMM. One set of consumers (Segment A) transitions among three latent states of purchase propensity: Low, Medium and High. A second segment of customers transitions among only the low and medium states. We denote p_s in Equation (1) the purchase probability for each state s such that:

$$p^A = \begin{bmatrix} p_L & p_M & p_H \end{bmatrix} = \begin{bmatrix} 0.1 & 0.5 & 0.9 \end{bmatrix}$$

That is, when a customer is in the *Low*, *Medium*, or *High* state, she has a 10%, 50%, and 90% probability of purchasing, respectively. In addition, consumers in Segment A could transition among the three latent states according to the following transition matrix Q^A that describes the transition probabilities among states (see Equation 2).

$$Q^A = \begin{bmatrix} q_{LL} & q_{LM} & q_{LH} \\ q_{ML} & q_{MM} & q_{MH} \\ q_{HL} & q_{HM} & q_{HH} \end{bmatrix} = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.05 & 0.80 & 0.15 \\ 0.05 & 0.10 & 0.85 \end{bmatrix}$$

On the other hand, customers in Segment B only have two states: *Low* and *Medium*, with purchase probabilities:

$$p^B = \begin{bmatrix} p_L & p_M \end{bmatrix} = \begin{bmatrix} 0.1 & 0.5 \end{bmatrix}$$

Customers in Segment B transition between these two states according to the transition matrix Q^B:

$$Q^B = \begin{bmatrix} q_{LL} & q_{LM} \\ q_{ML} & q_{MM} \end{bmatrix} = \begin{bmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{bmatrix}$$

Because the researcher often does not observe a-priori which customer belongs to which segment, the traditional approach in the marketing literature has been to estimate a single HMM with the same number of states across customers but with heterogeneity in the HMM parameters. Suppose the selected model (using some model selection criteria) corresponds to the three-state HMM.[3] This model constraints all customers to have three HMM states, but it allows customers to have different values for p and Q. If the researcher allows for flexible heterogeneity structure, the HMM with three states should capture correctly the individual parameters of customers in Segment A (we will see later that this is not always the case). However, it is not clear how the three-state HMM will capture the behavior of consumers in Segment B with only 2 states.

One option for the model to rationalize the behavior of customers in Segment B is to estimate for these customers the same purchase probabilities ($\widehat{p}^B$) as those of the customers in Segment A,

$$\widehat{p}^B = \begin{bmatrix} 0.1 & 0.5 & 0.9 \end{bmatrix}$$

but estimate a transition matrix ($\widehat{Q}^B$) such that customers in Segment B never transition to the high state,

[3]As we will show in Section 3, model selection criteria recommend choosing the more expensive model (i.e., the model with more states) even when only a small fraction of the customers transition among a larger number of states.

$$\widehat{Q}^B = \begin{bmatrix} 0.85 & 0.15 & 0 \\ 0.15 & 0.85 & 0 \\ \hat{q}_{HL} & \hat{q}_{HM} & \hat{q}_{HH} \end{bmatrix}$$

The pitfall of this specification is that the parameters of the last row of the transition matrix are unidentified. That is, because Segment B customers never visit the High state, any set of values for $\hat{q}_{HL}$, $\hat{q}_{HM}$ and $\hat{q}_{HH}$ that are between 0 and 1 and sum to one would rationalize the data. For example, both these transition matrices

$$\widehat{Q}^B = \begin{bmatrix} 0.85 & 0.15 & 0 \\ 0.15 & 0.85 & 0 \\ 0.05 & 0.10 & 0.85 \end{bmatrix} \text{ and } \widehat{Q}^B = \begin{bmatrix} 0.85 & 0.15 & 0 \\ 0.15 & 0.85 & 0 \\ 0.33 & 0.33 & 0.33 \end{bmatrix}$$

can be valid estimates from the model. Similarly, the value of $p_H = 0.9$ is not identified because any purchase probability of this state would be consistent with the data.

Alternatively, the model could rationalize the behavior of Segment B customers differently. The model could identify that all states are reachable by these customers; however, there are two states that capture essentially the same purchase propensity. For example, in the estimated model the first state can have a purchase probability of 10%, and two states with purchase probability of roughly 50%. Specifically:

$$\widehat{p}^B = \begin{bmatrix} 0.1 & 0.5 - \varepsilon & 0.5 + \varepsilon \end{bmatrix}$$

$$\widehat{Q}^B = \begin{bmatrix} 0.85 & \hat{q}_{LM} & \hat{q}_{LH} \\ \sim 0.15 & \hat{q}_{MM} & \hat{q}_{MH} \\ \sim 0.15 & \hat{q}_{HM} & \hat{q}_{HH} \end{bmatrix}$$

In such a case, $\hat{q}_{LL}$ is estimated to be 0.85 and $\hat{q}_{ML} = \hat{q}_{HL} \approx 0.15$, but the customer could move in any manner between Medium and High states, which capture nearly identical behavior, leading again to an identification problem. For example, both the transition matrices below can be valid estimates from the model.

$$\widehat{Q}^B = \begin{bmatrix} 0.85 & 0.075 & 0.075 \\ 0.15 & 0.7 & 0.15 \\ 0.15 & 0.15 & 0.7 \end{bmatrix} \text{ and } \widehat{Q}^B = \begin{bmatrix} 0.85 & 0.075 & 0.075 \\ 0.15 & 0.25 & 0.6 \\ 0.15 & 0.6 & 0.25 \end{bmatrix}$$

In sum, the model for the Segment B customers (with two states) is not identified when a 3 state HMM is estimated for all customers. Note that this identification problem cannot be corrected by collecting more observations per customer.

Because the researcher does not know which customers belong to Segment A or to Segment B, this lack of identification could be problematic when reporting aggregate estimates and when delivering insights from model estimates. For example, following the first identification example above, the population mean parameters of the last row of the transition matrix will be biased because they mix the correct estimates of Segment A and the unidentified estimates of Segment B. Similarly, the individual level estimates for Segment B customers, and any inference based on these estimates, are likely to be biased. In the next section, we use a simulation study to assess the degree of the identification problem illustrated by this example.

3 Simulation exercise

Because in typical empirical applications the researcher does not observe the "true" number of states that each customer transitions among, we design a simulation exercise, in which we know the true number of states for each customer to assess the implications of the assumption of homogeneity in number of HMM states. The objective of this simulation is to measure the potential impact of estimating an HMM that is homogeneous in the number of states when customers have different number of states.

We assume that there are two customer segments following the model described in the numerical example discussed in Section 2.3. The behavior of the first segment (Segment A) follows an HMM with three states (Low, Mid, and High) and the behavior of the second segment (Segment B) follows an HMM with two states (Low, and Mid). We simulate data for $I = 500$ customers and $T = 45$ time periods (25 periods for calibration, 10 for validation, and 10 for testing). At each time period the customer is making $N_{it} = 2$ purchase decisions following the state dependent behavior that corresponds to her state at that time period. We define $\lambda \in [0, 1]$ as the proportion of customers in Segment B (customers with two states). Consequently, $I \cdot (1 - \lambda)$ customers in Segment A transition among the Low, Mid, and High states, whereas $I \cdot \lambda$ customers in segment B transition only between the Low and the Mid states. Table 1 shows the corresponding initial, state-dependent, and transition probabilities used to simulate the behavior of each segment.

Table 1: Values for the HMM components used in the simulation exercise

	HMM component	2 states (λ)	3 states ($1-\lambda$)
π	Initial state probabilities	$\begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$
p	State dependent probabilities	$\begin{bmatrix} 0.1 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.1 & 0.5 & 0.9 \end{bmatrix}$
Q_i	Transition probabilities	$\begin{bmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.05 & 0.80 & 0.15 \\ 0.05 & 0.10 & 0.85 \end{bmatrix}$

Note that to better disentangle heterogeneity in the number of states from heterogeneity in the parameters we keep the state dependent probabilities of the data generation process homogeneous within each segment. However, we generate individual-level transition matrices using a multivariate Normal distribution on the corresponding unconstrained parameters.

3.1 Model estimation

We estimate an HMM with a Binomial state dependent distribution as described in Section 2.1 to recover the model's parameters. Throughout the paper we use a Hamiltonian Monte Carlo (HMC) procedure, particularly the No U-Turn Sampling algorithm (NUTS) available in Stan (Carpenter et al., 2016) to draw the model parameters from the posterior distribution (see Appendix C). We use a hierarchical specification to account for heterogeneity in the HMM transition probabilities. Because the researcher does not know a-priori the customer membership to each segment, she typically pools the data of all customers and estimates an HMM with heterogeneous parameters but with a common number of states across customers. For the purpose of this simulation exercise we specify state dependent parameters and initial state parameters to be homogeneous, while we allow transition matrices to be heterogeneous across customers. Consequently, the HMM has a set of individual parameters and a set of common parameters across customers. Specifically, we define θ_i as the vector of all individual level parameters, and Φ the vector of all homogeneous parameters. In addition, as common practice in the literature, we assume θ_i are independent and identically distributed Multivariate Normal with mean μ_θ and covariance matrix Σ_θ. See Appendix B for further details of the model specification, and prior distributions.

3.2 Results

3.2.1 Bias towards more states

The first step in estimating a HMM is to choose the number of states. It is common to estimate a HMM with varying number of states and choose the model that best fits the data based on some model selection criteria. Diverse model selection criteria are often used to select the number of states (assuming the same number of states for all customers).

Because in our simulation customers could have either two or three states, it is instructive to investigate how the simulated proportion of customers with two states, λ, affects the number of states chosen by the model selection criteria. Intuitively, the model selection should favor the model that best represents the true number of states for the majority of the customers.

We analyze the number of states recommended by alternative model selection criteria when the proportion of customers transitioning among three states (Segment A) and the proportion of customers transitioning among two states (Segment B) are equal ($\lambda = 0.5$). In Table 2, we show in-sample log-likelihood (LL), log marginal density (LMD), Watanabe-Akaike information criterion (WAIC), as well as log likelihood in the validation sample. Because in-sample likelihood-based measures such as LMD or WAIC may not sufficiently penalize for additional parameters, we use the log posterior mean likelihood on the validation sample to select the appropriate number of states.[4] As can be seen in in the second row in Table 2, both penalized model selection criteria (LMD and WAIC) favor the HMM with 3 states. The predictive Validation LL also favors a 3-states HMM. Thus, the model selection criteria suggest choosing the specification that fits better the customers with the larger number of states (more complex behavior).

An interesting question is how many states would the model selection criteria recommend when the majority of the customers belong to Segment B? Would the model selection criteria recommend the simpler model with only two states that fits the majority of the customers? To investigate this issue, we vary the proportion of Segment B customers (λ) from 50% to 98% and calculate both in-sample likelihood based criteria as well

[4]We show in Appendix D.1 that when all customers have the same number of states, in-sample likelihood-based model selection criteria such as LMD and WAIC tend to choose a model with more states than simulated.

Table 2: Number of states selection for $\lambda = 50\%$ of customers with 2 states. Note: The best model in each column is in bold

	In sample			**Validation**
Number of States	LL	LMD	WAIC	LL
1	-14234.93	-14238.17	28478.65	-3980.95
2	-11855.63	-12006.04	24186.30	-3339.71
3	-11612.95	**-11850.43**	**23805.67**	**-3290.16**
4	**-11576.55**	-11865.49	23822.12	-3290.66

as log-likelihood for the validation sample for the two and three state HMMs (see Table 3).[5] Even when most customers truly transition among two states, most model selection criteria suggest a model with three states. For instance, when only 5% of the sample corresponds to customers who transition among three states, LMD, WAIC, and validation log-likelihood still suggest that the best model is the 3-states HMM. Thus, in such a scenario, the model selection criteria recommend estimating the incorrect and unidentified model for 95% of the customers. The main driver of this result is the ability of the 3-state HMM to explain the data for both segments, whereas the 2-state HMM fits poorly the behavior of Segment A. We note that this result is unlikely to be due to lack of penalization on the number of parameters as the recommended number of states is given by likelihood performance on periods out of the calibration sample.

Table 3: Number of states selection as a function of λ. Note: The best model in each column is in bold

	In sample			**Validation**
Number of States	LL	LMD	WAIC	LL
70% with 2 states				
2	-11771.93	-11921.59	23993.47	-3305.18
3	**-11580.13**	**-11782.59**	**23659.69**	**-3232.04**
80% with 2 states				
2	-11680.37	-11839.79	23797.24	-3306.58
3	**-11512.72**	**-11706.79**	**23504.87**	**-3257.65**
90% with 2 states				
2	-11579.15	-11726.82	23524.47	-3240.68
3	**-11484.34**	**-11647.77**	**23351.28**	**-3230.38**
95% with 2 states				
2	-11640.97	-11800.44	23643.98	-3376.78
3	**-11610.17**	**-11745.94**	**23545.50**	**-3367.03**
98% with 2 states				
2	-11451.27	-11599.23	23193.01	**-3256.54**
3	**-11407.52**	**-11587.73**	**23182.27**	-3259.50

[5] For completeness we also estimated a 1-state and 4-state HMMs but omit them from the table for clearer representation. Those specifications were not selected by the penalized model selection criteria.

3.2.2 Recovery of states

The numerical example in Section 2.3 suggests that if customers have different number of states then the parameters of an estimated HMM assuming the same number of states across customers may be unidentified. To investigate whether that is the case, we show in Table 4 the parameter estimates for the case with equal proportion of customers with 2 and 3 states ($\lambda = 50\%$). Specifically, in Table 4b we report the state dependent probabilities posterior mean and the 95% posterior interval, which are homogeneous across customers; and on Table 4c we report the posterior mean and the 95% posterior interval of the average of the transition matrix across the population. From the estimates of p, we conclude that the model recovers well the conditional behavior associated with the true three states with probabilities 0.1, 0.5, and 0.9. Given that the model does not allow for heterogeneity in p, it must be that the estimated model restricts customers with truly two states to rarely transition to the highest state, as stated in the numerical example.

In order to draw insights or the transition matrix, we need to consider that the model allows for heterogeneity in transition probabilities, and Table 4c shows the average across a population, which consists of two different segments. We discuss this further in the next section.

Table 4: 3-state HMM parameter estimates for $\lambda = 50\%$

(a) Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\pi_{\mathbf{s}}$	0.446	0.415	0.139
	[0.366 0.527]	[0.311 0.513]	[0.083 0.199]

(b) State dependent probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\mathbf{p}_{\mathbf{s}}$	0.112	0.497	0.892
	[0.096 0.127]	[0.461 0.531]	[0.870 0.914]

(c) Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
	1	0.836	0.147	0.017
		[0.797 0.870]	[0.106 0.190]	[0.004 0.039]
Q	**2**	0.135	0.734	0.132
		[0.101 0.173]	[0.674 0.788]	[0.085 0.179]
	3	0.045	0.318	0.636
		[0.025 0.077]	[0.136 0.473]	[0.492 0.803]

3.2.3 Interpreting aggregate estimates

Most marketing papers that use an HMM to represent customers' dynamic behavior report and generate insights from population estimates. See for example Table 4 in Netzer et al. (2008), Table 6 in Montoya et al. (2010) , Table 4 in Schweidel et al. (2011), Table 5 in Kumar et al. (2011), Table 5 in Ascarza and Hardie (2013),[6] Table 7 in Zhang et al. (2014) (see Figure 1). This common practice may yield misleading insights if different customers transitioned among different number of states.

[6]Note that Ascarza and Hardie (2013) report the average and 95% interval of individual posterior means instead of population means. However, inferences based on this table suffer from the same potential bias as in the other papers.

Figure 1: Sample of tables of population mean transition probabilities in marketing papers

(a) Netzer et al. (2008)

Table 4 The Mean Posterior Transition Matrices*

	No interactions		
	t		
t − 1	Dormant	Occasional	Active
Dormant	**90%** [89%–90%]	**10%** [10%–11%]	**0%** [— – —]
Occasional	**14%** [14%–15%]	**58%** [56%–59%]	**28%** [27%–30%]
Active	**3%** [3%–3%]	**29%** [28%–31%]	**68%** [66%–69%]

*95% confidence interval in parenthesis.

(b) Montoya et al. (2010)

Table 6 Posterior Means of the Transition Matrix Probabilities Across Physicians

No marketing activities			Detailing only			Sampling only		
0.75	0.25	0.00	0.62	0.38	0.00	0.70	0.30	0.00
0.17	0.78	0.05	0.16	0.79	0.05	0.13	0.81	0.06
0.15	0.46	0.39	0.15	0.45	0.40	0.10	0.41	0.49

Note. The detailing and sampling matrices are calculated assuming the firm allocates the average number of details and samples to each physician.

(c) Schweidel et al. (2011)

Table 4 Transition Probabilities

	To			
From	Full-size	Mid-size	Economy	End
	(a) Posterior means and 95% intervals for transition probabilities during promotion			
Full-size	0.48 (0.33, 0.69)	0.34 (0.25, 0.39)	0.16 (0.05, 0.30)	0.02 (0.00, 0.04)
Mid-size	0.18 (0.09, 0.31)	0.68 (0.56, 0.81)	0.12 (0.04, 0.29)	0.03 (0.00, 0.10)
Economy	0.04 (0.01, 0.11)	0.11 (0.03, 0.24)	0.83 (0.62, 0.97)	0.02 (0.00, 0.05)
End	0	0	0	1

(d) Kumar et al. (2011)

Table 5 Transition Probabilities in the Absence of Marketing Dollars

	To state 1 (%)	To state 2 (%)	To state 3 (%)
From state 1	94	5	1
From state 2	90	6	4
From state 3	95	3	2

(e) Ascarza et al. (2013)

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

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

(f) Zhang et al. (2014)

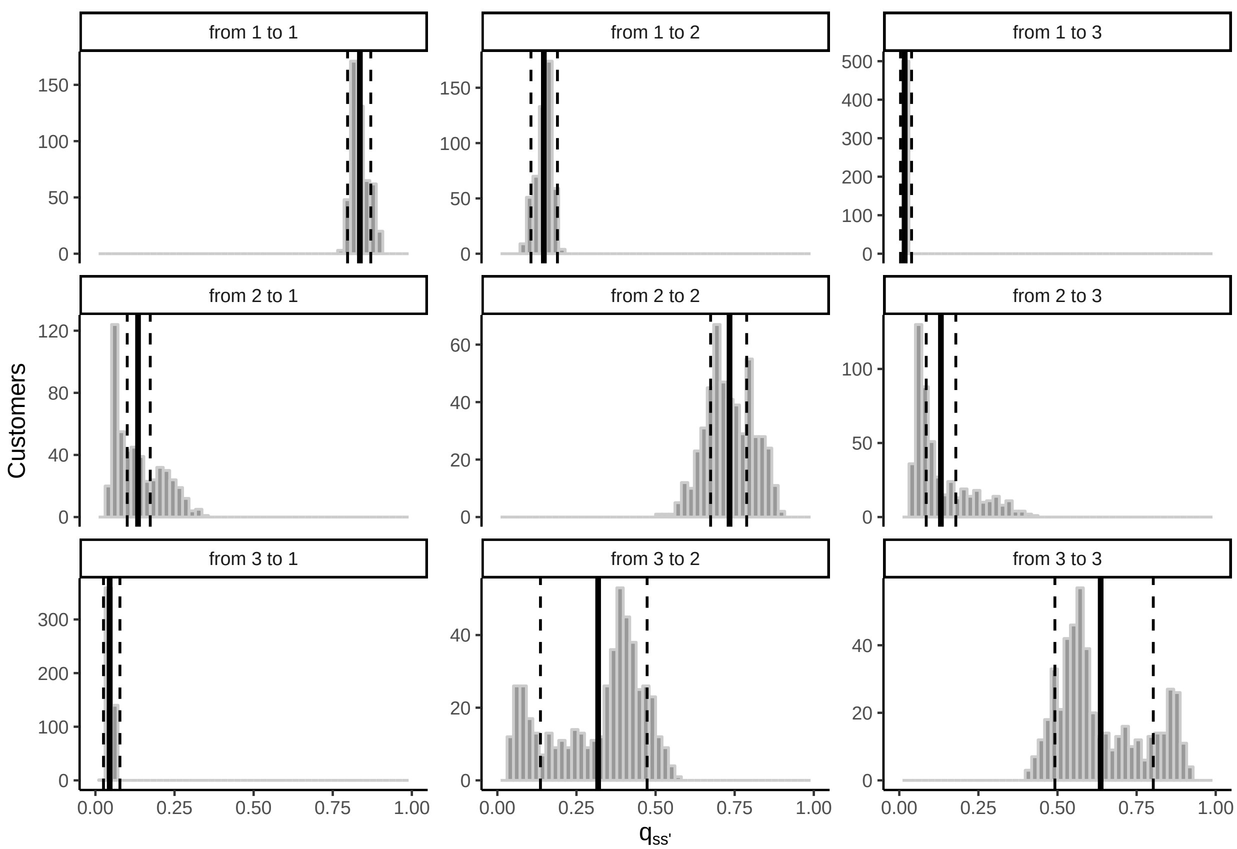
Table 7 Posterior Mean of the Transition Matrix Across Buyers

	10% price decrease		Average price (reference price)		10% price increase	
	Relaxed ($j+1$)	Vigilant ($j+1$)	Relaxed ($j+1$)	Vigilant ($j+1$)	Relaxed ($j+1$)	Vigilant ($j+1$)
Relaxed (j)	0.886	0.114	0.857	0.143	0.789	0.211
Vigilant (j)	0.095	0.905	0.077	0.923	0.043	0.957

Then, how would a population table look like when customers have different number of states? Table 4c shows the posterior mean and 95% posterior intervals of the average across the two segments. This average in Table 4c reflects the "composite" or weighted average of the transition matrix of customers with three states and the transition matrix of customers with two states who rarely transition to the third state. Both the estimated $\widehat{q}_{13}$ and $\widehat{q}_{23}$ are in between the simulated transition probability to the third state for customers with three states ($q_{13} = 0.05$, $q_{23} = 0.15$) and an hypothetical transition probability equal to zero for customers with only two states. As described earlier in our numerical example in Section 2.3, the parameters in the third row of the transition matrix (transitions from State 3) are unidentified for customers with only two states. Indeed, the corresponding aggregate estimates do not match those of the three states customers and particularly the probability of staying at state 3 is biased downwards.

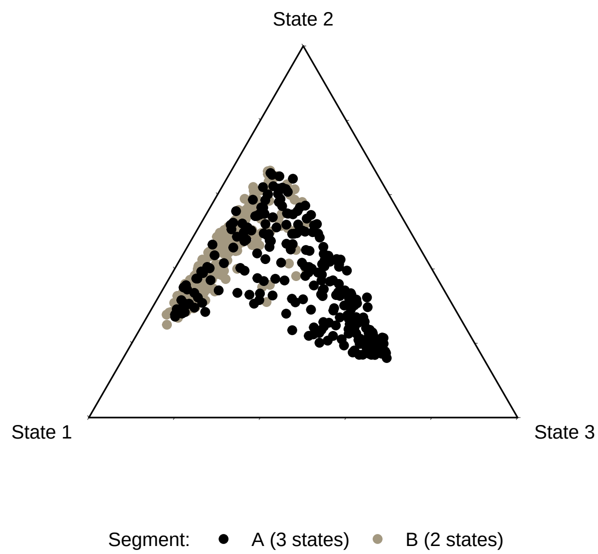
Thus, the population-based transition matrix in Table 4c (even though it was estimated with heterogeneity in transitions) represents neither the true corresponding values for customers with three states nor those of the customers with two states. As a result, one may make erroneous inferences from this aggregate transition matrix. For example, one may conclude that the focal firm is not doing a good job in keeping its most engaged customer active because the most profitable states (State 3 in our case), is less sticky than the two other states. To explore this aspect, in Figure 2 we report the heterogeneity of the transition probabilities. We observe that, opposite to what is suggested in the aggregated Table 4c, for those customers who visit the high probability of purchase state, the company is doing very well in keeping them in the high purchase state (State 3 is as sticky as the two other states for customers in Segment A), and we observe that the parameters are biased by a large group of customers who rarely visit that state.

Figure 2: Histogram of the individual posterior means of the transition probabilities



Another way of summarizing the heterogeneity in customers' dynamics is by computing the steady-state probabilities which represent the long-term behavior of each customer. Figure 3 illustrates this heterogeneity for the simulated sample. It shows that there is a group of customers who are more likely to transition among two states (mainly customers in segment B with relatively low estimated probability of being in state 3) whereas there are other customers who are more dynamic and are more likely to be transitioning among the three states.

Figure 3: Individual steady-state probabilities. Customers with 3 states are colored in black and customers with 2 states are colored in gray.



Note that the problem of interpretability of the aggregate result is likely to be even more severe when one accounts for heterogeneity in both Q and p in a continuous or discrete fashion, because the model would provide more flexibility for the unidentification problem to arise. We encourage researchers to closely look at and report the heterogeneity in the transition matrix parameters (as in Figure 2) and be cautious about interpreting the aggregate transition matrix when the heterogeneity is substantial. We further explore the bias in parameter estimates for the customers with the correct estimated model in Section 3.2.4.

3.2.4 Bias for "correctly" estimated customers

Arguably, the misleading insights drawn from aggregated transition matrices could be just a consequence of averaging two different populations, while indeed individual level estimates are unbiased for customers in segment A with truly 3 states. We show that this may not be the case. Indeed, we find evidence that the parameter estimates for customers with truly 3 states are also biased.

Because in the simulated data we know a-priori which customers belong to which segment, we can explore the estimated transition matrix, Q_i, averaging the segments with two and three states separately. We report the mean and 95% confidence intervals across draws from the corresponding posterior distributions for those averages for Segment A (3 states) and Segment B (2 states) customers in Table 5a and Table 5b, respectively.

Table 5: 3-state HMM transition probabilities by segment for $\lambda = 50\%$

(a) Segment A (3 states) transition matrix, mean and 95% intervals for the average across draws

		1	**2**	**3**
Q	**1**	0.832	0.151	0.018
		[0.781 0.872]	[0.105 0.202]	[0.004 0.041]
	2	0.095	0.717	0.188
		[0.063 0.135]	[0.640 0.795]	[0.110 0.260]
	3	0.050	0.223	0.726
		[0.030 0.076]	[0.120 0.312]	[0.644 0.820]

(b) Segment B (2 states) transition matrix, mean and 95% intervals for the average across draws

		1	**2**	**3**
Q	**1**	0.840	0.143	0.017
		[0.804 0.873]	[0.103 0.185]	[0.004 0.036]
	2	0.174	0.751	0.075
		[0.129 0.225]	[0.685 0.806]	[0.042 0.118]
	3	0.040	0.414	0.546
		[0.019 0.084]	[0.145 0.647]	[0.329 0.793]

First, as expected, we note from Table 5b that the third row of the transition matrix (transitions from State 3) for customers in Segment B seems to be unidentified. The estimated $\widehat{q}_{33}$ is significantly lower than the true value 0.85 (which drives the overall average in Table 4c downwards) and has a wide 95% posterior interval. This result is a consequence of the posterior distribution of q_{33} for Segment B customers being mainly driven by the shrinkage toward Segment A customers. Second, if we consider now the average transition matrix for Segment A customers in Table 5a, even though $\widehat{q}_{33}$ is higher than the corresponding $\widehat{q}_{33}$ for Segment B customers in Table 4c, this probability is also biased downwards (significantly lower than the true value 0.85) for the customers that truly have 3 states.

What would happen if the proportion of customers with 2 states increases? We expect that with more customers that have a lack of identification in their transition matrix, the bias in the parameters for the customers with 3 states increases. We show in Table 6 the average transition matrix for customers in Segment A (3 states) when the proportion of customers with 2 states is 50%, 80%, 90% and 95%.[7] These results show

[7]Note that in all these scenarios, the 3-states HMM is chosen as the best model.

that the bias for the customers with truly 3 states increases as the proportion of customers with unidentified parameters increases.

Table 6: 3-state HMM transition probabilities of customers in Segment A for $\lambda \in \{50\%, 80\%, 90\%, 95\%\}$

(a) $\lambda = 50\%$: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
	1	0.832	0.151	0.018
		[0.781 0.872]	[0.105 0.202]	[0.004 0.041]
Q	**2**	0.095	0.717	0.188
		[0.063 0.135]	[0.640 0.795]	[0.110 0.260]
	3	0.050	0.223	0.726
		[0.030 0.076]	[0.120 0.312]	[0.644 0.820]

(b) $\lambda = 80\%$: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
	1	0.839	0.140	0.021
		[0.784 0.880]	[0.096 0.195]	[0.008 0.040]
Q	**2**	0.130	0.829	0.041
		[0.088 0.178]	[0.747 0.878]	[0.009 0.114]
	3	0.060	0.258	0.682
		[0.021 0.134]	[0.121 0.373]	[0.582 0.797]

(c) $\lambda = 90\%$: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
	1	0.837	0.153	0.010
		[0.769 0.889]	[0.098 0.220]	[0.002 0.025]
Q	**2**	0.131	0.833	0.037
		[0.090 0.178]	[0.773 0.876]	[0.011 0.082]
	3	0.123	0.206	0.671
		[0.028 0.281]	[0.060 0.370]	[0.544 0.815]

(d) $\lambda = 95\%$: Population mean transition matrix posterior mean and posterior 95% intervals

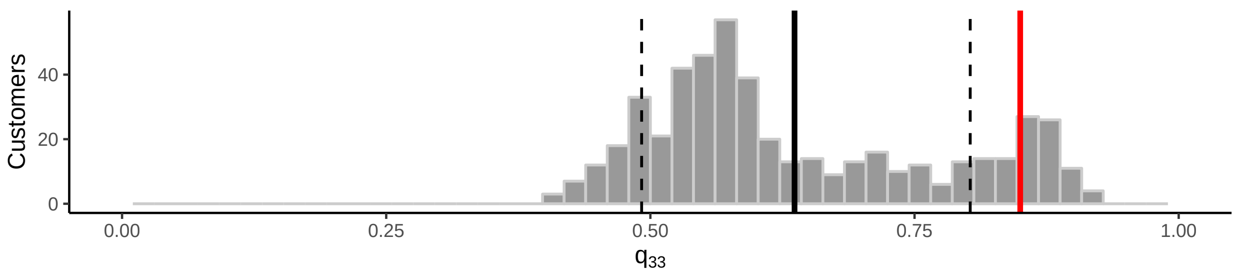
		1	**2**	**3**
	1	0.851	0.137	0.012
		[0.799 0.888]	[0.099 0.189]	[0.003 0.025]
Q	**2**	0.133	0.836	0.031
		[0.083 0.202]	[0.752 0.889]	[0.005 0.082]
	3	0.264	0.251	0.485
		[0.090 0.475]	[0.030 0.461]	[0.348 0.649]

We further investigate the individual bias by computing the posterior mean of the transition probabilities for each individual. Figure 4 shows the histogram of the individual posterior means for q_{33}. Figures 4a, 4b, and 4c depict these values for all, Segment A, and Segment B customers, respectively.

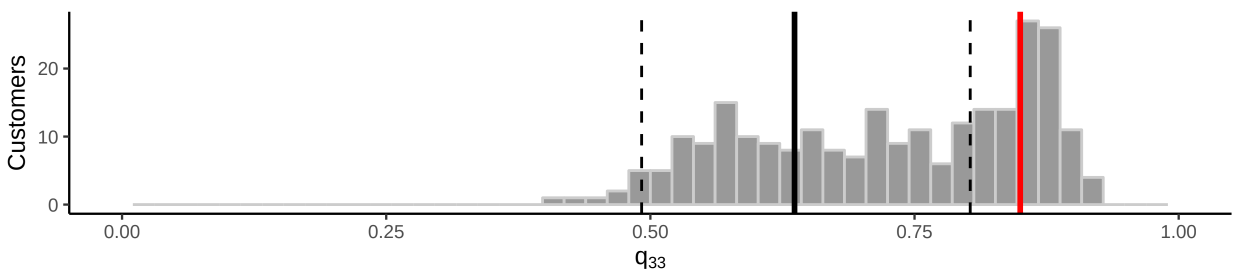
Figure 4 is consistent with the unidentification problem and the individual-level bias for most customers. Figure 4a shows how the posterior mean (and confidence interval in dashed black lines) excludes a significant portion of customers. The location of $\widehat{q}_{33}$ for customers in Segment B is close to 0.5 (Figure 4c), which drives the overall population mean downwards. More importantly, a large proportion of customers of Segment A have biased estimates with values significantly lower than the true 0.85 (Figure 4b).

Figure 4: Histogram of the individual posterior means of q_{33}. The solid black line represents the population posterior mean whereas the dashed black lines represent the population posterior 95% intervals. The solid red line represents the true value of $q_{33} = 0.85$ for Segment A customers.

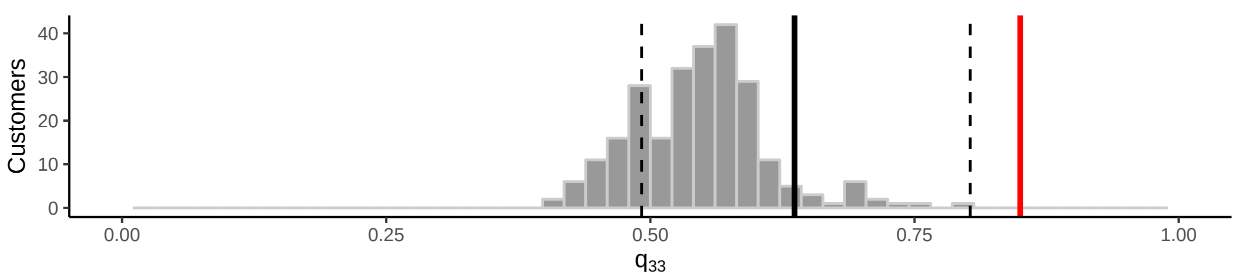
(a) All customers



(b) Segment A (3 states)



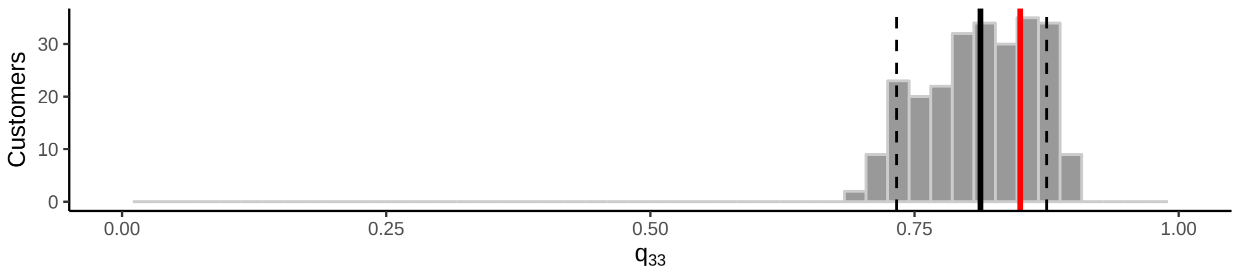
(c) Segment B (2 states)



Note that when we estimate an HMM using only the customers in Segment A (which we can do in a simulated context), the bias reported in Figure 4b disappears (see Figure 3).[8] This means that this bias on customers with truly 3 states is caused by the pooled estimation of Segment A and Segment B customers by a 3-states HMM model.

[8]Specifically, we estimate the model considering the $I \cdot (1 - \lambda)$ customers for the scenario with $\lambda = 50\%$. We do this to keep the number of customers with 3 states constant. Note that in this case, we have less data than in the baseline case that includes both segments. The details on the selection of the number of states and the parameter estimates can be seen in Appendix D.

Figure 5: Histogram of the individual posterior means of q_{33} when estimating the model using only Segment A customers. The solid black line represents the population posterior mean whereas the dashed black lines represent the population posterior 95% intervals. The solid red line represents the true value of $q_{33} = 0.85$ for Segment A customers.

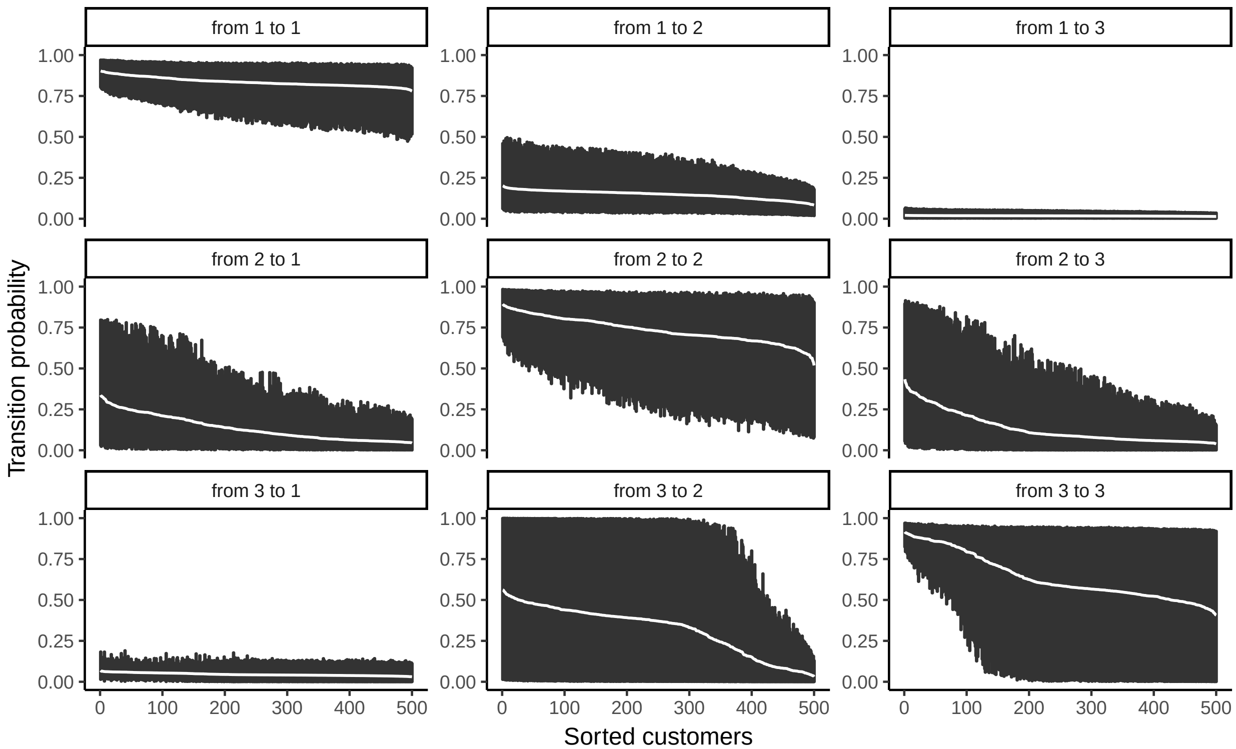


Therefore, the pooling of customers with different number of states may produce an important bias in both incorrectly and correctly estimated customers. Indeed, the uncertainty in the parameter estimates is a result of this lack of identification (see Figure 6). Figure 6a reports the posterior mean and 95% confidence intervals of the individual transition probabilities for all customers sorted in decreasing estimated probability. For some transition probabilities we observe large confidence intervals for many customers that prevent to statistically conclude any behavior. For instance, note the dashed area for the corresponding transition probabilities q_{22} and q_{33}. In those cases, the estimated posterior distributions move fully in the $[0, 1]$ interval. By plotting separately Segment A and Segment B customers, we confirm that the uncertainty is more severe for Segment B customers (see the corresponding last row of the transition matrix in Figure 7b)

Note that the uncertainty reduces importantly if we estimate a model only considering Segment A customers (with 3 states) although we use only half the observations than in the pooled case (see Figure 6b and compare it to Figure 7a).

Figure 6: Posterior mean and 95% confidence interval for individual transition probabilities. For each transition probability, customers are sorted decreasingly on their posterior mean. The shaded area represents the 95% intervals and the white line represents the posterior mean.

(a) Using both segments A (3 states) and B (2 states)



(b) Using only segment A (3 states)

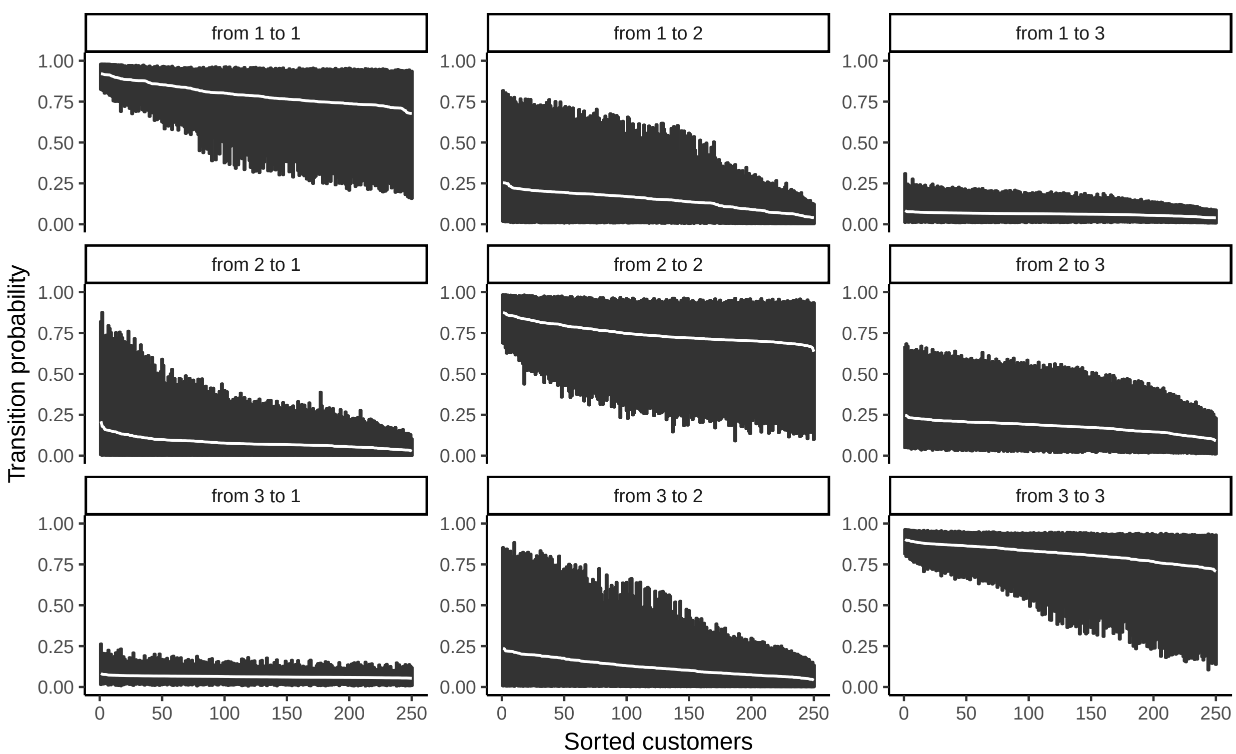
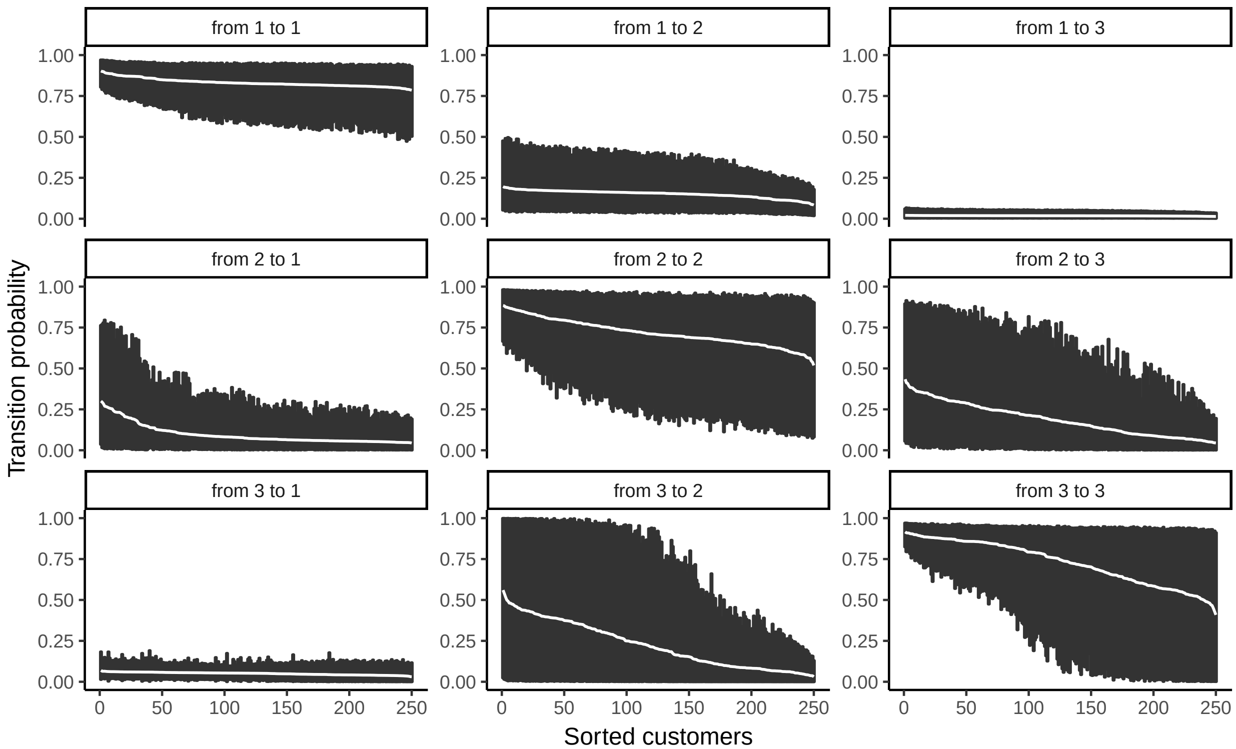
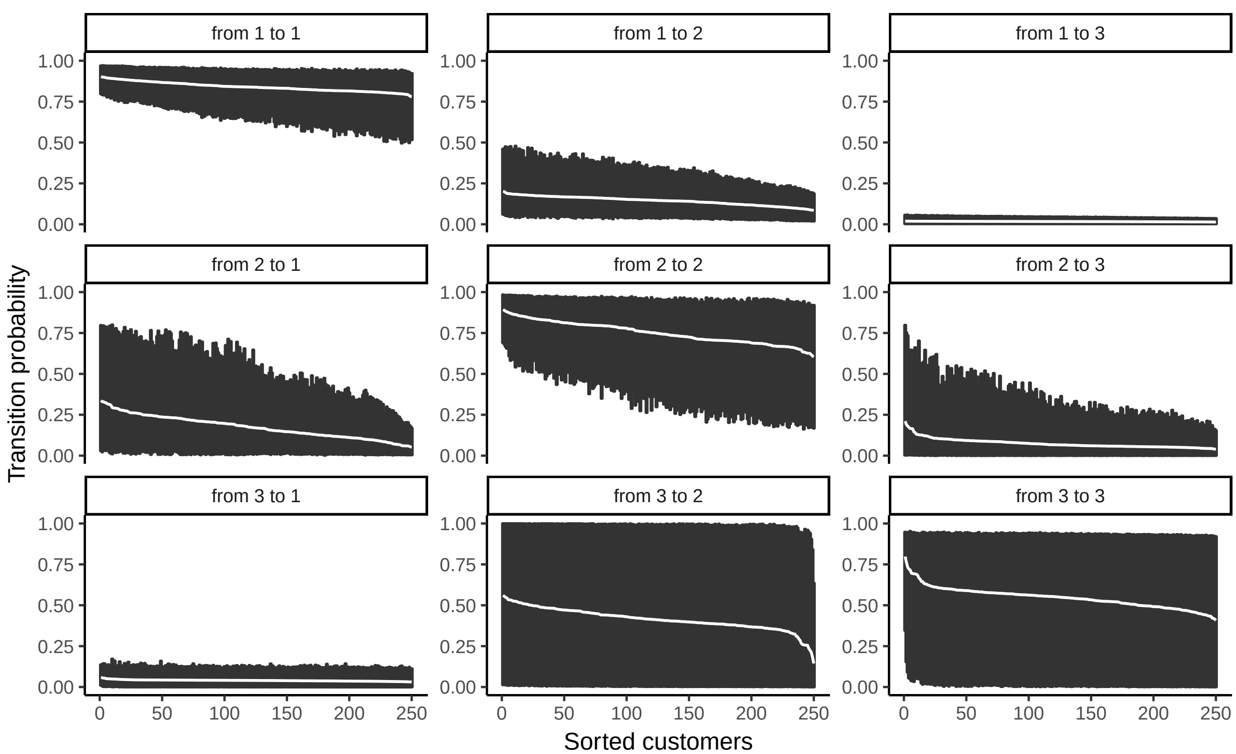


Figure 7: Posterior mean and 95% confidence interval for individual transition probabilities. For each transition probability, customers are sorted decreasingly on their posterior mean. The shaded area represents the 95% intervals and the white line represents the posterior mean.

(a) Segments A (3 states)



(b) Segments B (2 states)



4 Mixture of HMM

As we have seen in Section 3, not accounting for heterogeneity in the number of states often leads to identification problems, biased estimates, and potentially misleading insights. In this section we present our approach to allow for heterogeneity in the number of states by proposing a mixture of HMMs.

4.1 Model

The idea behind our proposed mixture of hidden Markov models (MHMMs) is to create a model that combines HMMs with varying number of states and allow each customer to probabilistically belong to an HMM process with the number of states that best reflects her behavior. Following Equation (3) we can define the likelihood of an HMM with S_m states as $L_{i,HMM}(Y_i|Q_i^m, p_i^m, S_m)$, where Q_i^m and p_i^m are the corresponding parameters of customer i for an HMM with S_m states. Our MHMM assumes that each customer has a probability, λ_m, of belonging to an HMM with S_m states. Let $m = \{1, \ldots, M\}$ be a class that follows an HMM with S_m number of states[9] and λ_m the probability that a customer belongs to class m. Then the likelihood of the MHMM can be written as:

$$L_i(\{p^m, Q_i^m\}_{m=1}^M, \lambda|Y_{i,1:T}) = \sum_{m=1}^{M} \lambda_m L_{i,HMM}(p_i^m, Q_i^m|Y_{i,1:T}, NS_i = S_m)$$

Similar to the case of a single HMM, we use the procedure described in Section 2.3 to estimate the parameters of the MHMM.

4.2 Results

We estimate the MHMM on the simulated data described in the previous section for the case with 50% customers in each segment. We show the fit measures for the 3-states HMM model chosen considering the validation sample and the MHMM with 1, 2, and 3 states (Table 7). We compare these models using the holdout sample, which we have not used to choose the number of states within the HMM models. The log-likelihood in the holdout sample suggests a better predictive performance of the MHMM model.

[9]For simplicity, we assume $S_m = m$, but more generally we can choose the number of states by cross validation.

Table 7: Model comparison for $\lambda = 50\%$ of customers with 2 states. Note: The best model in each column is in bold

		In sample			Validation	Holdout
Model	Number of States	LL	LMD	WAIC	LL	LL
HMM	3	**-11612.95**	**-11850.43**	**23805.67**	-3290.16	-3751.95
MHMM	**1, 2, and 3**	-11799.09	-11939.09	23922.88	**-3251.30**	**-3711.42**

To go beyond model fit, we investigate whether the MHMM captures well the behavior of customers with 2 and 3 states. Table 8 represents the parameter estimates for the MHMM for the case with 50% customers in each segment. Consistent with the simulated populations, the mixture membership probabilities assign customers to either the 2- or 3-state HMMs with similar probabilities. Only a small fraction (approximately 1%) of the customers are erroneously allocated, a-priori, to a 1-state HMM (see Table 8a). Looking at the state dependent behavior (p), and the transition probabilities (Q) of the component with 2 states (Tables 8d and 8e) and the component with 3 states (Tables 8g and 8h) the MHMM successfully recovers the behavior of the customers for each segment. Indeed, all true parameters fall within the corresponding confidence intervals.

Table 8: MHMM parameter estimates for $\lambda = 50\%$

(a) Membership probabilities to each MHMM component. Posterior mean and posterior 95% intervals

	1 state	**2 states**	**3 states**
$\lambda_\mathbf{m}$	0.014	0.507	0.480
	[0.000 0.046]	[0.405 0.602]	[0.382 0.583]

(b) 1 state component: State dependent probabilities posterior mean and posterior 95% intervals

	1
$\mathbf{p_s}$	0.528
	[0.082 0.976]

(c) 2 state component: Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**
$\pi_\mathbf{s}$	0.509	0.491
	[0.374 0.633]	[0.367 0.626]

(d) 2 state component: State dependent probabilities posterior mean and posterior 95% intervals

	1	**2**
$\mathbf{p_s}$	0.115	0.535
	[0.089 0.139]	[0.493 0.575]

(e) 2 state component: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**
	1	0.840	0.160
		[0.790 0.881]	[0.119 0.210]
Q	**2**	0.176	0.824
		[0.134 0.224]	[0.776 0.866]

(f) 3 state component: Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\pi_\mathbf{s}$	0.432	0.321	0.247
	[0.307 0.551]	[0.169 0.475]	[0.150 0.356]

(g) 3 state component: State dependent probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\mathbf{p_s}$	0.113	0.501	0.889
	[0.089 0.138]	[0.431 0.563]	[0.868 0.911]

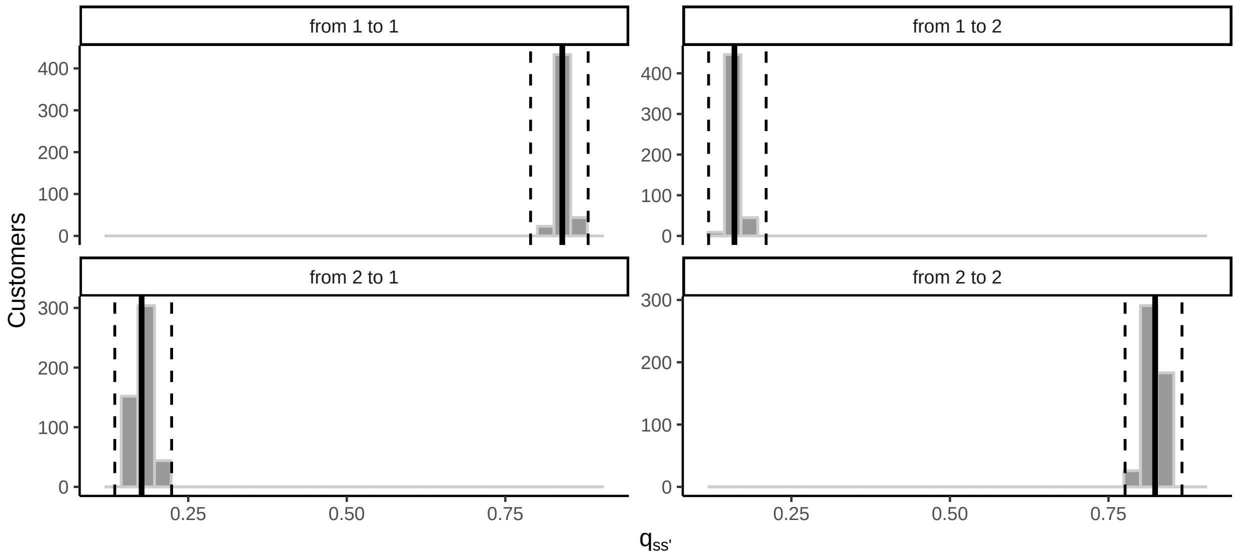
(h) 3 state component: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
	1	0.801	0.137	0.062
		[0.701 0.880]	[0.060 0.242]	[0.021 0.124]
Q	**2**	0.043	0.762	0.195
		[0.002 0.142]	[0.621 0.868]	[0.102 0.310]
	3	0.070	0.085	0.845
		[0.042 0.099]	[0.029 0.172]	[0.767 0.893]

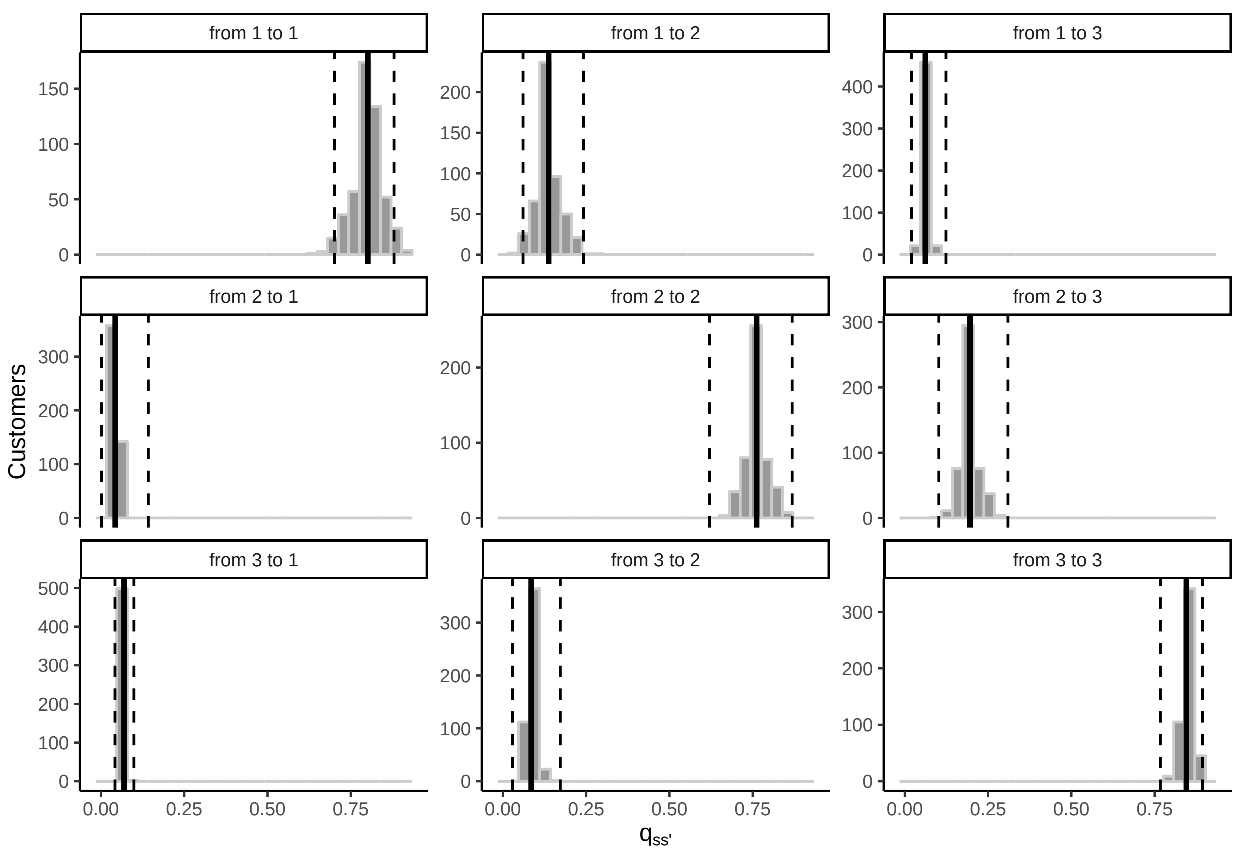
Figure 8 shows the heterogeneity in the estimated transition probabilities for the components with 2 and 3 states of the MHMM. We observe that the heterogeneity within each component is not as large as in the pooled 3-sates HMM (Figure 2). Indeed, the fact that most of the individual transition probabilities fall within the 95% confidence intervals of the population means, suggests that the pooling within each component do not yield severe bias.

Figure 8: Histogram of individual posterior mean of transition probabilities for each component of the MHMM. In solid black line is the population posterior mean, and in dashed lines the 95% confidence interval

(a) 2 states component



(b) 3 states component



We now evaluate whether the MHMM is capable of allocating each customer to the class with the appropriate number of states. Table 9 presents the confusion matrix comparing the "true" number of states based on the simulated data and the posterior membership to each class based on the MHMM's prediction. To predict each customer membership, we consider the highest probability to each class, by using

$$\begin{aligned} p(NS_i = S_m | Y_{i,1:T}, \{\lambda_m, p_i^m, Q_i^m\}_{m=1}^M) &\propto p(NS_i = S_m) \cdot p(Y_{i,1:T}, \{p_i^m, Q_i^m\}_{m=1}^M | NS_i = S_m) \\ &\propto \lambda_m \cdot L_{i,HMM}(p_i^m, Q_i^m | Y_{i,1:T}, NS_i = S_m). \end{aligned}$$

Table 9: Confusion matrix of customer allocation of HMM component for each segment

		True	
		2 states (Segment B)	3 states (Segment A)
MHMM component	1-state component	0	0
	2-states component	226	67
	3-states component	24	183

As can be seen in Table 9 the MHMM does a reasonably good job in allocating customers to the HMM with the appropriate number of states. Indeed, 81.8% ($= \frac{226+183}{500} \cdot 100$) of the customers are allocated correctly. The model is also able to recover the parameters at the individual level. To illustrate this characteristic, we show in Figure 9 the individual-level estimates of q_{33} for all customers, similarly as in Section 3.2.4.[10]

Figure 9: Histogram of individual-level posterior mean of q_{33} for the MHMM 3-states component. The solid black line represents the population posterior mean whereas the dashed black lines represent the population posterior 95% intervals. The solid red line represents the true value of $q_{33} = 0.85$ for Segment A customers.

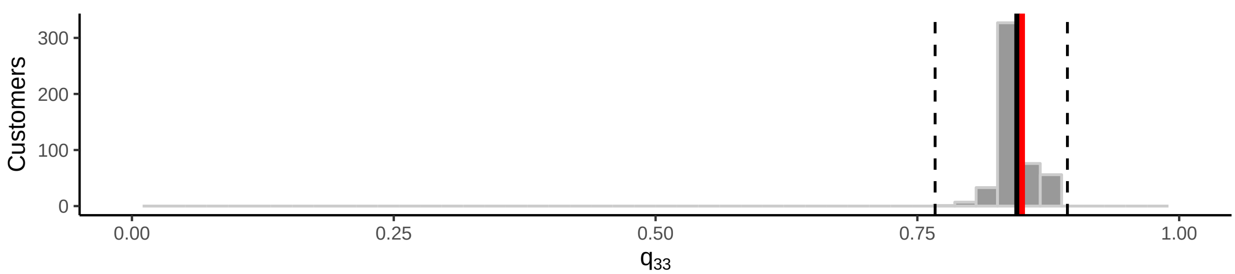


Figure 9 shows that the model is able to accurately capture the behavior of customers with 3 states. Note that we plot the estimated q_{33} for all customers as the MHMM provides an estimate for each individual regardless of her posterior membership probability. This result suggests that customers in Segment B are being correctly shrinkaged to the mean of the Segment A estimates and consequently the parameter estimates

[10]We select to show q_{33} because this parameter highlights better the unidentification problem in the transition matrix for the 3-states HMM.

for customers with 3 states do not suffer from bias.

Therefore, these results provide compelling evidence that the proposed MHMM not only predicts the data well, but unlike the standard HMM, it is also capable of correctly capturing the underlying dynamics of customers with varying number of hidden states. In the next section we use secondary data from an online role playing game to compare the proposed MHMM to an standard HMM that assumes the same number of states for all customers.

5 Empirical application

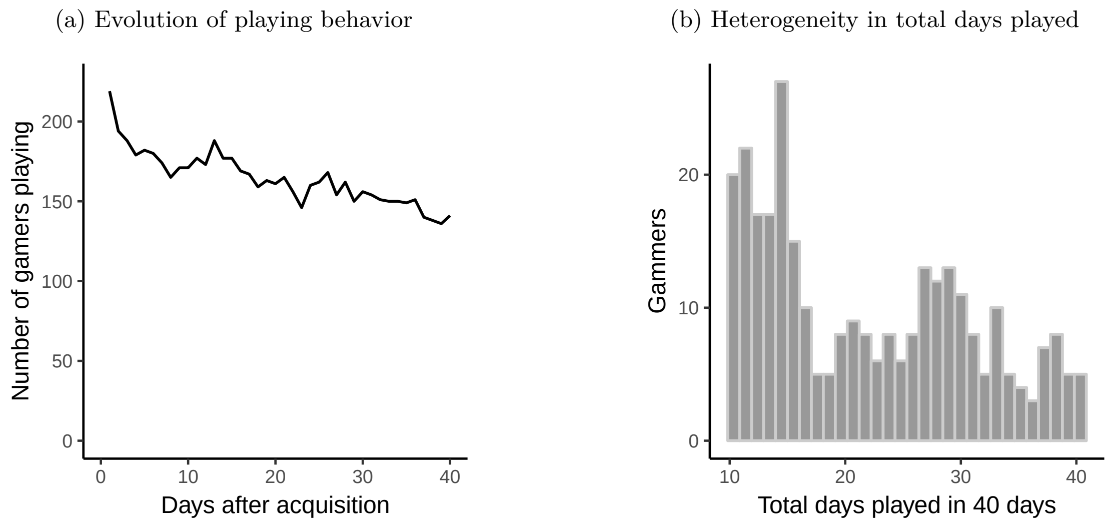
In this section we use secondary data about users' activity in an online role playing game to contrast the HMM and the MHMM in an actual marketing setting. While we do not know the true number of states for each customer in this dataset, it allows us to compare the performance of the HMM and MHMM both in terms of predictive ability and the insights each model generates.

5.1 Data

Our data comprise online game behavior in a role playing online game from a random sample of 300 gamers over a 60-day period after each gamer was acquired. Our full observation period is between April 1, 2008 and December 31, 2008. We observe when each player started the game for the first time and follow them for a 60-day period. We selected *active* gamers that played at least 10 days during their first 40 days (note that more than 40% of gamers did not play more than two days during those 40 days). This subset of players accounts for only 34% of all gamers that started playing after April 1, 2008, but they represents 77% of the total number of days played by all gamers acquired after April 1, 2008. We use the first 40 days for calibration, the next 10 days for validation (for selecting the number of hidden states of the HMM), and the last 10 days for testing the best HMM model against the proposed MHMM. Consistent with common metrics in game user analysis (Huang et al., 2018), our aim is to predict whether the gamer plays on any given day.

Figure 10 summarizes the playing behavior for the selected sample of 300 gamers. Figure 10a shows the commonly observed overall decline in playing behavior over the curse of the gamers life post acquisition. Figure 10b suggests a high degree of heterogeneity in the playing behavior with some players playing every single day for the entire 40 days and others as few as 10 days. Thus, an appropriate model of usage behavior should be able to capture both the heterogeneity and the dynamics in usage behavior over time.

Figure 10: Summary of daily gaming behavior



5.2 HMM and MHMM specifications

Following the description in Section 2.1.1 we specify the HMM of gamers' behaviors as follows:

- Initial probabilities: We estimate the probabilities directly from the data such that π_s is the probability that a gamer starts in state s with $\pi_s \in [0, 1]$ and $\sum_s \pi_s = 1$.
- Transition probabilities: We parametrize each row of the transition matrix using the softmax function (see Appendix B.2).
- Conditional gaming behavior: We specify a Binomial distribution with $N = 1$. Specifically

$$P(Y_{it} = y_{it}|Z_{it} = s) = m_{it|s}(y_{it}) = \text{Binomial}(y_{it}|N = 1, p_{its}) \tag{4}$$

  where $p_{its} = \text{logit}^{-1}(u_i + \phi_s + b_s X_{it})$. Thus u_i represents an individual-specific random effect, ϕ_s captures the state level, and b_s is the effect of our covariate when the gamer is in the state s. In this application X_{it} is a binary variable that indicates for gamer i if the day t corresponds to a Friday, Saturday or Sunday.

After specifying the HMM, we follow the description in Section 4.1 to specify the corresponding MHMM. In particular, for each component of the MHMM we specify an HMM with one, two and three states as described above. The parameters for the HMM and the MHMM are estimated as described in Section 3.1.

5.3 Results

5.3.1 Model selection for the HMM

We first determine the best specification for the HMM. To infer the number of states that best represents the data, we estimated the HMM for varying number of states. Based on the Validation log-likelihood measure we selected a 3-states HMM (see Table 10). Note that, as in our simulation exercise, the in-sample fit measures suggest a more complex specification.

Table 10: Selection of number of hidden states for the HMM. Note: The best model in each column is in bold

	In sample			**Validation**
Number of states	LL	LMD	WAIC	LL
1	-7006.59	-7087.67	14293.49	-1828.26
2	-5968.33	-6128.45	12501.01	-1424.01
3	-5904.30	-6084.65	12422.38	**-1390.34**
4	**-5817.53**	**-6041.75**	**12407.65**	-1390.93

5.3.2 Parameter estimates for the 3-state HMM

In Table 11 we report the population estimates of the 3-state HMM whereas in Figure 11 we report the individual posterior means for the state dependent and for the transition probabilities. First, the results suggest that the three states of the HMM would correspond to a low, medium and high states of gaming behavior with playing probabilities in any particular day of 0.089, 0.579, and 0.866, respectively (see Table 11b). Note that the heterogeneity in the conditional behavior, p_s, when a gamer is in the second state is substantial (see also Figure 11a). Second, the results for π_s in Table 11a suggest that most gamers start in the high state. Finally, regarding gamers' dynamics, Table 11c shows a large uncertainty in the population transition matrix, Q, particularly for state 2. In addition, Figure 11b shows important heterogeneity in the transition probabilities across gamers, particularly in the third state. Overall, these results seem to suggest that the second state is less sticky than the other two states and once the gamer is in that state she is almost as likely to stay as she is to move to the other states, and in particular to State 3. For completeness we also report in Table 11d the estimates for the effect of weekend days which indicate that during Friday, Saturday or Sunday a gamer is less likely to play if she is in the low or medium state but more likely to play if she is in the high state.

Table 11: 3-state HMM parameter estimates

(a) Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\pi_\mathbf{s}$	0.137	0.129	0.734
	[0.038 0.216]	[0.031 0.247]	[0.643 0.814]

(b) Population mean state dependent probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\mathbf{p_s}$	0.089	0.579	0.866
	[0.048 0.126]	[0.231 0.833]	[0.830 0.902]

(c) Population mean transition matrix posterior mean and posterior 95% intervals

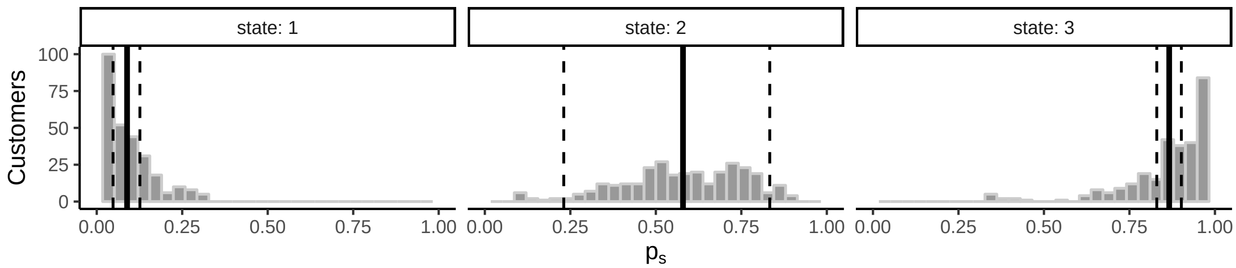
		1	**2**	**3**
	1	0.881	0.034	0.086
		[0.803 0.926]	[0.000 0.166]	[0.013 0.126]
Q	**2**	0.064	0.519	0.417
		[0.002 0.169]	[0.362 0.669]	[0.264 0.595]
	3	0.095	0.118	0.788
		[0.002 0.148]	[0.065 0.206]	[0.750 0.820]

(d) Weekend effect posterior mean and posterior 95% intervals

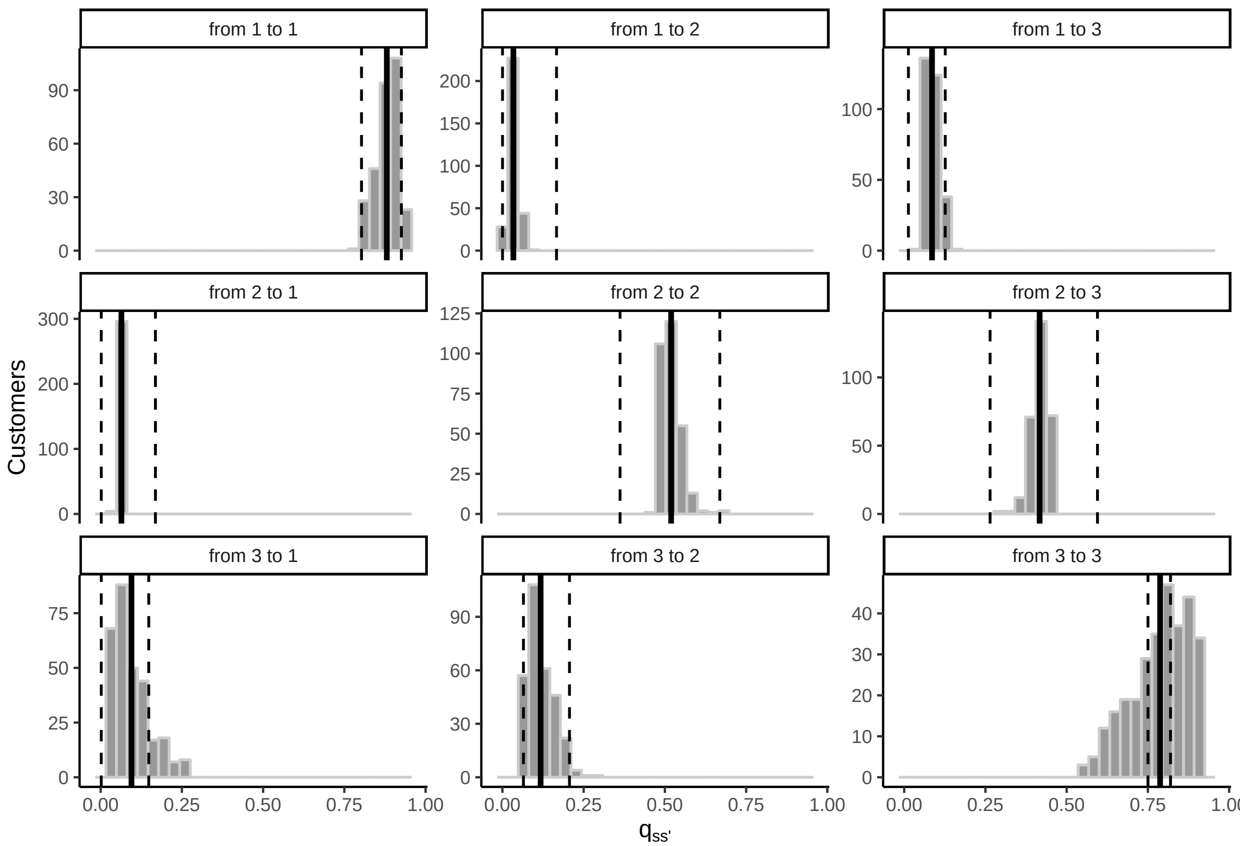
	1	**2**	**3**
$\mathbf{b_s}$	0.502	-4.529	3.501
	[−0.114 1.540]	[−9.320 − 2.200]	[2.065 7.506]

Figure 11: Histogram of individual posterior means of the state-dependent probabilities and the transition probabilities for 3-states HMM. The solid black line represents the population posterior mean whereas the dashed black lines represent the population posterior 95% intervals.

(a) State dependent probabilities



(b) Transition probabilities



5.3.3 Parameter estimates for the MHMM

Table 12 reports the parameter estimates for the proposed MHMM summarizing the posterior means and 95% confidence intervals for the population mean. Figure 12 shows the histogram of the individual conditional probabilities and Figure 13 shows the histogram of the individual transition probabilities. The aggregated and the individual level estimates show similar patterns. Therefore, we will use both sources of information to describe the behavior implied by the estimated MHMM.

Table 12a reports the population membership probabilities, λ_m, and indicates that most gamers (75.4%) have 3 states whereas 18.5% have 2 states and 6.1% have only one state. Let's now proceed to describe the behavior captured by each component.

1-state gamers. These static gamers show a medium to level of playing behavior overall with probability of playing, $p = 0.395$ (see Table 12b).

2-state gamers. These dynamic gamers have two distinct states, characterized by a low level of playing behavior when they are in the low state, with $p_1 = 0.190$, and a very high level of playing behavior when they are in the high state, with $p_2 = 0.969$ (see Table 12d). All these gamers start in the high state (see Table 12c) and stay in that state with high probability, $q_{22} = 0.913$. However, if the move to the low state, they also stay there with high probability, $q_{11} = 0.787$ (see Table 12e).

3-state gamers. These dynamic gamers have three distinct states, characterized by an inactive behavior when they are in the low state, with playing probability $p_1 = 0.025$, a low level of playing behavior when they are in the second state, with $p_2 = 0.254$, and a very high level of playing behavior when they are in the third state, with $p_3 = 0.941$ (see Table 12g). Most of these gamers (60.8%) start in the high state (see Table 12f) and stay in that state with relatively high probability, $q_{33} = 0.790$. In addition, 39.2% of the gamers start in the second state but once gamers transition to that state they stay there with also relatively high probability, $q_{22} = 0.731$. Finally, only 2.5% of the gamers start in the inactive state and stay there with high probability, $q_{11} = 0.902$. Therefore, the 3-state gamers are more dynamic than the 2-state gamers as the probabilities in the diagonal of the transition matrix are less sticky than the corresponding probabilities for the 2-state gamers.

We note that the individual estimates illustrated in Figures 12 and 13 confirm the characterization implied by the population estimates.

Table 12: MHMM parameter estimates

(a) Component probabilities posterior mean and posterior 95% intervals

	1 state	**2 states**	**3 states**
$\lambda_{\mathbf{m}}$	0.061	0.185	0.754
	[0.024 0.116]	[0.105 0.289]	[0.636 0.848]

(b) 1 state component: Population mean state dependent probabilities posterior mean and posterior 95% intervals

	1
$\mathbf{p_s}$	0.395
	[0.173 0.676]

(c) 2 state component: Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**
$\pi_{\mathbf{s}}$	0.000	1.000
	[0.000 0.000]	[1.000 1.000]

(d) 2 state component: Population mean state dependent probabilities posterior mean and posterior 95% intervals

	1	**2**
$\mathbf{p_s}$	0.190	0.969
	[0.054 0.352]	[0.929 0.999]

(e) 2 state component: Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**
Q	**1**	0.787	0.213
		[0.669 0.865]	[0.135 0.331]
	2	0.087	0.913
		[0.048 0.146]	[0.854 0.952]

(f) 3 state component: Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\pi_{\mathbf{s}}$	0.000	0.392	0.608
	[0.000 0.000]	[0.277 0.535]	[0.465 0.723]

(g) 3 state component: Population mean state dependent probabilities posterior mean and posterior 95% intervals

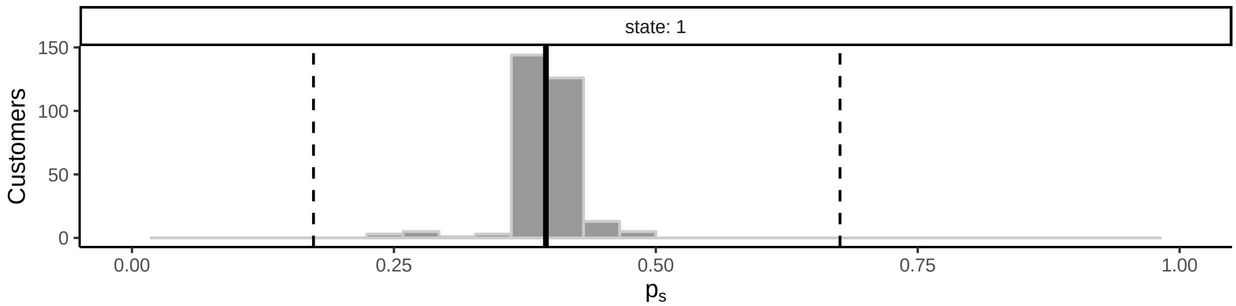
	1	**2**	**3**
$\mathbf{p_s}$	0.025	0.254	0.941
	[0.000 0.069]	[0.181 0.360]	[0.892 1.000]

(h) 3 state component: Population mean transition matrix posterior mean and posterior 95% intervals

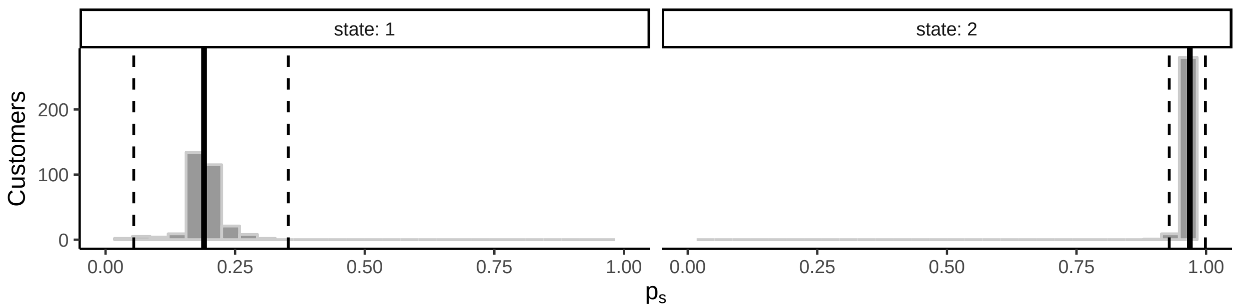
		1	**2**	**3**
Q	**1**	0.902	0.042	0.056
		[0.724 0.963]	[0.000 0.227]	[0.017 0.100]
	2	0.083	0.731	0.186
		[0.034 0.161]	[0.602 0.843]	[0.102 0.279]
	3	0.002	0.208	0.790
		[0.000 0.012]	[0.155 0.274]	[0.723 0.841]

Figure 12: Histogram of individual level posterior mean of state dependent probabilities for each component of the MHMM. In solid black line is the population posterior mean, and in dashed lines the population mean 95% CPI.

(a) 1 state component



(b) 2 states component



(c) 3 states component

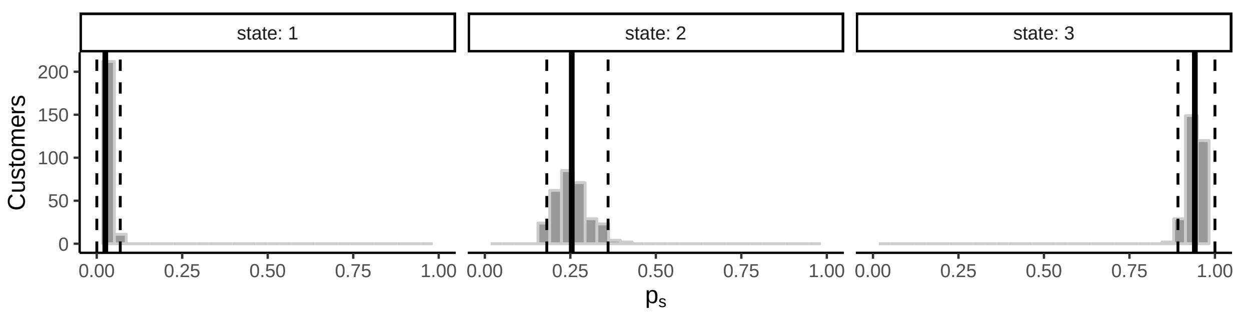
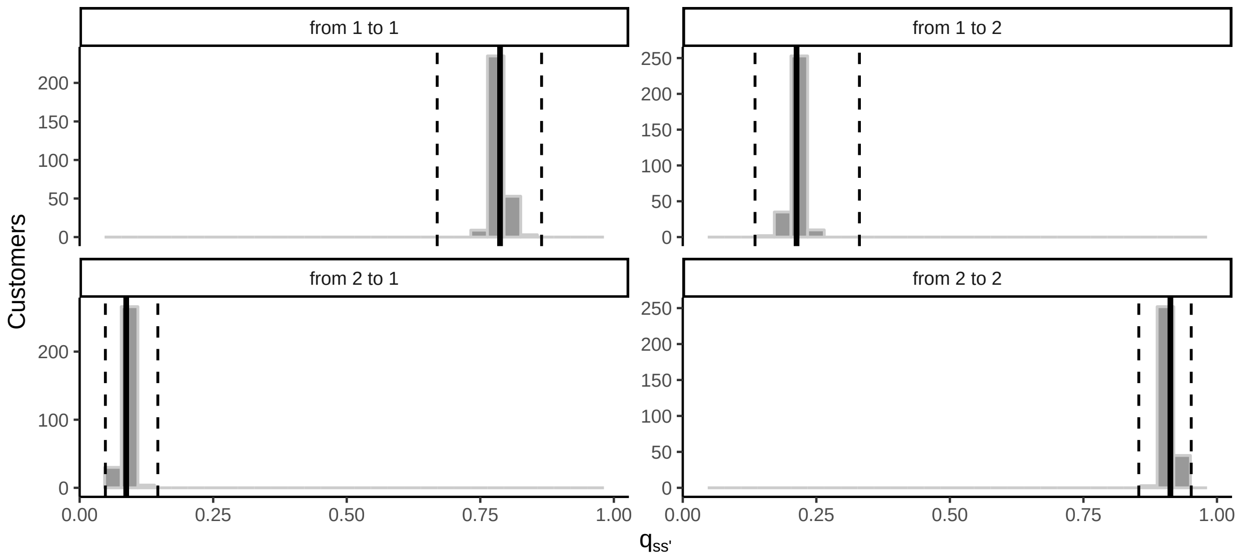
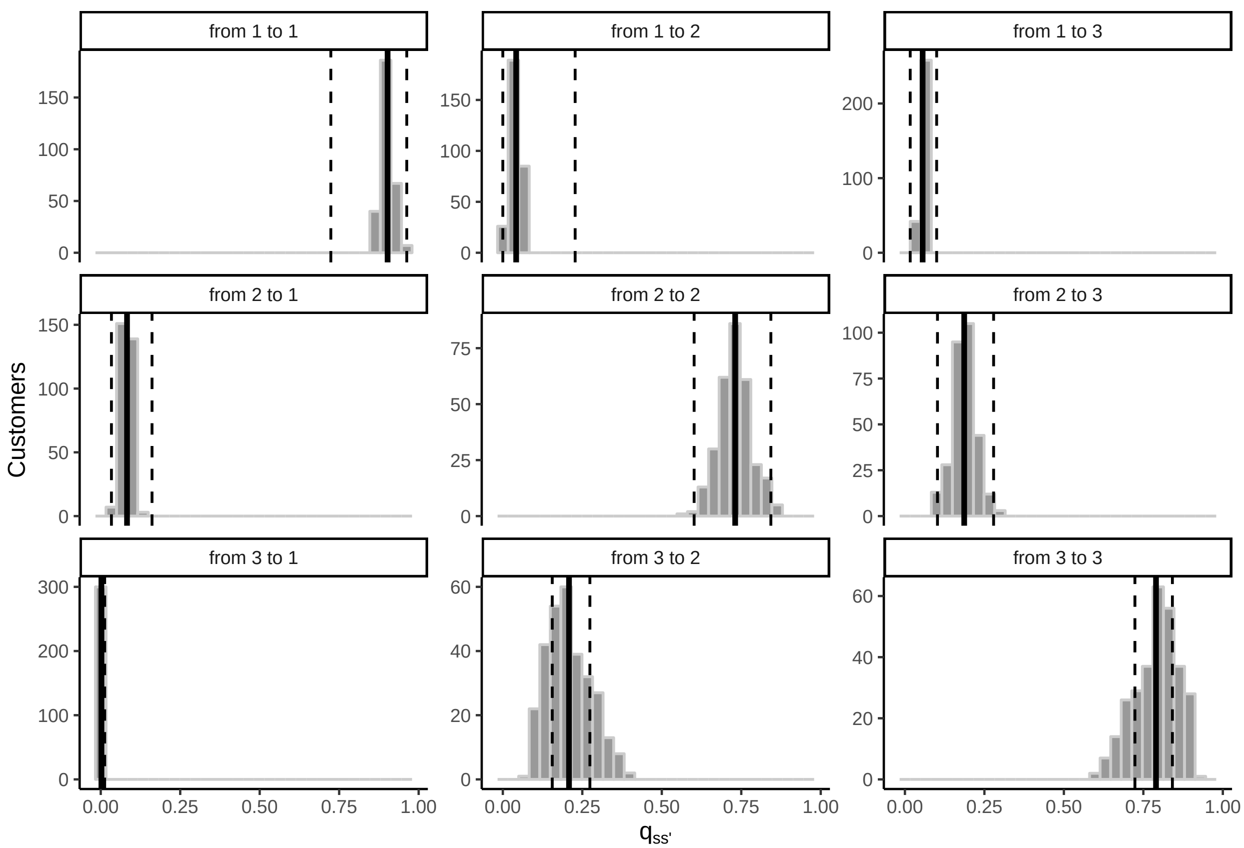


Figure 13: Histogram of individual level posterior mean of transition probabilities for each component of the MHMM. In solid black line is the population posterior mean, and in dashed lines the population mean 95% CPI.

(a) 2 states component



(b) 3 states component



5.3.4 Comparison between HMM and MHMM

We now proceed to compare the results of the HMM and MHMM by first looking at the predictive performance and then by contrasting the characterization of the dynamics implied by each approach. In Table 13 we report the in-sample fit and out-of-sample predictions of the selected 3-state HMM and the proposed MHMM. We note that the HMM does better in terms of in-sample fit whereas the proposed MHMM has a better performance in the Validation data (that was used to select the number of states in the HMM) and the holdout data that were never used by neither of these models.

Table 13: Model selection: HMM vs. MHMM. Note: The best model in each column is in bold

		In sample			**Validation**	**Holdout**
Model	Number of states	LL	LMD	WAIC	LL	LL
HMM	3	**-5904.30**	**-6084.65**	**12422.38**	-1390.34	-1355.54
MHMM	**1, 2, and 3**	-5979.15	-6205.65	12594.02	**-1380.35**	**-1342.09**

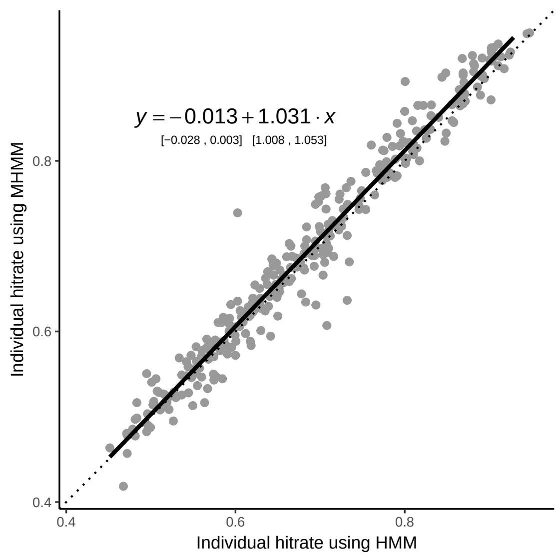
We further compare the predictive performance of the two selected models by reporting the RMSE between the predicted and the true number of customers playing each day during the holdout period (see Table 14). We note that considering all gamers the MHMM shows better predictive ability (lower RSME, 12.96 vs 14.31). Now, if we use the MHMM to split the sample between the gamers with different number of states, the RMSE of the HMM is slightly better for gamers with 1 and 2 states (4.68 vs 4.99) whereas the RMSE of the proposed MHMM is better for gamers with 3 states (10.49 vs 12.39). That is, for gamers with simpler dynamics, both models yield fairly similar predictions whereas for gamers with richer dynamics, the MHMM provides a better prediction.

Table 14: RMSE of the number of gamers playing per period using: (1) all gamers, (2) gamers identified by the MHMM with 1 or 2 states, and (3) gamers identified by the MHMM with 3 states. Posterior mean and 95% confidence intervals are reported

	All		**1 and 2 states**		**3 states**	
Model	Post. Mean	CPI	Post. Mean	CPI	Post. Mean	CPI
HMM	14.31	[11.21 , 17.64]	4.68	[3.48 , 6.09]	12.39	[9.37 , 15.54]
MHMM	12.96	[10.20 , 16.31]	4.99	[3.87 , 6.43]	10.49	[7.97 , 13.15]

To study the superior predictive performance of the MHMM we compute the hit rate between the predicted behavior (by each model) and the true behavior for the holdout period. Figure 14 contrasts the implied hit rates of each model at the individual level. We observe that overall the two models yield fairly similar results but the advantage of the MHMM is stronger for gamers that are relatively easier to predict (for which both models yield high hit rates). Statistically, the slope of a regression of individual hit rates of the MHMM (y) on the individual hit rates of the HMM (x) yield a slope significantly higher than 1 (at the 95% significance level) and a non-significant intercept, confirming the overall superior individual-level predictive ability of the MHMM relative to the HMM.

Figure 14: Individual level hit rate using HMM vs. MHMM.



The HMM and MHMM suggest that players star at a state with a high level of engagement and transition over time to lower states. This may be indicative of churn behavior, which is an extremely important metric in the context of online games. Accordingly, we further explore the ability of the models to predict gamers' defection. For that analysis, we consider a defected gamer if she does not play during the last 10 days of the holdout period. Figure 15 summarizes the predictive performance of each model with the corresponding ROC curves.

Figure 15: ROC curve for churn prediction: HMM vs. MHMM.

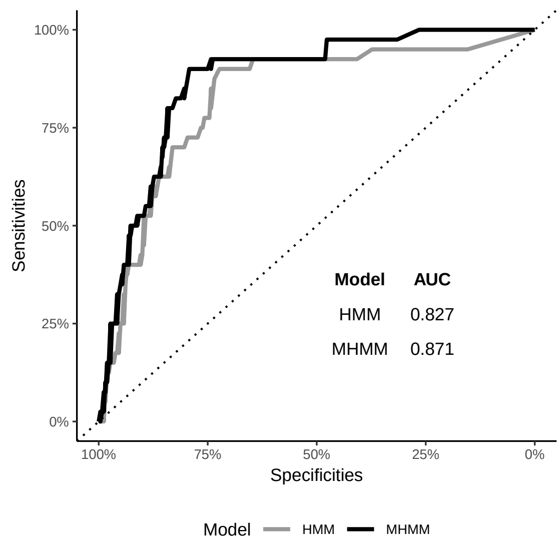


Figure 15 confirms the superior performance of the MHMM in predicting churn. Indeed, the area under the curve (AUC) of the MHMM is a 5.3% higher than that of the HMM.

Regarding the dynamics implied by both approaches the difference that stand out the most is the characterization of the second state. We note that the HMM suggests a medium state (probability of playing of $p = 0.579$) that is quite unsticky (probability of staying in that state of $q_{22} = 0.519$). In contrast, the MHMM does not find such state. The 3-state components of the MHMM suggests that gamers are transition among a non-playing, low-playing or intensive-playing, all of which are quite sticky. Thus, one could conclude that compared to our proposed MHMM, the HMM traditionally used in the literature fails by mixing *heterogeneity in dynamics* with *dynamics*. That is, as it is unable to capture the heterogeneity in dynamics, the model rationalizes such behavior by suggesting more dynamic behavior than the one implied by an MHMM that captures such heterogeneity in dynamics.

6 Discussion

HMMs have proved to be an effective modeling approach to describe customer dynamic behavior. Accordingly, there have been diverse applications of this method in Marketing covering different topics ranging from eye-tracking to B2B contexts. Although most of the applications allow for flexible heterogeneity in customer behavior it is usually assumed that all customers have the same degree of dynamics by constraining the customer sample to have the same number of hidden states.

In this work we first analyze through a series of simulations the implications of such an assumption. We simulate a population with two segments with different number of states (two and three hidden states) and then estimate the typical HMM that does not allow for such heterogeneity. We find that in contrast to selecting the model that best represents the behavior of the majority of the population, the existence of few customers with a high number of states dictates model selection and favors choosing expensive models. We observe that population estimates may not summarize well customers' behavior and must be interpreted with caution as such results may lead to misleading conclusions. Additionally, we observe biased estimates even for customers with the same number of states as estimated due to the impact of the customers with misspecified and unidentified on the parameter estimates of the entire population.

Note that the implied population estimates (usually taken from the hierarchical structure of the HB modeling approach) may not summarize well the behavior of the population. Especially the dynamics implied by the transition matrix Q. The transition matrix derived by using the population parameters mixes customers that reach all states with customers that never reach a certain state. These latter customers have un-identified parameters that bias the population estimates. If the researcher wants to report one transition matrix for the population (for HMM or even MHMM approach) we suggest reporting an histogram of individual level posterior mean estimates, to show the degree of heterogeneity in the population.

We propose a mixture model (MHMM) approach that flexibly captures heterogeneity in the number of states and at the same time allows for heterogeneity in the parameters. The proposed approach obtains good results by identifying correctly the behavior of the simulated segments.

The MHMM proved to be a flexible approach to account for heterogeneity in the number of states both in simulated and an online game empirical application. The application to the online gaming application further highlights the different dynamic patterns that may emerge from using a MHMM relative to a HMM.

However, we caution that using MHMM could become expensive in terms of number of parameters to estimate if one allows for more segments with higher number of states (with four or more hidden states). That being said, in most applications seen in the marketing literature, the estimated HMM with common number of state for the entire population has been relatively (typically 3 states). An alternative approach to the MHMM would could be to use a Hierarchical Dirichlet Process to let the data select the number of hidden states as part of the inference. However, application of such an approach with individual level selection of mass points can be computationally challenging. We leave this modeling approach for future research.

Overall, despite the deserved importance and care being given in the marketing literature and practice to capturing individual-level heterogeneity, application of HMMs in marketing commonly assume heterogenous parameters, but homogeneity in the number of states for all consumers. We demonstrate the risks of assuming such homogeneity and propose a possible solution. We hope this research will pave the road for further exploring individual-level dynamics using HMMs.

References

Ansari, A., Montoya, R., and Netzer, O. (2012). Dynamic learning in behavioral games: A hidden Markov mixture of experts approach. *Quantitative Marketing and Economics*, 10(4):475–503.

Ascarza, E. and Hardie, B. G. S. (2013). A Joint Model of Usage and Churn in Contractual Settings. *Marketing Science*, 32(February):570–590.

Ascarza, E., Netzer, O., and Hardie, B. (2018). Some Customers Would Rather Leave Without Saying Goodbye. *Marketing Science*, 37(1):54–77.

Brangule-Vlagsma, K., Pieters, R. G. M., and Wedel, M. (2002). The dynamics of value segments: Modeling framework and empirical illustration. *International Journal of Research in Marketing*, 19(3):267–285.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Li, P., and Riddell, A. (2016). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, VV(Ii).

Du, R. Y. and Kamakura, W. A. (2006). Household Life Cycles and Lifestyles in the United States. *Journal of Marketing Research*, 43(1):121–132.

Ebbes, P., Grewal, R., and DeSarbo, W. S. (2010). Modeling strategic group dynamics: A hidden Markov approach. *Quantitative Marketing and Economics*, 8(2):241–274.

Eddy, S. R. S. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9):755–763.

Fader, P. (2012). Customer Centricity: Focus on the Right Customers for Strategic Advantage, Second Edition | Wharton Digital Press.

Hamilton, J. (2008). Regime-switching models. *The new palgrave dictionary of economics*, pages 0–15.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357.

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press.

Huang, Y., Jasin, S., and Manchanda, P. (2018). Level up: Leveraging skill and engagement to maximize player game-play in online video games. *Working paper*.

Hughes, J. P. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*.

Kumar, V., Sriram, S., Luo, a., and Chintagunta, P. K. (2011). Assessing the Effect of Marketing Investments in a Business Marketing Context. *Marketing Science*, 30(5):924–940.

Lemmens, A., Croux, C., and Stremersch, S. (2012). Dynamics in the international market segmentation of new product growth. *International Journal of Research in Marketing*, 29(1):81–92.

Li, S., Sun, B., and Montgomery, A. L. (2011). Cross-Selling the Right Product to the Right Customer at the Right Time. *Journal of Marketing Research*, 48(4):683–700.

Liechty, J., Pieters, R., and Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika*, 68(4):519–541.

Luo, A. and Kumar, V. (2013). Recovering Hidden Buyer–Seller Relationship States to Measure the Return on Marketing Investment in Business-to-Business Markets. *Journal of Marketing Research*, L(February):143–160.

Ma, L., Sun, B., and Kekre, S. (2015). The Squeaky Wheel Gets the Grease — An Empirical Analysis of Customer Voice and Firm Intervention on Twitter. *Marketing Science*, 34(5):627–645.

Mamon, R. and Elliott, R. J. (2007). *Hidden Markov models in finance*. Springer.

Mark, T., Lemon, K. N., and Vandenbosch, M. (2014). Customer Migration Patterns: Evidence from a North American Retailer. *The Journal of Marketing Theory and Practice*, 22(3):251–270.

Mark, T., Lemon, K. N., Vandenbosch, M., Bulla, J., and Maruotti, A. (2013). Capturing the evolution of customer-firm relationships: How customers become more (or less) valuable over time. *Journal of Retailing*, 89(3):231–245.

Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science*, 23(4):579–595.

Montoya, R. and Gonzalez, C. (in press). A Hidden Markov Model to Detect on-shelf out-of-stocks. *Manufacturing and Service Operations Management*.

Montoya, R., Netzer, O., and Jedidi, K. (2010). Dynamic Allocation of Pharmaceutical Detailing and Sampling for Long-Term Profitability. *Marketing Science*, 29(5):909–924.

Moon, S., Kamakura, W. a., and Ledolter, J. (2007). Estimating Promotion Response When Competitive Promotions Are Unobservable. *Journal of Marketing Research*, 44(3):503–515.

Netzer, O., Ebbes, P., and Bijmolt, T. H. (2017). Hidden markov models in marketing. In *Advanced methods for modeling markets*, pages 405—-449. Springer.

Netzer, O., Lattin, J. M., and Srinivasan, V. (2008). A Hidden Markov Model of Customer Relationship Dynamics. *Marketing Science*, 27(2):185–204.

Paas, L. J. ., Vermunt, J. K. ., and Bijmolt, T. H. . A. . (2007). Discrete Time , Discrete State Latent Markov Modelling for Assessing and Predicting Household Acquisitions of Financial Products. *Journal of the Royal Statistical Society*, 170(4):955–974.

Park, S. and Gupta, S. (2011). A Regime-Switching Model of Cyclical Category Buying. *Marketing Science*, 30(3):469–480.

Poulsen, C. S. (1990). Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing*, 7(1):5–19.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Romero, J., van der Lans, R., and Wierenga, B. (2013). A partially hidden markov model of customer dynamics for CLV measurement. *Journal of Interactive Marketing*, 27(3):185–208.

Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2014). Model Selection Using Database Characteristics : Developing a Classification Tree for Longitudinal Incidence Data. *Marketing Science*, 33(2):188–205.

Schweidel, D. a., Bradlow, E. T., and Fader, P. S. (2011). Portfolio Dynamics for Customers of a Multiservice Provider. *Management Science*, 57(3):471–486.

Schweidel, D. A. and Knox, G. (2013). Incorporating Direct Marketing Activity into Latent Attrition Models. *Marketing Science*, 32(3):471–487.

Scott, L. (2002). Bayesian methods for hidden Markov models, recursive computing in the 21st century. *J.Amer.Statist.Assoc.*, 97:337–351.

Shachat, J. and Wei, L. (2012). Procuring Commodities: First-Price Sealed-Bid or English Auctions? *Marketing Science*, 31(2):317–333.

Shi, S. W. and Wedel, M. (2013). Information Acquisition During Online Decision Making : A Model-Based Exploration Using Eye-Tracking Data. *Management Science*, (January 2016).

Shi, S. W. and Zhang, J. (2014). Usage Experience with Decision Aids and Evolution of Online Purchase Behavior. *Marketing Science*, 33(6):871–882.

Stüttgen, P., Boatwright, P., Monroe, R. T., Stüttgen, P., Boatwright, P., and Monroe, R. T. (2012). A Satisficing Choice Model. *Marketing Science*, (July 2016).

van der Lans, R., Pieters, R., and Wedel, M. (2008a). Competitive Brand Salience. *Marketing Science*, 27(5):922–931.

van der Lans, R., Pieters, R., and Wedel, M. (2008b). Eye-Movement Analysis of Search Effectiveness. *Journal of the American Statistical Association*, 103(482):452–461.

Wedel, M., Pieters, R., and Liechty, J. (2008). Attention switching during scene perception: How goals influence the time course of eye movements across advertisements. *Journal of Experimental Psychology-Applied*, 14(2):129–138.

Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden Markov model. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*.

Zhang, J. Z., Netzer, O., and Ansari, A. (2014). Dynamic Targeted Pricing in B2B Relationships. *Marketing Science*, 33(October):317–337.

Zhang, J. Z., Watson, G. F., Palmatier, R. W., and Dant, R. P. (2016). Dynamic Relationship Marketing. *Journal of Marketing*.

Zucchini, W. and MacDonald, I. (2009). Hidden Markov Models for Time Series. *Hand, The*.

Appendix

A HMM papers in marketing

Table 15: Non-comprehensive list of marketing papers using HMM (Netzer et al., 2017)

Article	Heterogeneity in: transition probabilities	Heterogeneity in: state dependent probabilities	Number of states	Number of states selected based on:
Poulsen (1990)	No	No	2	Estimation
Brangule-Vlagsma et al. (2002)	No	No	6	Estimation
Liechty et al. (2003)	No	No	2	Theory
Montgomery et al. (2004)	Yes	Yes	2	Estimation
Du and Kamakura (2006)	No	No	13	Estimation
Paas et al. (2007)	No	No	9	Estimation
Moon et al. (2007)	No	Yes	2	Theory
Netzer et al. (2008)	Yes	No	3	Estimation
Wedel et al. (2008)	No	No	2	Theory
van der Lans et al. (2008b)	No	Yes	2	Theory
van der Lans et al. (2008a)	No	Yes	2	Theory
Montoya et al. (2010)	Yes	Yes	3	Estimation
Ebbes et al. (2010)	No	No	3	Estimation
Schweidel et al. (2011)	Yes	No	4	Estimation
Park and Gupta (2011)	No	Yes	2	Theory
Li et al. (2011)	Yes	Yes	3	Estimation
Kumar et al. (2011)	No	Yes	3	Estimation
Lemmens et al. (2012)	No	Yes	3	Estimation
Stüttgen et al. (2012)	Yes	Yes	2	Theory
Ansari et al. (2012)	Yes	Yes	2	Theory
Shachat and Wei (2012)	No	No	3	Theory
Ascarza and Hardie (2013)	Yes	Yes	3	Estimation
Romero et al. (2013)	No	Yes	7	Estimation
Shi and Wedel (2013)	No	No	3	Estimation
Luo and Kumar (2013)	Yes	Yes	3	Estimation
Mark et al. (2013)	No	No	4	Estimation
Mark et al. (2014)	No	No	3	Estimation
Shi and Zhang (2014)	Yes	No	3	Estimation
Zhang et al. (2014)	Yes	Yes	2	Estimation
Schwartz et al. (2014)	Yes	Yes	2	Theory
Ma et al. (2015)	Yes	Yes	3	Estimation
Zhang et al. (2016)	No	No	4	Estimation
Ascarza et al. (2018)	Yes	Yes	3	Estimation

B HMM model specification

We detail the parameter transformations and priors for the general model specification in Section 2.1.

B.1 State dependent probabilities

We define the increasing vector of probabilities p with $p_1 \leqslant p_2 \leqslant \ldots \leqslant p_S$, such that

$$p_s = \sum_{k=1}^{s} u_k$$

with $\vec{u}$ a vector in the $S-$dimensional simplex

$$(\vec{u} \in \mathbb{R}^{S+1},\ u_s \geqslant 0,\ \sum_s u_s = 1)$$

B.2 Transition probabilities

We parametrize $Q_{is\bullet}$, the s'th row of the transition matrix for customer i using a vector $\gamma_{is} \in \mathbb{R}^{S-1}$ where:

$$q_{iss'} = \begin{cases} \frac{\exp(\gamma_{iss'})}{\sum\limits_{k=1}^{S-1} \exp(\gamma_{isk})+1} & \text{if } s' \in \{1, \ldots, S-1\} \\ \frac{1}{\sum\limits_{k=1}^{S-1} \exp(\gamma_{isk})+1} & \text{if } s' = S \end{cases}$$

B.3 Heterogeneity (hierarchical component)

We denote $\theta_i = \Big(\gamma_{i11}, \ldots, \gamma_{iSS-1}\Big)$, and we a Gaussian distribution to account for unobserved heterogeneity,

$$\theta_i \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

where μ_θ and Σ_θ are the mean vector and covariance matrix respectively.

We further parametrize to fasten computation. We define $\tau = \sqrt{\text{diag}(\Sigma_\theta)}$ the vector of standard deviations, Ω the correlation matrix for covariance Σ, and L_Ω its corresponding Cholesky decomposition. We also define z_i the standardized individual level parameters, such that,

$$\theta_i = \mu_\theta + \mathrm{diag}(\tau) \cdot L_\Omega \cdot z_i$$

with $z_i \sim \mathcal{N}(0, I)$.

In addition, to ensure a positive τ, we parametrize each component k of τ using the vector $\widetilde{\tau}$ such that,

$$\tau_k = 2.5 \cdot \tan(\widetilde{\tau}_k).$$

B.4 Priors

B.4.1 Initial probabilities

For $\pi \in \mathbb{R}^S$ we use uniform priors on the simplex,

$$\pi \sim Dirichlet(1).$$

B.4.2 State dependent probabilities probabilities

For $u \in \mathbb{R}^{S+1}$ we use uniform priors on the S-dimensional simplex,

$$u \sim Dirichlet(1).$$

B.4.3 Transition probabilities

We use the following priors for μ_θ, $\widetilde{\tau}$, and L_Ω

$$\mu_\theta \sim \mathcal{N}(0, 2 \cdot I)$$

$$\widetilde{\tau}_k \sim U(0, \pi/2)$$

Note that in this case, π stands for the number $\pi \approx 3.14\ldots$.

$$L_\Omega \sim \text{LKJ\_corr\_cholesky}(2)$$

where LKJ_corr_cholesky$(\cdot)$ stands for the distribution ofthe cholesky factor of a mtrix that distributes according to an LKJ correlation distribution (LKJ_Corr$(\Omega|c) \propto \det(\Omega)^{c-1}$)

C HMM Stan code

```
data {
  int<lower=1>          S;                    // num states
  int<lower=1>          I;                    // num customers
  int<lower=1>          T;                    // num calibration periods
  int<lower=0>          Y[I,T];               // observed behavior
  int<lower=1>          T_val;                // num periods for validation
  int<lower=0>          Y_val[I,T_val];       // observed behavior for validation
  int<lower=1>          T_test;               // num periods for testing
  int<lower=0>          Y_test[I,T_test];     // observed behavior for testing
  int<lower=1>          K;                    // Binomial number trials
}
parameters {
  matrix[S*(S-1), I]                      z;          // indep normals for transition prob.
  cholesky_factor_corr[S*(S-1)]           L_Omega;    // cholesky corr. for transition prob.
  row_vector[S*(S-1)]                     mu_theta;   // mean of unconstrained transition prob.
  vector<lower=0,upper=pi()/2>[S*(S-1)] tau_unif;   // scaled variance of transition prob.
  simplex[S+1]                            mu_unif_phi; // transformed state dependent prob.
  simplex[S]                              ppi;        // initial prob.
}
transformed parameters{
  matrix[I,S*(S-1)]   theta;                          // individual transition parameters
  vector<lower=0>[S*(S-1)] tau;                       // variance of transition prob.
  matrix[S,S] log_Q[I];                               // log transition prob.

  for (s in 1:S*(S-1))
    tau[s] = 2.5 * tan(tau_unif[s]);
  theta = rep_matrix(mu_theta,I) + (diag_pre_multiply(tau,L_Omega) * z)';
  for (i in 1:I){
    for (k in 1:S){
      row_vector[S-1] ttheta;
      ttheta = theta[i,((S-1)*(k-1)+1):((S-1)*k)];
```

```
      log_Q[i,k,]=to_row_vector(log_softmax(to_vector(append_col(ttheta,0))));
      // add the zero at the end, so vector is size S,
      //    and compute softmax (multinomial logistic link)
    }
  }
}
model {
  to_vector(z) ~ normal(0, 1);
  L_Omega ~ lkj_corr_cholesky(2);
  mu_theta ~ normal(0, 2);
  mu_unif_phi~dirichlet(rep_vector(1.0,S+1));
  {
    vector[S] mu_phi = head(cumulative_sum(mu_unif_phi),S);
    // Forward algorithm
    for (i in 1:I){
      real acc[S];
      real gamma[T,S];
      for (k in 1:S)
          gamma[1,k] = log(ppi[k])+ binomial_lpmf(Y[i,1]|K,mu_phi[k]);
      for (t in 2:T) {
        for (k in 1:S) {
          for (j in 1:S){
              acc[j] = gamma[t-1,j] + log_Q[i,j,k] + binomial_lpmf(Y[i,t]|K,mu_phi[k]) ;
          }
          gamma[t,k] = log_sum_exp(acc);
        }
      }
      target +=log_sum_exp(gamma[T]);
    }
  }
}
generated quantities{
  matrix[S*(S-1),S*(S-1)] Omega;       // Correlation of transitions
```

```
matrix[S,S] Q[I];                     // Transitions
vector[I] log_like;                   // In sample log-likelihood
vector[I] log_like_val;               // Validation log-likelihood
vector[I] log_like_test;              // Holdout log-likelihood
int<lower=1,upper=S> y_star[I,T];     // Most likely state (Viterbi)
real log_p_y_star[I];                 // Probability of most likely state (Viterbi)
vector[S] mu_phi = head(cumulative_sum(mu_unif_phi),S);  // State dependent prob.

Omega = L_Omega * L_Omega';
Q = exp(log_Q);

  for (i in 1:I){
      int bacS_ptr[T, S];
      real best_logp[T, S];
      real best_total_logp;
      real acc[S];
      real gamma[T+T_val+T_test,S];

      // Forward algorithm and Viterbi algorithm
      for (k in 1:S){
          gamma[1,k] = log(ppi[k])+ binomial_lpmf(Y[i,1]|K, mu_phi[k]);
          best_logp[1, k] = gamma[1,k];
      }
      for (t in 2:T) {
          for (k in 1:S) {
              best_logp[t, k] = negative_infinity();
              for (j in 1:S){
                  real logp;
                  logp = best_logp[t-1, j]+ log_Q[i,j,k] +
                         binomial_lpmf(Y[i,t]|K, mu_phi[k]);
                  acc[j] = gamma[t-1,j] + log_Q[i,j,k] + binomial_lpmf(Y[i,t]|K, mu_phi[k]);
                  if (logp > best_logp[t, k]) {
                      bacS_ptr[t, k] = j;
```

```
                best_logp[t, k] = logp;
            }
        }
    gamma[t,k] = log_sum_exp(acc);
    }
}
log_p_y_star[i] = max(best_logp[T]);
for (k in 1:S)
    if (best_logp[T, k] == log_p_y_star[i])
        y_star[i,T] = k;
for (t in 1:(T - 1))
    y_star[i,T - t] = bacS_ptr[T - t + 1,y_star[i,T - t + 1]];
log_like[i] =log_sum_exp(gamma[T]);

// Validation
    for (k in 1:S){
        for (j in 1:S)
            acc[j] = gamma[T,j] + log_Q[i,j,k] +
                      binomial_lpmf(Y_val[i,1]|K,mu_phi[k]) ;
        gamma[T+1,k] = log_sum_exp(acc);
    }

    for (t in 2:T_val) {
        for (k in 1:S) {
            for (j in 1:S){
                acc[j] = gamma[T+t-1,j] + log_Q[i,j,k] +
                          binomial_lpmf(Y_val[i,t]|K,mu_phi[k]);
            }
            gamma[T+t,k] = log_sum_exp(acc);
        }
    }
    log_like_val[i] =log_sum_exp(gamma[T+T_val])-log_like[i];
```

```
        // Holdout
            for (k in 1:S){
                for (j in 1:S)
                    acc[j] = gamma[T+T_val,j] + log_Q[i,j,k] +
                              binomial_lpmf(Y_test[i,1]|K,mu_phi[k]) ;
                gamma[T+T_val+1,k] = log_sum_exp(acc);
            }

            for (t in 2:T_test) {
                for (k in 1:S) {
                    for (j in 1:S){
                        acc[j] = gamma[T+T_val+t-1,j] + log_Q[i,j,k] +
                                  binomial_lpmf(Y_test[i,t]|K,mu_phi[k]);
                    }
                    gamma[T+T_val+t,k] = log_sum_exp(acc);
                }
            }
            log_like_test[i] =log_sum_exp(gamma[T+T_val+T_test])-
                              log_like[i]-log_like_val[i];
    }
}
```

D Results for no heterogeneity in the number of states

We are interested in showing that the results from Section 3.2 and Section 4.2 are purely driven by the heterogeneity in number of states. Therefore, we estimate the model using only customers from Segment A.

D.1 Model selection

Table 16: Number of states selection when all individuals have 3 states

	In sample			**Validation**	**Holdout**
Number of States	LL	LMD	WAIC	LL	LL
HMM					
1	-7336.31	-7337.24	14680.51	-2065.25	-2349.62
2	-5920.08	-6010.19	12042.73	-1667.51	-1888.95
3	**-5790.17**	**-5924.84**	**11882.09**	**-1635.87**	**-1858.56**
4	-5780.35	-5922.29	11880.63	-1636.62	-1860.35
5	-5778.06	-5911.16	11879.52	-1637.35	-1864.72
MHMM					
3	-5803.49	-5937.56	11897.02	-1636.01	-1858.60

D.2 Parameter estimates

Table 17: 3-state HMM parameter estimates for $\lambda = 0\%$

(a) Initial state probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\pi_\mathbf{s}$	0.408	0.320	0.272
	[0.314 0.504]	[0.187 0.452]	[0.178 0.369]

(b) State dependent probabilities posterior mean and posterior 95% intervals

	1	**2**	**3**
$\mathbf{p_s}$	0.114	0.524	0.892
	[0.094 0.134]	[0.471 0.576]	[0.871 0.913]

(c) Population mean transition matrix posterior mean and posterior 95% intervals

		1	**2**	**3**
Q	**1**	0.794	0.147	0.060
		[0.718 0.857]	[0.073 0.231]	[0.026 0.106]
	2	0.078	0.745	0.178
		[0.025 0.148]	[0.623 0.848]	[0.096 0.277]
	3	0.063	0.125	0.812
		[0.036 0.093]	[0.048 0.213]	[0.733 0.875]