Structural Heterogeneity in Hidden Markov Models: Heterogeneity in the Number of States

Nicolas Padilla[1], Ricardo Montoya[2], and Oded Netzer[3]

[1]*London Business School*
[2]*School of Management, Pontificia Universidad Catolica de Chile*
[3]*Columbia Business School, Columbia University*

February 12, 2026

ABSTRACT

Hidden Markov models (HMMs) are widely used in marketing to capture customer behavior dynamics by modeling transitions among latent states such as attention, web search, customer relationships, and purchase intent. While HMMs typically incorporate parametric heterogeneity, they often assume a fixed number of latent states across customers, which we refer to as "structural homogeneity." This paper investigates the bias and economic impact of this assumption, examines whether parameter heterogeneity can offset it, and explores alternative approaches to address it.

Using extensive Monte Carlo simulations and two empirical applications, we show that assuming structural homogeneity can lead to identification issues, misinterpretation of customer behavior, and suboptimal marketing decisions with significant economic costs. In a pharmaceutical marketing context, we demonstrate that ignoring state heterogeneity can underestimate the impact of detailing and sampling by 56.2% and 41.2%, respectively. Targeting based on models that account for structural heterogeneity significantly improves profitability compared to policies based on standard HMMs.

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

Hidden Markov models (HMMs) have been widely used across disciplines to model sequences of observations generated by transitions among a set of latent states. In marketing, HMMs have gained substantial traction: our review identifies at least 40 published studies employing HMMs in the marketing literature.[1] In these applications, HMMs have been primarily used to model how customers (and sometimes firms) evolve over time across a set of latent states. A key difference between the applications of HMMs in marketing and other fields is that in the latter, HMMs model 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).[2]

Accounting for heterogeneity across customers has both statistical and substantive merits. Statistically, accounting for parameter heterogeneity across customers in dynamic models is crucial to disentangle heterogeneity from dynamics properly (Heckman et al., 1981). Substantively, consumer heterogeneity is fundamental to the marketing concept, providing the basis for market segmentation, targeting, and positioning (Kamakura et al., 1996). While most HMMs in marketing allow for parametric heterogeneity, these models assume the same number of latent states across customers. This assumption, which we refer to as “structural homogeneity” forces the model to assume that all customers share the same underlying structure. Whereas parametric heterogeneity has been extensively studied, structural heterogeneity has received little attention in the marketing literature, especially in dynamic models.

In this paper, we examine the potential biases, identification problems, and economic detriments introduced by assuming that all customers have the same number of states in the presence of structural heterogeneity. We then study how alternative modeling approaches, both those that explicitly account for such heterogeneity and those that do not, handle settings in which different customers may have different numbers of HMM states. We argue and demonstrate that the common practice of estimat-

[1] See Table 1, adapted from Netzer et al. (2017), for a selected list of HMMs in marketing and management.

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

ing a single model with the same number of states for all customers, while flexibly accounting for heterogeneity in model parameters, can lead to identification problems, misleading inferences, and suboptimal marketing decisions. Importantly, this problem arises even when the vast majority of customers have the same number of states. We further illustrate the economic impact of these misleading inferences using two empirical marketing applications.

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 who transition only between low and medium levels of relationship. Estimating a common model for both groups with three HMM states can lead to parameter identification problems for customers who transition between only two states, even when parametric 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 their behavior in multiple 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 the next section.

We first explore, through an extensive simulation exercise, the extent of the problem of ignoring structural heterogeneity under different conditions of heterogeneity, dynamics, data length, and the impact of marketing actions. We confirm that ignoring heterogeneity in the number of states when such heterogeneity exists can lead to an identification problem, particularly for 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 recommend more expensive models with more states. Therefore, one 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 individual-level estimates even for customers with the same number of states as 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 population shrinkage. Moreover, we show that the common practice of reporting population-level estimates (e.g., Netzer et al., 2008; Montoya et al., 2010; Kumar et al., 2011; Schweidel et al., 2011; Ansari et al., 2012), despite accounting for parametric heterogeneity, could lead to misleading insights in the presence of structural heterogeneity, much like in cases where heterogeneity distributions are multimodal or highly skewed. Finally, we show that, in the presence of structural heterogeneity, targeting policies derived from an HMM that ignores this heterogeneity are substantially less effective than those based on the correct model.

We demonstrate that a more flexible parametric heterogeneity, such as discrete–continuous mixtures (e.g., Allenby et al., 1998; Kappe et al., 2018), in which the segments are structurally identical, but differ in the values of their response parameters within a unified model framework, cannot address the issues raised by structural heterogeneity. On the other hand, models that account for structural heterogeneity, such as a mixture of HMMs (MHMM), go beyond the discrete-continuous specification of heterogeneity, accommodating structural differences among segments (e.g., one segment has three states, and the other has two states), which can be framed as a particular case of structural heterogeneity (Kamakura et al., 1996; Yang and Allenby, 2000; Gu and Yang, 2010). We demonstrate that such models better recover the underlying dynamics, increase predictive performance, and improve the effectiveness of marketing actions.

We examine the implications of ignoring and accounting for structural heterogeneity in the number of states on relevant marketing actions and business outcomes using two empirical applications. First, we investigate a marketing resource allocation problem using pharmaceutical data on physicians' prescription behavior. We find that ignoring structural heterogeneity leads to biased estimates of the effectiveness of marketing actions (detailing and sampling). Furthermore, we find that properly accounting for structural heterogeneity in HMMs can yield a substantial increase in incremental profits relative to a targeting policy that does not account for structural heterogeneity. In a second empirical application, in the context of a role-playing online game, we demonstrate that not accounting for structural heterogeneity may lead to an overestimation of player dynamics due to mixing heterogeneity and dynamics. This, in turn, leads to inaccurate predictions of customer churn and lifetime value. These applications highlight the importance of accounting for structural heterogeneity in marketing contexts in which capturing individual-level heterogeneity and dynamics is crucial.

This paper contributes to the marketing literature by identifying a previously overlooked source of bias in a widely used modeling framework for empirical marketing research: HMMs estimated on panel data. We show that imposing structural homogeneity in the number of latent states, even when allowing for rich parametric heterogeneity, can lead to identification problems, distorted inference, and suboptimal marketing decisions. More broadly, we clarify the distinction between structural and parametric heterogeneity in dynamic models and assess the effectiveness of alternative approaches for accommodating each. Substantively, our findings highlight the risks of ignoring structural heterogeneity when evaluating the impact of marketing actions in dynamic settings, with direct implications for targeting, resource allocation, churn prediction, and customer lifetime value estimation.

The rest of the paper is organized as follows. We first briefly outline the HMM, and illustrate using a numerical example the identification problem that can arise from ignoring structural heterogeneity. We then describe several simulation experiments that show the conditions under which the assumption of homogeneity in the number of states can lead to identification problems, misleading inferences, and suboptimal marketing decisions. We next use data from a pharmaceutical marketing application and an online role-playing game to examine the identification problem and illustrate its impact on marketing decisions and business outcomes. We conclude with a discussion of the main contributions of this paper and directions for future research.

STRUCTURAL AND PARAMETRIC HETEROGENEITY IN HMMS

As detailed in Table 1 there are dozens of papers applying HMMs to understand customers' underlying latent dynamics. Examples of these include: 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; Liberali and Ferecatu, 2022), 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; Ferecatu and De Bruyn, 2022) and households lifecycle stages (Du and Kamakura, 2006). HMMs have also been 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 is not consumers but other units where accounting for heterogeneity is important. For example, Ebbes et al. (2010) looked at how the competitive landscape of different firms (banks) changed over time. Moon et al. (2007) used an HMM to uncover firms' latent competitive promotions. Lemmens et al. (2012) examined evolving segments of countries. Montoya and Gonzalez (2019) analyzed product sales for multiple SKUs to identify on-shelf out-of-stocks.

In dynamic models such as HMMs, it is crucial to account for customer heterogeneity. The marketing literature has studied two sources of customer heterogeneity in consumer models: *parametric* and *structural* heterogeneity. Parametric heterogeneity refers to individual differences in the HMM parameter values, e.g., differences in brand or attribute preferences and responses to the marketing mix. Structural heterogeneity refers to differences in the structure of the model, accounting for the fact that some customers might follow different decision processes (Kamakura et al., 1996; Allenby et al., 1998; Yang and Allenby, 2000; Gu and Yang, 2010). In the context of HMMs, heterogeneity[3] can be conceptualized using these two sources. Parametric heterogeneity in HMMs refers to differences in customers' initial probabilities, transition probabilities, and parameters governing state-dependent distributions. Structural heterogeneity in HMMs, on the other hand, encompasses variations beyond these parameters. Including, for example, differences in the functional form of the heterogeneity distributions, a type of non-structural heterogeneity that is common to many statistical models with potentially heterogeneous units. Unlike distributional variations, the number of latent states is a fundamental structural component of HMMs that dictates the model's complexity and its ability to capture various underlying dynamic patterns. The researcher's decision on the number of states has a substantial impact on the

[3]Even though dynamics could also be understood as "within-individual heterogeneity" over periods, we use the term heterogeneity in this paper to refer to customer (cross-sectional) heterogeneity.

Table 1: A list of marketing papers using HMMs

Article	Parametric Heterogeneity in: transition probabilities	Parametric 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
Zhang et al. (2017)	Yes	Yes	3	Estimation
Hui (2017)	Yes	Yes	3	Theory
Ascarza et al. (2018)	Yes	Yes	3	Estimation
Lee et al. (2018)	Yes	Yes	3	Estimation
Gopalakrishnan et al. (2021)	No	No	2	theory
Ferecatu and De Bruyn (2022)	Yes	Yes	3	Theory
Liberali and Ferecatu (2022)	No	No	2,3	Estimation
Naumzik et al. (2022)	Yes	No	3	Estimation
Zhang et al. (2022)	Yes	Yes	2	Estimation
Alós-Ferrer and Garagnani (2023)	Yes	Yes	4	Theory

Note: A non-comprehensive list. Adapted and expended from (Netzer et al., 2017).
The table does not include papers that used HMM solely as a benchmark for the modeling approach proposed in the focal paper.

inferences regarding customer dynamics, the model's predictive ability, and the resulting recommended managerial decisions.

To fully account for customer heterogeneity in HMMs, researchers would ideally estimate a separate model for each customer. However, due to the limited number of observations per customer, researchers typically pool the information across customers when accounting for parametric heterogeneity. Indeed, of the 42 marketing papers that employ an HMM reported in Table 1, 28 accounted for some form of unobserved heterogeneity. The most common approaches for capturing heterogeneity in the HMM parameters across time series (customers) are hierarchical Bayesian modeling (e.g., Netzer et al., 2008; Schwartz et al., 2014; Scott, 2002) and latent class (e.g., Schweidel and Knox, 2013), or a combination of both (Kappe et al., 2018). However, while the aforementioned studies have often accounted for heterogeneity in the HMM parameters (allowing for *parametric heterogeneity*), 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, researchers in marketing and other fields typically assume that all agents transition among the same number of states, hence neglecting *structural heterogeneity.*

Before further illustrating the implications of the assumption of common HMM states across customers, we briefly outline a typical HMM specification which we use in the remainder of this paper.

Traditional HMM specification

We assume that a customer i ($i = 1, \ldots, I$) exhibits behavior Y_{it}, (e.g., purchase) at each time period ($t = 1, \ldots, T$). The customer behavior at each period is governed by the customer's state in that period ($Z_{it} = s, \; s = 1, \ldots, S$)[4] and, potentially, can also be affected by a set of marketing actions X_{it}.

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 $P(Z_{i1} = s) = \pi_{is}$. The vector of initial state distributions for all states is defined as $\boldsymbol{\pi_i} = (\pi_{i1}, \ldots, \pi_{iS})$.

2. The transitions: Customer i transitions from state s at time t to state s' at time $t+1$ with probability $P(Z_{it+1} = s' | Z_{it} = s, X_{it}) = q_{itss'}$.

[4]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.

3. The state-dependent behavior: The probability of observing behavior $Y_{it} = y_{it}$ for customer i at time t given that they are in state s at time t is $P(Y_{it} = y_{it}|Z_{it} = s, X_{it}) = m_{it|s}(y_{it})$.

The state-dependent behavior can be represented by any discrete or continuous probability distribution function. However, for the purpose of the simulation and empirical applications 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. Thus, 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 period t, Z_{it} and marketing actions X_{it}. Thus, we can write the state-dependent behavior as:

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

where $p_{is}(X_{it})$ is the probability of purchasing given that the customer i is at state s. We parameterize the vector $\mathbf{p}_i = \{p_{i1}, p_{i2}, \ldots, p_{iS}\}$ to increase in its components ($p_{i1} \leq p_{i2} \leq \ldots \leq p_{iS}$) to avoid the label-switching problem (See Web Appendix A.1).

We can define the state-dependent behavior in a matrix form using the following diagonal matrix: $M_{it} = \text{diag}\begin{pmatrix} m_{it|1} & \ldots & m_{it|S} \end{pmatrix}$. Finally, we define the transition matrix $\mathbf{Q}_{it}$ that contains the transition probabilities between time $t-1$ and time t as:

$$\mathbf{Q}_{it} = \begin{bmatrix} q_{it11} & \cdots & q_{it1S} \\ \vdots & \ddots & \vdots \\ q_{itS1} & \cdots & q_{itSS} \end{bmatrix}. \tag{2}$$

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

$$q_{iss'} = \begin{cases} \frac{\exp(\gamma_{iss'})}{\sum_{\ell=1}^{S-1} \exp(\gamma_{is\ell})+1} & \text{if } s' \in \{1, \ldots, S-1\} \\ \frac{1}{\sum_{\ell=1}^{S-1} \exp(\gamma_{is\ell})+1} & \text{if } s' = S \end{cases} \tag{3}$$

For the specification with time-varying marketing actions X_{it} in the transition matrix, we use an ordered logistic specification where transition probabilities $q_{its\ell}(X_{it}) = P(z_{it+1} = \ell|z_{it} = s, X_{it})$ are specified by:

$$q_{its1} = \frac{\exp\left(\gamma_{is1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}$$

$$q_{its\ell} = \frac{\exp\left(\gamma_{isl} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is\ell} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)} - \frac{\exp\left(\gamma_{is\ell-1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is\ell-1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)} \qquad \ell = 2, \ldots, S-1 \quad (4)$$

$$q_{itsS} = 1 - \frac{\exp\left(\gamma_{isS-1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{isS-1} - X'_{it} \cdot \boldsymbol{\rho}_{is}\right)}$$

and $\gamma_{is1} < \gamma_{is2} < \ldots < \gamma_{isS-1}$.

Likelihood. Putting the three components together, we integrate out the hidden states (Zucchini and MacDonald, 2009) to obtain the likelihood function of observing the behavior of customer i over time,

$$\mathcal{L}_i = p(Y_i | \boldsymbol{\pi_i},\ \mathbf{Q}_i,\ \mathbf{p}_i) = \boldsymbol{\pi_i} M_{i1} \left(\prod_{t=2}^{T_i} Q_{it} M_{it} \right) \mathbf{1}', \tag{5}$$

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

Numerical example

Similar to the stylized example briefly described in the introduction, let's 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:

$$\mathbf{p}^A = \begin{bmatrix} p_L & p_M & p_H \end{bmatrix} = \begin{bmatrix} 0.1 & 0.5 & 0.9 \end{bmatrix}, \quad \mathbf{p}^B = \begin{bmatrix} p_L & p_M \end{bmatrix} = \begin{bmatrix} 0.1 & 0.5 \end{bmatrix}.$$

That is, when a customer in Segment A is in the *Low*, *Medium*, or *High* state, they have 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 transition matrix $\mathbf{Q}^A$. In contrast, consumers in Segment B can only transition between the low and medium states with transition matrix $\mathbf{Q}^B$.

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

Because the researcher often does not observe 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 preference heterogeneity in the HMM parameters. Suppose the selected model (using some model selection criteria) corresponds to the three-state HMM.[5] This model constrains all customers to have three HMM states, but it allows customers to have different values for $\mathbf{p}$ and $\mathbf{Q}$. If the researcher allows for a flexible heterogeneity structure, the HMM with three states should correctly capture 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 two 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 ($\widehat{\mathbf{p}}^B$) as those of the customers in Segment A, but estimate a transition matrix ($\widehat{\mathbf{Q}}^B$) such that customers in Segment B never transition to the high state, such that

$$\widehat{\mathbf{p}}^B = \begin{bmatrix} 0.1 & 0.5 & 0.9 \end{bmatrix}, \text{ and } \widehat{\mathbf{Q}}^B = \begin{bmatrix} 0.85 & 0.15 & 0 \\ 0.15 & 0.85 & 0 \\ \hat{q}_{HL} & \hat{q}_{HM} & \hat{q}_{HH} \end{bmatrix}.$$

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

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

can be valid estimates to rationalize the data. Similarly, the value of $p_H = 0.9$ for customers in Segment B 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

[5] As we will show in Section "What happens if ignoring structural heterogeneity in HMMs?", 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.

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 a purchase probability of roughly 50%. Specifically:

$$\widehat{\mathbf{p}}^B = \begin{bmatrix} 0.1 & 0.5-\varepsilon & 0.5+\varepsilon \end{bmatrix}, \text{ and } \widehat{\mathbf{Q}}^B = \begin{bmatrix} 0.85 & \hat{q}_{LM} & \hat{q}_{LH} \\ \sim 0.15 & \hat{q}_{MM} & \hat{q}_{MH} \\ \sim 0.15 & \hat{q}_{HM} & \hat{q}_{HH} \end{bmatrix}.$$

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 transition matrices below can be valid estimates from the model.

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

In sum, the model for Segment B customers (with two states) is not identified when a three-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 the 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 Section "What happens if ignoring structural heterogeneity in HMMs?", we report the results of an extensive simulation study to assess the degree of the identification problem illustrated by this example.

Model estimation

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 their posterior distributions. We use a hierarchical specification to account for parametric heterogeneity. Because the researcher does not know a priori the customer membership to each segment, the researcher typically pools the data of all customers and estimates an HMM with heterogeneous parameters but with a common number of states across customers. Throughout the paper, we specify the initial state probabilities as homogeneous, while allowing the state-dependent and transition probabilities 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 as the vector of all global 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 Web Appendices A for further details of the model specification and prior distributions.

WHAT HAPPENS IF IGNORING STRUCTURAL HETEROGENEITY IN HMMS?

To address the challenge researchers face in empirical applications, where the "true" number of states each customer transitions among remains unobservable, we have designed an extensive simulation exercise. In the simulation, we know the exact number of states for each customer, which enables us to explore the implications of assuming homogeneity in the number of HMM states. Therefore, the main 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 numbers of states. Specifically, we study the impact on the potential lack of identification of the model parameters and model prediction. We also explore how different data-generating processes and model assumptions may affect these outcomes. In addition to the bias in parameter estimation and its corresponding insights, we extend the simulation exercise to explore the potential implications of making marketing decisions based on these biased models. We describe the main elements and findings of the simulation exercise next (for a full description of the simulation and results, see Web Appendix B).

Simulation design

Data generating process. We simulate data from data-generating processes (DGP) under multiple scenarios. We first describe our baseline scenario and then how alternative scenarios deviate from the baseline.

Baseline scenario. We assume that there are two customer segments following the model (and values) described in the numerical example discussed in the previous section. The behavior of the first segment (Segment A) follows an HMM with three states (Low, Mid, and High), whereas 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). We define λ as the percentage of Segment B customers among all customers, which we set at $\lambda = 50\%$ for this baseline scenario. We generate individual-level state-dependent and transition probabilities with continuous parametric (non-structural) heterogeneity within each segment through multivariate Gaussian distributions on the unconstrained parameters.

Additional scenarios. In addition to the *Baseline* scenario, we generate several alternative scenarios that deviate from the *Baseline* across multiple factors:

1. the degree of structural heterogeneity (using 1 and 2 states for Segments B and A, respectively; instead of 2 and 3 states);
2. the structural heterogeneity mixture percentage ($\lambda = 70\%$, 80%, 90%, 95%, and 98%; instead of $\lambda = 50\%$);[6]
3. the non-structural parametric heterogeneity among state-dependent and transition probabilities (*No heterogeneity* and *High heterogeneity*; instead of Low heterogeneity);
4. the degree of dynamics (*High dynamics*; instead of Low dynamics); and
5. the length of the time series data used for calibration (*Long* series of $T = 40$ periods; instead of

[6] We do not examine cases where $\lambda < 50\%$, as lower values of λ would clearly lead the model selection criteria to favor a higher number of states.

$T = 25$).

We also assess in this simulation the implications of structural heterogeneity on marketing decisions, where we examine the targeting strategies derived from models' predictions following an intervention.

To robustly analyze each of the scenarios, given the inherent uncertainty of the estimation procedure and the sampling error from the data-generating process, we generate 100 dataset replications for each scenario. For each replication, we estimate the specified HMM using a fully Bayesian framework as described next. Overall, across conditions and replications, we run more than 4,200 simulated scenarios, each involving a separate model estimation.

The implications of ignoring structural heterogeneity

We explore three main consequences that may arise from ignoring structural heterogeneity: (i) bias towards selecting HMMs with more states, (ii) misleading insights from population estimates, and (iii) bias in parameter estimates even for those estimated with the correct number of states.

Bias towards more states. The first step in estimating an HMM is determining the number of states that best represent customers' dynamics. The common procedure is to estimate several HMMs, each with a different number of states, and choose the best model specification based on some model selection criteria. Specifically, HMM applications in marketing have consistently relied on multiple information criteria computed using the calibration sample, such as the deviance information criterion (DIC), the log marginal density (LMD), and the Watanabe-Akaike information criterion (WAIC). However, typical in-sample penalized fit criteria in Markov-switching models, such as HMMs, often fail to accurately determine the correct number of states (Smith et al., 2006). In fact, we show in Web Appendix B.1 that in the context of these HMMs, even when all customers have the same number of states, these criteria either underpenalize (DIC) or overpenalize (LMD and WAIC) for model complexity. Conversely, out-of-sample criteria, such as the expected log predictive density in the validation set (Val-LL) and the (posterior mean) root mean square error of the outcome in the validation set (Val-RMSE), successfully recover the correct simulated number of states when there is no structural heterogeneity. Consequently, throughout the paper, we use these two model selection criteria to determine the number of states in each replication.

In the *Baseline* scenario, we find that when the proportion of customers transitioning among three states (Segment A) and two states (Segment B) are equal ($\lambda = 50\%$), both model selection criteria Val-LL and Val-RMSE choose the model with 3 states on all 100 replications.

Next, we explore how many states the model selection criteria would recommend when the majority of the customers belong to Segment B (the segment with fewer states). Would the model selection criteria recommend the model that fits the majority of the customers? To investigate these questions, we vary the percentage of Segment B customers (λ) progressively from 50% to 98% and calculate the proportion of replications where the 3-state HMM is favored over the 2-state HMM for each criterion. Table 2 shows that, even when most customers truly transition among two states, the 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, Val-LL and Val-RMSE still suggest that the best model is the 3-states HMM across 92% and 80% of the replications, respectively; which is 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 customers.[7]

Parameter recovery: Interpreting aggregate estimates. The numerical example in the previous section suggests that if customers have different numbers of states, then the parameters of an estimated HMM that assumes the same number of states across customers may be unidentified. We investigate whether allowing for flexible parametric heterogeneity across consumers can mitigate the problem. Table 3 shows the parameter estimates for the *Baseline* scenario. Specifically, we report the posterior mean and the 95% posterior credible interval of: (a) the initial probabilities, (b) the average across the population of the state-dependent probabilities, (c) and the average of the transition probabilities across the

[7] We note that the phenomenon of model selection criteria favoring more flexible (and computationally expensive) specifications in the presence of structural heterogeneity is not unique to HMMs. In fact, we replicate this result in a similar example using a hierarchical Bayesian Linear Regression in Web Appendix D.

Table 2: Percentage of replications in which model selection criteria favor a 3-state HMM over a 2-state HMM, by the share of 2-state customers (λ).

Proportion with 2 states	Replications with 3 states as best model	
	Val-LL	Val-RMSE
Baseline		
$\lambda = 50\%$	100%	100%
Alternative structural heterogeneity mixture		
$\lambda = 70\%$	100%	100%
$\lambda = 80\%$	99%	100%
$\lambda = 90\%$	97%	99%
$\lambda = 95\%$	92%	80%
$\lambda = 98\%$	44%	25%

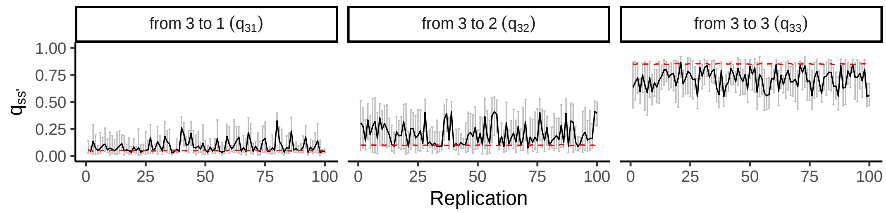
population.

Table 3: 3-state HMM parameter estimates for one replication of the *Baseline* scenario

	π_s	$\mathbf{Q}$			$\mathbf{p_s}$
		State 1	State 2	State 3	
State 1	0.354 [0.247, 0.457]	0.797 [0.738, 0.845]	0.177 [0.120, 0.241]	0.027 [0.007, 0.058]	0.118 [0.085, 0.150]
State 2	0.491 [0.342, 0.625]	0.124 [0.088, 0.168]	0.806 [0.731, 0.858]	0.070 [0.031, 0.124]	0.500 [0.435, 0.560]
State 3	0.155 [0.071, 0.261]	0.058 [0.029, 0.141]	0.308 [0.051, 0.510]	0.634 [0.445, 0.868]	0.854 [0.790, 0.905]

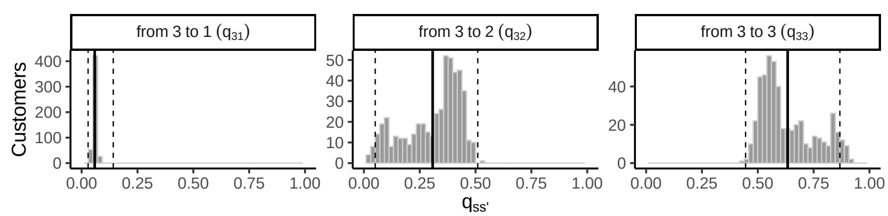
From the estimates of p_s, we conclude that the model recovers well the conditional behavior associated with the true three states with probabilities 0.1, 0.5, and 0.9 (see $\mathbf{p}^A$ in the "Numerical example" subsection). This result is robust across replications. However, when we explore the population mean transition probabilities in Table 3, we find that these averages combine the transition matrices of customers with three states and those with two states who rarely transition to a third state. The estimated $\widehat{q}_{13}$ and $\widehat{q}_{23}$ fall between the true transition probabilities for customers with three states ($q_{13} = 0.05$, $q_{23} = 0.15$) and a hypothetical zero probability for customers with two states. As mentioned in the "Numerical example" subsection, the parameters in the third row of the transition matrix (transitions from State 3) are unidentified for customers with only two states. Figure 1 shows that this result is consistent across replications (i.e., the estimated posterior mean of q_{33}, in black, is consistently below the true values, in red, across replications).

Figure 1: Baseline condition. Recovery of population transition probabilities from the third state across 100 replications. The black line represents the posterior mean, and the gray bars the 95% posterior credible interval. True simulated values for Segment A customers with three states are in red.



Thus, the population-average transition matrix in Table 3 and Figure 1, despite accounting for parametric heterogeneity in transition probabilities, does not represent the true values for either three-state or two-state customers. This can lead to erroneous inferences. For example, one may wrongly conclude that the firm is not doing a good job in keeping its most engaged customers active, as the most profitable state (State 3) appears less sticky than the two other states. Figure 2 further explores this, showing the heterogeneity of the transition probabilities for the same replication as in Table 3. It reveals a multimodal behavior: for some customers, the firm is doing very well in keeping them in the high-purchase state, while others have a highly dynamic third state.

Figure 2: Histogram of the individual posterior means of the transition probabilities from the third state for one replication of the *Baseline* scenario.



Arguably, for deriving insights about behavior, making predictions, and informing marketing decisions, only individual-level estimates are relevant in a hierarchical model. However, it is important to note that many marketing studies using HMMs continue to report and draw insights from population-level transition matrices rather than individual-level 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),[8] Table 7 in Zhang et al. (2014) (see Figure E.1 in the

[8] 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.

Web Appendix E). Similarly to settings where the parametric heterogeneity distribution is multimodal or highly skewed, this common practice may yield misleading insights when structural heterogeneity is present. Accordingly, we encourage researchers to closely examine and report heterogeneity in HMM parameters (as in Figure 2) and to exercise caution when interpreting aggregate parameters in the presence of substantial heterogeneity.

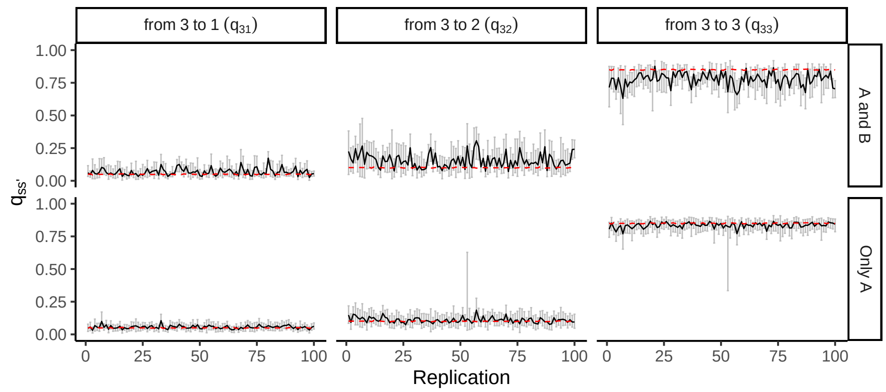
Bias for correctly estimated customers. While population-level transition matrices can yield misleading insights, one might expect the underlying individual-level estimates, such as those in Figure 2, to remain unbiased, at least for customers in Segment A who truly have three states. We show this is not the case. Specifically, we find that parameter estimates are biased even for customers with truly three states whose behavior is modeled using a correctly specified three-state HMM.

Because segment membership is known in the simulation, we can investigate the estimated transition matrix, Q_i, separately for customers in each segment. Additionally, we benchmark the estimates from the *Baseline* scenario against those from the correctly specified model for Segment A customers, namely, a model estimated using only Segment A customers (*Only A*). Doing so allows us to contrast the models' estimates with and without structural heterogeneity for the same set of customers.

We start visualizing this bias by averaging the individual-level estimates for customers in Segment A from the Baseline model. The top panel in Figure 3 shows the posterior mean and 95% credible interval of the average individual transition probabilities from State 3 across Segment A customers for each replication. For q_{33}, the posterior mean (in black) consistently underestimates the true values (in red). While the posterior uncertainty resulting from the estimation may hide this downward bias for any specific replication, we can observe this consistent bias across replications. In contrast, this bias disappears for a model estimated using only customers in Segment A (without structural heterogeneity), despite using less data (see the bottom panel of Figure 3). We observe similar bias for q_{32}.

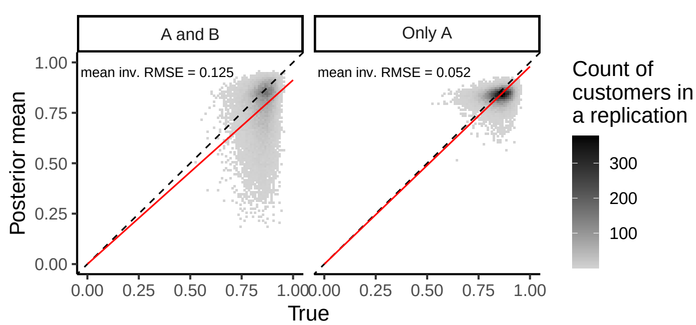
The probabilities in the top panel of Figure 3 are averages across all Segment A customers, thus the individual-level differences may be larger. Figure 4 shows the true vs. the individual-level posterior mean of each customer in Segment A across all replications. The left panel shows the individual recovery when all customers, *A and B*, are used to estimate the HMM, while the right panel shows the

Figure 3:
Recovery of Segment A transition probabilities from the third state across 100 replications. The top panel shows Segment A estimates when the model is estimated using all data (*A and B*). The bottom panel shows Segment A estimates when the model is estimated using only Segment A customers (*Only A*). The black line represents the posterior mean, and the gray bars the 95% posterior credible interval. True simulated values are in red.



recovery when *Only A* customers are used to estimate the HMM. The figure indicates that the individual estimates of those with three states are downward biased when estimating the HMM without considering the underlying structural heterogeneity. This individual bias disappears when estimating the model using only Segment A data without structural heterogeneity.

Figure 4: True vs. posterior mean of the individual q_{i33} estimates for Segment A. The left panel corresponds to the case when both segments are used in the estimation, whereas the right panel corresponds to the case when only Segment A customers are used in the estimation. Each observation represents a customer in a replication. A darker color indicates a higher count.

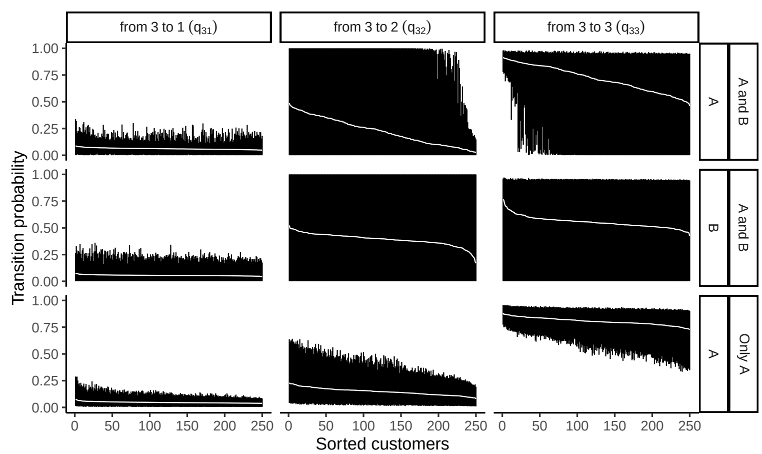


What if the proportion of customers with two states increases? We expect that having more customers lacking identification in their transition matrix would increase the bias in the parameters for the customers with 3 states. Our results confirm that the bias for three-state customers grows as the proportion of customers with unidentified parameters increases (see Web Appendix B.3.2).

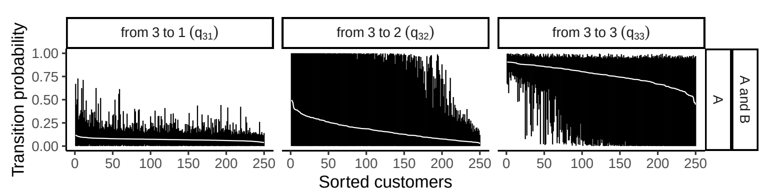
In a Bayesian framework, a lack of identification leads to high posterior uncertainty in the parameter estimates. Consequently, to further illustrate this identification problem, we choose one replication from the *Baseline* scenario and show in Figure 5a the individual level 95% posterior credible intervals of the third row of the transition matrix. First, the middle panel shows that the posterior uncertainty of Segment B customers is exceptionally high; the posterior distributions fully cover the $[0,1]$ interval for q_{22} and q_{33}, which is clear evidence of lack of identification. Second, this lack of identification is pooled to Segment A customers (top panel) as the majority of customers have high uncertainty for q_{22} and q_{33}. These wide credible intervals prevent any meaningful inference regarding these transitions. Finally, the uncertainty reduces substantially when estimating the model with only Segment A customers (bottom panel), despite using half the observations.

Figure 5: Posterior mean and 95% confidence interval for individual transition probabilities from state 3. In each panel, customers are sorted in decreasing order of their posterior mean. The shaded area represents the 95% intervals and the white line represents the posterior mean.

(a) HMM. The top two panels show Segments A and B, respectively, when the model is estimated jointly (*A and B*). The bottom panel shows Segment A when estimated using only Segment A (*Only A*).



(b) RCM-HMM. The panel shows Segment A when the model is estimated jointly (*A and B*).



These results also suggest that pronounced individual-level posterior uncertainty in the transition probabilities may indicate potential structural heterogeneity. Although such uncertainty should not be interpreted as a definitive diagnostic, routinely inspecting individual-level credible intervals as part of post-estimation checks can help researchers detect situations in which the imposed state structure may be inconsistent with the underlying behavioral patterns.

In sum, estimating the HMM in the presence of structural heterogeneity in the number of states biases the transition matrix estimates, even for customers whose model is correctly specified. This bias disappears when the model is estimated only on customers with no structural heterogeneity (Segment A customers only). As detailed in the "Numerical example" subsection, the bias for three-state customers arises from pooling Segment A and Segment B customers in a three-state HMM, and the lack of identification in the last row of the transition matrix for Segment B customers.

Alternative simulated scenarios. We next examine variations of the Baseline DGP to explore boundary conditions for the bias arising from ignored structural heterogeneity. We summarize below the main findings and report the full analysis in Web Appendix B.3:

1. increasing the proportion of customers with fewer states amplifies the bias,
2. greater within-segment dispersion in customer parameters (continuous parametric heterogeneity) exacerbates the bias, whereas eliminating such dispersion attenuates it, though not completely,
3. extending the length of the time series, within ranges typical of marketing applications, has only a limited effect on reducing the bias,
4. when states are highly dynamic (i.e., less "sticky"), the bias is modestly reduced but remains present,
5. the bias is substantially attenuated when customer dynamics are weak, as in settings with a mix of static and low-dynamic customers (1 and 2 states).

Effect of structural heterogeneity on marketing decisions

Identification and bias in the estimated model's parameters are likely to lead to suboptimal marketing decisions. In this section, we use the simulation framework to assess how ignoring structural heterogeneity affects these decisions.

To motivate this exercise, consider a salesforce compensation scheme (see e.g., Daljord et al., 2016; Nahm et al., 2022). The company has two segments of salespeople, differentiated by the geographic regions in which they operate. In high-demand geographies, salespeople transition among three sales levels (low, medium, and high) based on their effort. In contrast, those in low-demand geographies transition only between low and medium sales states. To incentivize higher sales across both segments, the company is considering implementing an experiment following a compensation scheme where for salespeople in the treatment group/month bonuses are awarded based on surpassing the previous year's sales rather than meeting a fixed quota (e.g., Misra and Nair, 2011). For salespeople in the three states' geographies, this incentive scheme is likely to be highly effective for those in the low state and potentially for those in the medium state. However, it may discourage salespeople in the high state, as they cannot easily surpass their previous year's performance. For the same reasons, in regions with two states, the scheme would incentivize salespeople in the low state but disincentivize those in the medium state, which represents the highest state in these regions.

We incorporate the effect of the incentive scheme into the transition matrix in the simulation analysis. Specifically, we model exposure to the incentive in period t using a binary indicator, which enters the HMM transition probabilities through an ordered logit specification (Equation 4; see Table 4 for the simulated parameter values). We assume that the company implements a randomized experiment to learn salespeople's responsiveness to this marketing intervention. In each period t $(t = 1, \ldots, 40)$, the company targeted 50% of the salespeople (independently across periods). Additional details of the simulation design are provided in the Web Appendix B.6.

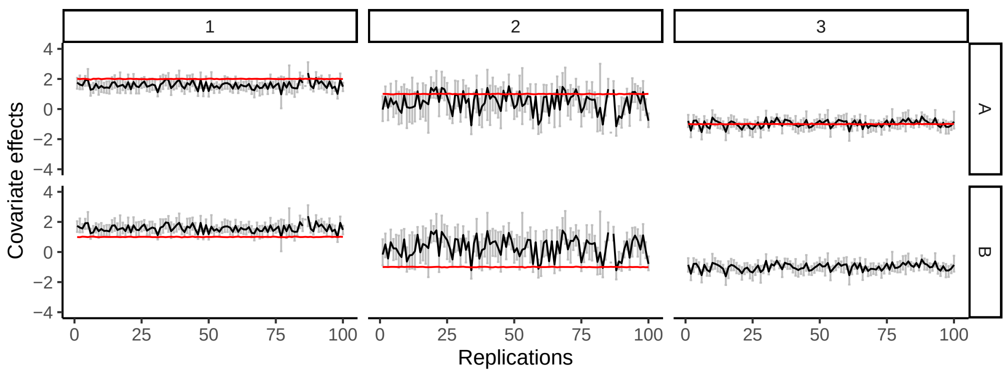
Table 4: Simulated segment-level transition matrices with incentive effects.

Segment		Without incentive $X = 0$			With incentive $X = 1$		
	from/to	1	2	3	1	2	3
A	1	0.85	0.10	0.05	0.44	0.30	0.26
	2	0.05	0.80	0.15	0.02	0.66	0.32
	3	0.05	0.10	0.85	0.13	0.19	0.68
B	1	0.85	0.15	.	0.68	0.32	.
	2	0.15	0.85	.	0.32	0.68	.

We study the performance of the traditional HMM on two dimensions: (1) the recovery of the marketing effects and (2) the implied targeting decisions.

Recovery of the marketing effects. Figure 6 shows that the state-of-the-art HMM consistently underestimates the effect of the marketing action for individuals with the correct number of states (Segment A) in the low and medium states across replications. At the same time, the HMM significantly overestimates the incentive's impact for individuals with the wrong number of states (Segment B). For the transition probability estimates, see Table B.15 in the Web Appendix.

Figure 6: Recovery of covariate effects for the HMM across 100 replications. Segments A and B are shown in the top and bottom rows, respectively. Each column represents a state. The black line represents the posterior mean, and the gray bars the 95% posterior credible interval on each replication. True simulated values are in red.



Targeting decisions. We assess the managerial consequences of ignoring structural heterogeneity by defining simple targeting rules based on the model's estimated parameters and then evaluating their performance under the true DGP. To ensure that the contrast between the HMM-based policies and the optimal targeting is not confounded by sampling or estimation error, we also estimate a benchmark model that replicates the structure of the true DGP (True-Structure model) where we separately estimate an HMM with two states for Segment B and an independent HMM with three states for Segment A, and derive the corresponding targeting policy for this correctly specified model. For each model, we compute an individual-level expected lift for each salesperson, defined as the expected difference in cumulative outcomes over the evaluation horizon when the incentive is given versus withheld. Specifically, we implement a one-time incentive at period $\bar{T} = 40$ and evaluate its effect over the subsequent $T_e = 12$ periods using posterior predictive expectations under the estimated model. Targeting decisions are therefore based on model-implied lifts, whereas the actual outcomes used to assess performance are simulated using the true underlying parameters. We provide details on how we compute the expected lift in Web Appendix B.6.3.

We consider two targeting policies: a *Positive Lift* policy, which targets all salespeople whose estimated lift is positive and therefore captures the model's ability to estimate the level of the true lift, and a *Top 10%* policy, which targets the 10% of salespeople with the highest estimated lift and thus isolates the model's ability to correctly rank individuals regardless of errors in the magnitude of the lift. Given that the data are simulated, we can benchmark the model-based targeting policies against the optimal policy computed using the true parameters. Using this benchmark, we evaluate each policy along two dimensions. First, we compare the set of salespeople targeted by each model-based policy with those who should be targeted under the optimal policy. Second, we compute the expected lift of each policy and express it relative to the expected lift attained under the optimal policy. Table 5 summarizes the results. The third and fourth columns report the average precision and recall across replications for both the HMM and the True-Structure model.[9] The final two columns report the average expected lift of each policy and its percentage shortfall relative to the optimal lift.

Table 5: Precision, recall, and lift of the model-based policies for different targeting scenarios across replications.

Targeting Policy	Model used	Precision	Recall	Lift	Dist. from optimal
Positive	HMM	0.846	0.858	87.73	8.46%
	True-Structure model	0.890	0.888	91.34	4.68%
Top 10%	HMM	0.679	0.679	47.69	15.75%
	True-Structure model	0.868	0.868	54.76	3.14%

We find that, when targeting all salespeople with a positive expected lift, the HMM correctly targets 84.6% of those it selects and identifies 85.8% of the salespeople who should be contacted. Nevertheless, the resulting HMM-based policy nearly doubles the distance to the optimal policy, relative to the policy based on the True-Structure model. When targeting the top 10%, performance deteriorates further: the HMM-based policy correctly identifies only 67.9% of the salespeople who should be targeted, yielding a total lift that is 15.75% below the optimal. Most of this gap reflects the consequences of ignoring structural heterogeneity rather than estimation error, as the True-Structure model is only

[9]Precision is the percentage of targeted salespeople who should be targeted; recall is the percentage of salespeople who should be targeted that the model successfully targets.

3.14% away from the optimal policy.

In sum, we demonstrated that ignoring the presence of structural heterogeneity leads to computationally more expensive models, incorrect aggregate inferences, and bias on the parameters of even those customers for which the model is correctly specified. Additionally, we find that these limitations have consequences for the targeting of marketing actions. In the next section, we analyze what researchers can do to address structural heterogeneity in HMMs.

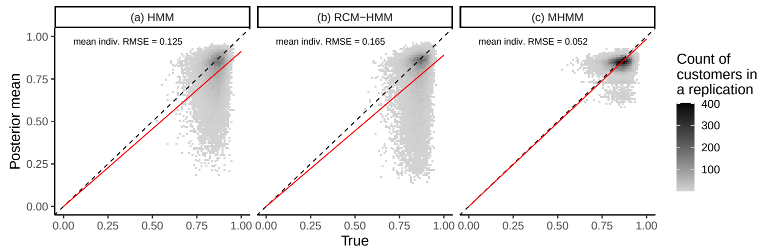
WHAT CAN BE DONE TO ADDRESS STRUCTURAL HETEROGENEITY?

More flexible parametric heterogeneity

One potential approach to addressing structural heterogeneity in a traditional HMM is to introduce greater flexibility in parametric heterogeneity, for example, by allowing for discrete mixture components. Accordingly, we consider a finite Gaussian mixture in which individual-level parameters are drawn from a multimodal prior. Let $\boldsymbol{\theta}_i$ denote the vector of individual-level parameters in the HMM. We specify a Gaussian mixture as the mixing distribution, $\boldsymbol{\theta}_i \sim \sum_{c=1}^{C} \psi_c \cdot \mathcal{N}(\boldsymbol{\mu}_{\theta,c}, \Sigma_{\theta,c})$ where $\boldsymbol{\mu}_{\theta,c}$ and $\Sigma_{\theta,c}$ are the mean vector and covariance matrix respectively for component c, and ψ_c is the mixture probability such that $\psi_c \geq 0, \quad \sum_c \psi_c = 1$. This specification yields a discrete-continuous representation of parameter heterogeneity and it is known as the Random Coefficient Mixture HMM (RCM-HMM; Kappe et al., 2018).

We find that the RCM-HMM exhibits the same downward bias in q_{33} as the standard HMM for Segment A customers with truly three states (see panels (a) and (b) of Figure 7), indicating that introducing discrete–continuous parametric heterogeneity does not resolve the structural heterogeneity problem. Consistent with this pattern, individual-level results show poor recovery of transition probabilities and, in some settings, such as high continuous heterogeneity or highly dynamic state switching, the RCM-HMM performs even worse than the HMM (Table B.6 in the Web Appendix). Moreover, posterior uncertainty at the individual level does not improve relative to the HMM (see Figure 5b). Full results, including additional figures and tables, are reported in Web Appendix B.4.1. Finally, this specification does not improve predictions. In fact, in most scenarios, the state-of-the-art HMM outperforms the RCM-HMM in out-of-sample metrics. See Table B.7 in the Web Appendix for details.

Figure 7: True vs. posterior mean of the individual q_{i33} for Segment A. We show the HMM in Panel (a), the RCM-HMM in Panel (b), and the MHMM in Panel (c). Each observation represents a customer in a replication. A darker color indicates a higher count.



Mixture of HMMs (MHMM)

Alternatively, one could explicitly account for structural heterogeneity in the model. The simplest approach to achieve that is to use a mixture of HMMs (Van de Pol and Langeheine, 1990; Helske and Helske, 2019), where each HMM component has a different number of states.

The idea behind the MHMM is to create a model that combines HMMs with varying numbers of states and that allows each customer to probabilistically belong to an HMM process with the number of states that best reflects their behavior. Following Equation (5) we can define the likelihood of an HMM with S_k states as $\mathcal{L}_{i,\text{HMM}}(\mathbf{Y}_i|\boldsymbol{\pi}_i^k, \mathbf{p}_i^k, \mathbf{Q}_i^k, S_k)$, where $\boldsymbol{\pi}_i^k$, $\mathbf{p}_i^k$, and $\mathbf{Q}_i^k$ are the corresponding parameters of customer i for an HMM with S_k states. The MHMM assumes that each customer has a probability, ω_k, of belonging to an HMM with S_k states. Let $k = \{1, \ldots, K\}$ be a class that follows an HMM with S_k states.[10] Then, the likelihood of the MHMM can be written as:

$$\mathcal{L}_i(\mathbf{Y}_{i,1:T}|\{\boldsymbol{\pi}_i^k, \mathbf{p}_i^k, \mathbf{Q}_i^k\}_{k=1}^K, \omega) = \sum_{k=1}^{K} \omega_k \cdot \mathcal{L}_{i,\text{HMM}}(\mathbf{Y}_{i,1:T}|\boldsymbol{\pi}_i^k, \mathbf{p}_i^k, \mathbf{Q}_i^k, NS_i = S_k) \tag{6}$$

Similar to the standard HMM, we estimate the MHMM within a hierarchical framework to allow for continuous parametric heterogeneity across customers. This hierarchical structure enables individual-level variation in the initial-state, transition, and state-dependent parameter probabilities for each HMM component. We rely on HMC, implemented in Stan, to obtain posterior draws of all model

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

parameters (see the "Model estimation" subsection).[11]

It is useful to note that the MHMM can be represented as a single, larger HMM whose transition matrix is block diagonal, with each block corresponding to the transition matrix of one HMM component and zeros in the off-diagonal blocks. In principle, one could attempt to estimate such a model by simply increasing the number of states in a conventional unconstrained HMM to account for structural heterogeneity. However, without explicitly imposing the block-diagonal restrictions, that is, without forcing the off-diagonal transition probabilities to be zero, the expanded HMM becomes even more weakly identified, which exacerbates the identification issues we discuss in our simulations (see Web Appendix C). Thus, researchers must impose this structure, either through the MHMM specification or by enforcing zero off-diagonal transitions in an enlarged HMM. Simply enlarging the HMM state space without these restrictions does not constitute a valid alternative for modeling structural heterogeneity. Consequently, a restricted-expanded HMM formulation simply offers an alternative way to compute the MHMM likelihood.

We study whether the MHMM can mitigate the biases reported in our simulation analysis. Table 6 reports population estimates of the 2-state and 3-state components of the MHMM for one replication of the *Baseline* scenario (see Web Appendix B.5 for all replications). Examining the state-dependent behavior (p_s), and the transition probabilities (Q) of the two-state component (Tables 6a and 6c) and the three-state component (Tables 6b and 6d), the MHMM effectively captures the behavior of customers in each segment. Indeed, most estimated parameters are close to the true parameters (reported in the "Numerical example" subsection), and the true parameters fall within the corresponding posterior credible intervals.[12]

Moving to individual-level parameters and across replications, panel (c) in Figure 7 shows that the individual downward bias of q_{33} for Segment A customers, evident in the HMM, consistently disappears with the MHMM. Indeed, we find that the RMSE between the true and the estimated parameters drops by more than half for the q_{i33} estimates (from 0.125 for the HMM to 0.052) for the MHMM.

[11]The `seqHMM` R package (Helske and Helske, 2019) includes a restricted version of the MHMMs. However, it only permits time-invariant covariates, and it captures parametric heterogeneity only through discrete latent classes, rather than through a flexible continuous hierarchical distribution.

[12]Only p_2 for those with 2 states is marginally outside the 95% posterior credible interval.

Table 6: MHMM state-dependent probabilities and transition estimates for 2-state and 3-state components one replication of the *Baseline* scenario. Population posterior means and posterior 95% intervals.

(a) 2-state component: State-dependent probs.

	1	2
p_s	0.095	0.566
	[0.064 0.125]	[0.507 0.614]

(b) 3-state component: State-dependent probs.

	1	2	3
p_s	0.114	0.514	0.916
	[0.074 0.162]	[0.428 0.593]	[0.892 0.939]

(c) 2-state component: Transition matrix

		1	2
Q	**1**	0.833	0.167
		[0.775 0.877]	[0.123 0.225]
	2	0.168	0.832
		[0.128 0.218]	[0.782 0.872]

(d) 3-state component: Transition matrix

		1	2	3
Q	**1**	0.792	0.170	0.039
		[0.676 0.878]	[0.084 0.282]	[0.008 0.083]
	2	0.066	0.772	0.162
		[0.015 0.134]	[0.672 0.843]	[0.100 0.243]
	3	0.048	0.116	0.837
		[0.019 0.085]	[0.056 0.191]	[0.770 0.882]

With the MHMM, we can also assess how often the model classifies customers into the component of the model with the correct number of states. In the baseline scenario, the MHMM correctly classifies 63.8% and 70.7% of the customers with two or three states, respectively. This classification accuracy improves with more data per customer (longer time series), reaching 74.9% and 83.0% for the two and three states, respectively, for the 40-period scenario. In Web Appendix B.5, we provide robustness checks showing that all these results hold consistently across replications and across all simulation scenarios described in the "Data generating process" subsection.

Finally, we analyze the MHMM's predictions of customers' future behavior compared to those of the traditional HMM. Evaluating the model's predictive performance is important, as these predictions are the key input for customer valuation (CLV) and retention strategies (churn prediction). To do so, for each model, we compute the expected log predictive density (Test-LL) and root mean squared error (Test-RMSE) on a set of test holdout periods. To obtain a robust comparison between the models, we compute the proportion of replications in which the MHMM provides better predictive results than the HMM. We find that the MHMM delivers superior predictive performance in the vast majority of cases: in 99% of the replications of the *Baseline* scenario, the MHMM achieves both higher Test-LL and lower Test-RMSE than the HMM, and across all simulation scenarios this proportion ranges between 85% and 100% (see Table B.13 in Web Appendix B.5.3).

Targeting simulation

We begin by assessing whether the RCM-HMM or the MHMM can more accurately recover marketing effects in the targeting simulation. Table 7 reports the RMSE of individual-level covariate effects for the RCM-HMM, the MHMM and the standard HMM. First, we find that the RCM-HMM does not improve the recovery of marketing effects of the HMM. Second, relative to both the HMM and the RCM-HMM, the MHMM performs comparably for customers in Segment A, while achieving substantially lower RMSE for customers in Segment B, indicating improved recovery of covariate effects for this segment. Estimated transition probabilities with and without incentives for all models are reported in Table B.15 in Web Appendix B.6.2.

Table 7: RMSE for covariate effects by segment and state for the HMM and MHMM

Model	Segment A			Segment B	
	St. 1	St. 2	St. 3	St. 1	St. 2
HMM	0.488	0.737	0.296	0.611	1.371
RCM-HMM	0.487	0.821	0.355	0.587	1.290
MHMM	0.505	0.786	0.323	0.368	0.392

We next assess whether the MHMM's superior recovery of marketing effects translates into improved targeting decisions. Table 8 extends Table 5 by incorporating results from the RCM-HMM and the MHMM. Consistent with the results in Figure 7, the RCM-HMM performs at par with the HMM in targeting: its precision and recall are nearly identical, and its distance from the optimal policy is almost indistinguishable from that of the HMM. More interestingly, we find that the MHMM substantially improves targeting performance under both targeting policies. When targeting all salespeople with a positive expected lift, the MHMM achieves 87.0% precision and 87.6% recall, bringing its performance much closer to that of the True-Structure model. The performance gap between HMM and MHMM widens with a more selective targeting policy: the MHMM attains an accuracy of 79.0%, compared to 67.9% for the HMM. Across both policies, the MHMM closes a substantial portion of the performance gap associated with the HMM, delivering lifts that are only 1.2–7.8 percentage points below those of the True-Structure model. Overall, these results indicate that a model that (probabilistically) accounts for structural heterogeneity closes most of the gap with the True-Structure model.

Table 8: Precision, recall, and lift of the model-based policies for different targeting scenarios across replications.

Targeting Policy	Model used	Precision	Recall	Lift	Dist. from optimal
Positive	HMM	0.846	0.858	87.73	8.46%
	RCM-HMM	0.844	0.860	87.72	8.47%
	MHMM	0.870	0.876	89.57	6.54%
	True-Structure Model	0.890	0.888	91.34	4.68%
Top 10%	HMM	0.679	0.679	47.69	15.75%
	RCM-HMM	0.690	0.690	48.03	15.14%
	MHMM	0.790	0.790	51.56	8.84%
	True-Structure Model	0.868	0.868	54.76	3.14%

The asymmetry of misspecification: including vs. ignoring structural heterogeneity

Although the MHMM clearly outperforms traditional specifications when structural heterogeneity is present, and it nests the model with no structural heterogeneity, there is a risk that the increased model complexity will overfit the data when structural heterogeneity is absent. Conceptually, this concern may be asymmetric: failing to account for structural heterogeneity could be highly detrimental, whereas allowing for it when absent may entail comparatively smaller costs.

To evaluate this potential asymmetry, in this section, we compare the ability of the MHMM and HMM-based models to recover the true parameters under DGPs that either include or omit structural heterogeneity. We consider two HMM-based DGPs. The first is a standard HMM in which all customers transition among three states with continuous parametric heterogeneity. The second follows a discrete–continuous heterogeneity structure akin to the RCM-HMM, with two customer segments, each comprising three states with different parameters and additional continuous heterogeneity within segments (see details in Web Appendix B.7).

Table 9 reports the average RMSE between the true and estimated transition probabilities of the HMM, the RCM-HMM and 3-state component of the MHMM across 100 replications, for three DGPs, an HMM, an RCM-HMM, and an MHMM.[13]

We can observe in Table 9 that, as expected, the model that matches the DGP recovers the model

[13]When the DGP is the MHMM, we only compute RMSEs for each model using those customers with truly three states (Segment A).

Table 9: Transition probabilities RMSE for different estimated models and data-generating processes

Estimated Model	DGP		
	HMM 1 Segment	RCM-HMM 2 Segments	MHMM 2 Components
HMM	0.042	0.055	0.072
RCM-HMM	0.053	0.051	0.087
MHMM	0.042	0.056	0.053

parameters best (has the lowest RMSE). More importantly, we can see that when the DGP comes from an HMM, the MHMM performs similarly to the HMM and significantly better than the RCM-HMM. When the DGP follows an RCM-HMM, the MHMM and HMM perform similarly and about 8-9% worse than the true DGP. However, consistent with our previous analyses, when the DGP includes structural heterogeneity, models that include parametric heterogeneity but not structural heterogeneity (HMM and RCM-HMM) perform 35% and 63% worse than the MHMM (last column of Table 9). Web Appendices B.4.2 and B.5 present similar results with respect to out-of-sample predictions. Thus, while the MHMM is capable of capturing parametric heterogeneity (continuous or discrete-continuous), the HMM (and RCM-HMM) cannot capture structural heterogeneity.

Overall, the empirical evidence points to a clear asymmetry: allowing for structural heterogeneity when it is absent entails little cost, but neglecting it when it is present compromises both parameter recovery and predictive accuracy. Thus, accounting for structural heterogeneity provides a robust safeguard against misspecification that traditional HMM formulations lack. The empirical applications that follow, therefore, includes the MHMM alongside a standard HMM to assess the incremental value of modeling structural heterogeneity.

EMPIRICAL APPLICATIONS

We illustrate the risk of ignoring structural heterogeneity and the value of accounting for structural heterogeneity through two empirical applications. The first application examines the impact of marketing actions on physicians' prescription behavior in a pharmaceutical context. The second application explores gamers' dynamics in an online gaming context, focusing on metrics like CLV and churn.

Application 1: Pharmaceutical marketing

Data. Our data include monthly prescription behavior for a random sample of 300 physicians, as well as monthly detailing and sampling activities over 24 months. Detailing involves face-to-face meetings where pharmaceutical representatives present drug information to physicians, while sampling involves providing free drug samples to physicians. Further details are available in Montoya et al. (2010).

HMM and MHMM specifications. We follow the model specification outlined in the "HMM components" subsection to include marketing actions in transitions and state-dependent prescription behavior.[14] Similarly to our simulations in the "Effect of structural heterogeneity on marketing decisions" subsection, we introduce time-varying covariates in the transition matrix using an ordered logistic specification following Equation (4). Additionally, we introduce covariates in the state-dependent behavior using a binomial distribution following Equation (1).

The binomial probability that physician i prescribes the focal drug in month t when at state s is given by $p_{its} = \text{logit}^{-1}(u_i + \phi_s + b_{is}X_{it})$, where u_i captures physician-specific heterogeneity, ϕ_s captures state-specific baseline behavior, and b_{is} captures the effect of the marketing activities (X_{it}) on the physician i's behavior when at state s.

To allow for the most flexible account for parametric heterogeneity for the HMM, we allow for a discrete-continuous mixture of Gaussians heterogeneity in the model parameters.

Selection of HMM states and number of Gaussian mixtures. We considered the first 21 months of the data for model selection. In particular, we used the first 18 months to estimate the different specifications of the HMM and the last three months in that period for validation purposes. Considering both validation expected log predictive posterior density (Val-LL) and the validation root mean square error (Val-RMSE), the best model corresponds to a 3-state HMM with a single Gaussian mixture (we show this in Table G.1 in the Web Appendix G.1). Consequently, we select this specification for further comparison with the MHMM.

[14] Note that, for the most part, we follow closely the specification proposed in Montoya et al. (2010) for these data. However, given our goal of disentangling structural heterogeneity from parametric heterogeneity, we extend such specification and allow for heterogeneity in the intercept of the conditional prescription behavior.

Heterogeneity in the number of states. Since the selected HMM has three states, we estimate an MHMM with 3 components: static, two-state HMM, and three-state HMM. We first investigate the degree of structural heterogeneity implied by the MHMM. According to the posterior membership distribution, we find that 10.4% of the physicians are static and have only one state, 35.3% have two states, and 54.3% have all three states. Thus, even though the HMM assumes that all physicians transition among three states, the MHMM suggests that nearly half of the physicians transition among fewer than three states.

HMM vs. MHMM results.

Physicians' prescription behavior. We first compare how the HMM and MHMM characterize the physicians' behavior conditional on a state.[15] For the HMM, we denote as L, M, and H the latent states with the lowest, medium, and highest prescription propensities ϕ_s, respectively. We use the same notation for the MHMM, with the caveat that static customers only have one state (which we label as "M"), and two-state customers have only low (L) and high (H) states. Table 10 reports the conditional prescription probabilities without marketing actions (No Marketing), with Detailing only (Detailing only) and with Sampling only (Sampling only).[16]

The results from the HMM suggest that, on average, physicians' prescription probability is close to zero in the low state and approximately 4% in the medium state. In the high state, physicians' prescription probability reaches up to 14%. The MHMM, instead, infers the existence of one segment of static physicians with a medium prescription intensity (5.7% prescription probability); a two-state segment with low (0.1%) and high (11.9%) prescription behaviors, and a three-state segment, with low, medium, and high intensity behaviors (0.2%, 4.4%, and 16.4%, respectively). Note that the high state for the 2-state physicians is less intense than the corresponding high state of 3-state physicians (11.9% vs. 16.4%). One of the differences between the implied HMM and MHMM can be verified in the high state. Whereas the MHMM separates the 2-state physicians with a relatively lower pre-

[15] To facilitate the comparison between the two models, we report in the tables the corresponding transformed parameters of each model. The raw parameter estimates are reported in Web Appendix G.2.

[16] As the covariates corresponding to the marketing activities were mean-centered, in Table 10 we illustrate the effect of detailing and sampling at 2 standard deviations from the corresponding raw means.

Table 10: Conditional probabilities of the
3-state HMM vs. MHMM. *L* denotes “Low”, *M* denotes “Medium”, and *H* denotes “High” prescription behaviors, respectively. The three columns “No marketing activities,” “Detailing only,” and “Sampling only” represent the mean share of new prescriptions with no detailing or sampling, with detailing, and with sampling, respectively.

(a) Conditional probabilities of the 3-state HMM

States	No Marketing	Detailing only	Sampling only
L	0.001	0.003	0.005
M	0.038	0.040	0.061
H	0.141	0.145	0.155

(b) Conditional probabilities of the MHMM

Number of states	States	No Marketing	Detailing only	Sampling only
1	M	0.057	0.075	0.065
2	L	0.001	0.004	0.024
	H	0.119	0.141	0.113
3	L	0.002	0.001	0.025
	M	0.044	0.040	0.070
	H	0.164	0.151	0.197

scription intensity (11.9%) from the 3-state physicians with a relatively higher prescription intensity (16.4%), the HMM averages these two populations and characterizes the behavior when in this state with a prescription intensity of 14.1%.

Short-term marketing effectiveness. It is noteworthy that for the HMM, sampling is consistently more effective than detailing across all states. The HMM suggests that sampling maintains its effectiveness regardless of the physician’s state. Conversely, the MHMM suggests a structural heterogeneity reaction to marketing actions. Physicians who are less dynamic in their prescription behavior (static or those with two states) tend to be more reactive to detailing than to sampling. Indeed, detailing is often used to convince physicians to adopt. On the other hand, dynamic physicians (those who transition among three states) tend to be more affected by sampling than detailing. These insights would remain obscured in models that fail to incorporate structural heterogeneity.

Dynamics. Table 11 shows the corresponding transition matrices for the HMM (Table 11a) and 2- and 3-state physicians for the MHMM (Table 11b and Table 11c, respectively). Comparing Table 11a (the HMM) with Table 11c (the three-state component of the MHMM), we see that, possibly due to its inability to capture structural dynamics, the HMM tends to over-emphasize dynamics in the transitions particularly to and from the extreme states. The diagonal elements of transition matrices in both components of the MHMM tend to be higher than those of the HMM, indicating that ignoring structural heterogeneity may lead to over-estimating customer dynamics.

Table 11: Population transition matrices for the 3-state HMM and MHMM. Posterior mean and 95% intervals.

(a) 3-state HMM

	L	**M**	**H**
L	0.688 [0.586, 0.779]	0.246 [0.165, 0.338]	0.066 [0.029, 0.136]
M	0.073 [0.018, 0.144]	0.504 [0.373, 0.630]	0.423 [0.292, 0.579]
H	0.144 [0.082, 0.218]	0.301 [0.224, 0.395]	0.555 [0.436, 0.662]

(b) MHMM - 2-state component

	L	**H**
L	0.707 [0.563, 0.815]	0.293 [0.185, 0.437]
H	0.313 [0.180, 0.459]	0.687 [0.541, 0.820]

(c) MHMM - 3-state component

	L	**M**	**H**
L	0.648 [0.465, 0.813]	0.320 [0.164, 0.463]	0.031 [0.002, 0.098]
M	0.078 [0.026, 0.184]	0.626 [0.485, 0.760]	0.296 [0.149, 0.448]
H	0.098 [0.027, 0.197]	0.337 [0.214, 0.475]	0.565 [0.389, 0.741]

Long-term effects. Table 12 displays the effects of marketing actions on transitions for the HMM and MHMM. These capture the long-lasting effects of detailing and sampling through changes in physicians' latent states. Table 12 shows that for both the HMM and MHMM, detailing is effective in moving physicians away from Low, particularly for those with two states, but not so effective in keeping physicians in the H state. Sampling is more effective than detailing in moving physicians away from L only for 3-state physicians in the MHMM, and it is also effective in keeping those physicians in the H state, an insight that is getting lost in the HMM estimates.

It is important to note that the HMM attenuates the effects of detailing and sampling compared to the corresponding effects in the MHMM. For instance, sampling is not effective in the H state for 2-state physicians but is effective for 3-state physicians. Thus, the effect that the HMM captures seems to be an average of the effects for these two populations.

Table 12: Long-term effects of marketing activities. Posterior means of the population transition matrices for HMM and MHMM

Model / component			No Marketing actions			Detailing only			Sampling only		
			L	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**
3-state HMM		**L**	0.688	0.246	0.066	0.540	0.344	0.116	0.545	0.336	0.119
		M	0.073	0.504	0.423	0.061	0.466	0.473	0.050	0.428	0.521
		H	0.144	0.301	0.555	0.153	0.306	0.541	0.147	0.298	0.555
MHMM	2-state		**L**		**H**	**L**		**H**	**L**		**H**
		L	0.707	.	0.293	0.479	.	0.521	0.661	.	0.339
			.	.	.	.	.	.	.	.	.
		H	0.313	.	0.687	0.297	.	0.703	0.345	.	0.655
	3-state		**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**
		L	0.648	0.320	0.031	0.590	0.370	0.041	0.426	0.492	0.081
		M	0.078	0.626	0.296	0.063	0.600	0.336	0.048	0.498	0.454
		H	0.098	0.337	0.565	0.123	0.369	0.508	0.084	0.310	0.606

The managerial implications of accounting for structural dynamics via the MHMM. It is clear from the reported estimates that accounting for structural heterogeneity via the MHMM leads to different inferences regarding customer dynamics and the impact of marketing actions on these dynamics relative to a model that does not account for structural heterogeneity. However, what managerial implications stem from these differences? In this section, we focus on two key implications: prediction ability and targeting of marketing actions.

Prediction ability. Recall that the first 21 months were used to calibrate the models and the last three periods for test purposes. Table 13 shows the in and out-of-sample performance of the HMM and MHMM. Consistent with the simulation results, the MHMM shows better out-of-sample performance (log-likelihood and RMSE in the test sample). This increased predictive ability of the MHMM over the HMM suggests that if the firm is interested in investing, for example, in high-activity physicians, using the MHMM is likely to lead to better predictions and targeting decisions.

Table 13: Prediction ability: HMM vs. MHMM

		Test	
Model	Number of states	LL	RMSE
HMM	3	-1,640.54	2.22
MHMM	1, 2 and 3	**-1,635.56**	**2.20**

Note: The best model in each column is in bold.

Targeting. To more directly analyze the impact of targeting, we project the predictions for 20 periods beyond the calibration periods and simulate, using the individual-level HMM and MHMM estimates, the effect of a single marketing action on physician prescription behavior. We show three scenarios. First, we assume no marketing actions for all 20 periods (*No Marketing*). Second, we assume detailing activity at period $t = 5$ after the calibration period and no activity otherwise (*Detailing only*). Finally, we assume sampling activity at period $t = 5$ after the calibration period and no activity otherwise (*Sampling only*). We simulate the effect of additional details or samples on the number of prescriptions in the short-term (first month) and long-term (following 14 months). To ensure no residual marketing impact from the calibration period, interventions are applied only in period $T + 5$, with no other interventions before or after this point.

We analyze the impacts of the different scenarios from period $T + 5$ to $T + 20$. For each model (HMM and MHMM), marketing action (detailing only or sampling only), and each time period (short-term, 1 month, and long-term, 14 months), we implement a targeting rule that selects the top 25% of physicians with the highest predicted impact. Results are reported in Table 14 for physicians targeted by both models. The models overlap on 56% of targeted physicians for detailing and 74% for sampling (see Table G.3.1 in Web Appendix G).[17]

Table 14 shows the average increase in the prescription probability across physicians compared to a baseline scenario of no interventions (lift). For instance, the HMM predicts that detailing generates a lift of 1.6% and 0.9% in the short- and long-term, respectively. For the same intervention, the MHMM predicts a lift of 2.9% and 0.9% in the short- and long-term, respectively. That is, considering the predicted total impact, the MHMM suggests an impact 56.2% higher. In the case of sampling, following

[17]The analysis for the subset of physicians targeted by either model (the union) provides similar insights but introduces a targeting dimension that may interact with the policy assessment.

Table 14: Prediction of the impact of targeting of the top 25% of physicians based on each policy over 15 periods after an intervention

Marketing action	Estimated model	Average lift: Short-term (1 month)	Long-term (14 months)	Total (15 months)	Difference v. HMM
Detailing	HMM	0.016	0.009	0.024	–
	MHMM	0.029	0.009	0.038	56.2%
Sampling	HMM	0.036	0.019	0.056	–
	MHMM	0.042	0.037	0.079	41.2%

Note 1: Targeted physicians receive a marketing intervention equal to $\mu + 2\sigma$ of the corresponding action (Detailing or Sampling) and no marketing intervention on the other action. The baseline condition is no detailing or sampling activities.
Note 2: For each marketing intervention, the targeting rule selects the top 25% of physicians with predicted higher impact according to both models.
Note 3: Lift is averaged across physicians. $\tau_i = \sum_t E(\Delta Y_{it})/N_{it}$.

the same analysis, the MHMM suggests an impact 41.2% higher.

The previous analysis demonstrates that targeting detailing or sampling based on MHMM relative to the HMM leads to higher lift. However, that analysis relies on the predictions made by each model. Therefore, to further contrast the managerial implications of accounting for structural heterogeneity in HMMs, we leverage the company's targeting policy in the validation period. Specifically, we computed the expected lift of each model (HMM or MHMM) of the observed allocation of detailing and sampling in the validation period (periods 22-24), and selected the corresponding top 25% with the highest expected lift (vs. a no-targeting policy). We then compare the actual observed behavior of the targeted physicians in the validation period. We consider the prescriptions written by the targeted physicians and the cost of the allocated resources.[18]

The results are shown in Table 15 under the *fixed reach* panel. In this targeting scenario, both models target 75 physicians (top 25%). By comparing their behaviors in the pre (19-21) and post (22-24) targeting periods, the HMM yields \$23,800 whereas the MHMM yields \$39,190 incremental profits. This implies that following the MHMM targeting yields incremental profits that are 64.66%

[18] Following Montoya et al. (2010) we consider that the retail price of a prescription is \$300, the cost of one detail is \$80, and the cost of one sample is \$30.

(=($39,190-$23,800)/$23,800) higher than the suggested HMM's targeting policy.

Table 15: Outcomes of targeted customers based on the HMM vs. MHMM. The top panel shows the scenario where reach is fixed across models to 25%. The bottom panel shows the scenario where the total budget is fixed at the budget spent by the HMM-driven policy, targeting the 25% customers with the highest return.

Model	Customers targeted	Quantity	Outcomes from targeted customers		
			Pre	Post	Difference
		periods:	19-21	22-24	
Fixed reach					
HMM	75	Prescriptions (#)	653	782	129
		Revenue ($)	195,900	234,600	38,700
		Cost ($)	68,910	83,810	14,900
		Profit ($)	126,990	150,790	23,800
MHMM	75	Prescriptions (#)	793	981	188
		Revenue ($)	237,900	294,300	56,400
		Cost ($)	66,270	83,480	17,210
		Profit ($)	171,630	210,820	39,190
Fixed budget					
HMM	75	Prescriptions (#)	630	749	119
		Revenue ($)	189,000	224,700	35,700
		Cost ($)	57,690	57,320	-370
		Profit ($)	131,310	167,380	36,070
MHMM	81	Prescriptions (#)	774	903	129
		Revenue ($)	232,200	270,900	38,700
		Cost ($)	59,300	57,230	-2,070
		profit ($)	172,900	213,670	40,770

In previous analyses, we maintained equal reach between the targeting policies of each model, but their costs may differ. Subsequently, we conducted a second analysis with relatively equal budgets for the targeting policies. We calculated the expected ROI (= lift/cost) for targeting each physician and selected the top 25% with the highest expected ROI based on either the HMM or the MHMM. The budget for the HMM was determined first, and the MHMM targeting was constrained to match this budget. The results, shown in Table 15 under the *fixed budget* panel, indicate that the HMM targets 75 physicians, whereas the MHMM targets 81. Comparing the pre (19-21) and post (22-24) targeting periods, the MHMM yields incremental profits that are 13.03% (=($40,770-$36,070)/$36,070) higher than the HMM.

Overall, the pharmaceutical empirical application highlights that accounting for structural heterogeneity not only leads to better inference and predictions but, more importantly, to better targeting of marketing actions.

Application 2: Online gaming

One possible limitation of the pharmaceutical empirical application is the limited dynamics in physician behavior over a period of 24 months. In this application, we focus on the adoption of online games at the daily level, which may give rise to a higher degree of dynamics. In the context of online games, some of the most important marketing metrics for platforms are customer churn, customer lifetime value, and remaining lifetime play for customer targeting purposes (Chen et al., 2018). We succinctly describe the main results of the second application and provide full details in Web Appendix H.

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. We use the first 40 days for calibration, the next 10 days for validation (and determine the HMM specification), and the last 10 days for testing the predictive ability of the models. Consistent with standard metrics in gaming analytics that focus on retention behavior (Huang et al., 2019), we focus on characterizing and predicting daily gaming behavior.

HMM and MHMM specifications. Following the general description in the "HMM components" subsection, we specify the HMM of gamers' behavior as follows: we estimate the initial probabilities from the data, we parametrize each row of the transition matrix using the softmax function, as in Equation (3). We specify the conditional probabilities using a Binomial distribution with $N = 1$ for the conditional gaming behavior (see Equation 1). To capture systematic changes in conditional gaming across weekdays, we include in the conditional distribution a weekend covariate that indicates if a particular day corresponds to a Friday, Saturday, or Sunday ($p_{ist} = \text{logit}^{-1}\left[\alpha_i + \delta_s + X_{it} \cdot \beta_{is}\right]$).

In the case of the MHMM, for each of its components, we specify an HMM with one, two, and three states as described above.[19] See the HMM and MHMM parameter estimates in Tables 16 and 17, respectively.

HMM and MHMM results. The validation expected log predictive posterior density (Val-LL) supports a 3-state HMM, indicating low, medium, and high gaming behavior states with daily playing

[19] We estimated an MHMM with up to three states because the corresponding recommended HMM has three states.

Table 16: 3-state HMM parameter estimates. Posterior means and posterior 95% intervals

	π_s	Q			$\mathbf{p_s}$
		State 1	State 2	State 3	
State 1	0.137 [0.038, 0.216]	0.881 [0.803, 0.926]	0.034 [0.000, 0.166]	0.086 [0.013, 0.126]	0.089 [0.048, 0.126]
State 2	0.129 [0.031, 0.247]	0.064 [0.002, 0.169]	0.519 [0.362, 0.669]	0.417 [0.264, 0.595]	0.579 [0.231, 0.833]
State 3	0.734 [0.643, 0.814]	0.095 [0.002, 0.148]	0.118 [0.065, 0.206]	0.788 [0.750, 0.820]	0.866 [0.830, 0.902]

Table 17: MHMM parameter estimates. Posterior means and posterior 95% intervals

(a) Component probabilities

	1 state	**2 states**	**3 states**
!k	0.061 [0.024 0.116]	0.185 [0.105 0.289]	0.754 [0.636 0.848]

(b) 1 state component

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

(c) 2 state component: Initial state probabilities, transition matrix, and state-dependent probabilities

	π_s	Q		$\mathbf{p_s}$
		State 1	State 2	
State 1	0.000 [0.000, 0.000]	0.787 [0.669, 0.865]	0.213 [0.135, 0.331]	0.190 [0.054, 0.352]
State 2	1.000 [1.000, 1.000]	0.087 [0.048, 0.146]	0.913 [0.854, 0.952]	0.969 [0.929, 0.999]

(d) 3 state component: Initial state probabilities, transition matrix, and state-dependent probabilities

	π_s	Q			$\mathbf{p_s}$
		State 1	State 2	State 3	
State 1	0.000 [0.000, 0.000]	0.902 [0.724, 0.963]	0.042 [0.000, 0.227]	0.056 [0.017, 0.100]	0.025 [0.000, 0.069]
State 2	0.392 [0.277, 0.535]	0.083 [0.034, 0.161]	0.731 [0.602, 0.843]	0.186 [0.102, 0.279]	0.254 [0.181, 0.360]
State 3	0.608 [0.465, 0.723]	0.002 [0.000, 0.012]	0.208 [0.155, 0.274]	0.790 [0.723, 0.841]	0.941 [0.892, 1.000]

probabilities of $p_1 = 0.089$, $p_2 = 0.579$, and $p_3 = 0.866$. Most gamers begin in the high state ($\pi_3 = 0.734$), which is consistent with the common pattern in online games of decreasing engagement over time. For the MHMM, the population membership probabilities, ω_k, indicate that most gamers (75.4%) have 3 states whereas 18.5% and 6.1% have 2 and 1 states, respectively. We now describe the behavior captured by each component. The static gamers (1 state) show a medium level of playing behavior overall with a playing probability of $p = 0.395$. The 2-states customers transition between a low level of playing behavior ($p_1 = 0.190$), and a very high level of playing behavior ($p_2 = 0.969$). All these gamers start in the high state ($\pi_2 = 1$) and stay in that state with high probability, $q_{22} = 0.913$. However, if they move to the low state, they also stay there with a relatively high probability, $q_{11} = 0.787$. The most dynamic gamers (3-states players) are characterized by inactive behavior when they are in the low state ($p_1 = 0.025$), a low level of playing behavior when they are in the second state ($p_2 = 0.254$), and a very high level of playing behavior when they are in the third state ($p_3 = 0.941$). Although no gamer in this group starts in the inactive state, when gamers transition to the inactive state, they stay there with a 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 (q_{22} and q_{33}) are less sticky than the corresponding probabilities for the 2-state gamers (q_{11} and q_{22}).

The comparison between the HMM and the MHMM reveals significant differences in player dynamics. The most notable distinction lies in the characterization of the second state. The HMM identifies a medium state ($p_2 = 0.579$) with low persistence ($q_{22} = 0.519$), whereas the MHMM does not recognize that state. Instead, the MHMM delineates three distinct states: non-playing, low-playing, and intensive-playing, all exhibiting high persistence. This discrepancy indicates that the traditional HMM, commonly used in the literature, fails to capture structural heterogeneity by conflating heterogeneity in dynamics with dynamics themselves. Consequently, the HMM suggests more dynamic behavior than the MHMM, which accounts for structural heterogeneity in player dynamics.

Predictive playing behavior. Table 18 compares the out-of-sample predictions of the selected 3-state HMM and the MHMM. The MHMM exhibits better predictive accuracy in the holdout data in both expected log predictive posterior density (LL) and RMSE. For instance, the MHMM has a lower RMSE (12.98 vs. 15.00) when comparing the predicted and actual number of customers playing each day during the holdout period.

Table 18: HMM and MHMM prediction ability.

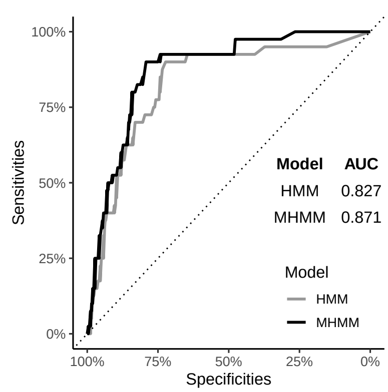
		Test	
Model	Number of states	LL	RMSE
HMM	3	-1,355.54	15.00
MHMM and Parametric	1, 2, and 3	**-1,328.07**	**12.98**

Note: The best model in each column is in bold.

Implications for customer relationship management.

Churn. Predicting churn is a pivotal aspect of customer relationship management, particularly in the online gaming industry (Hadiji et al., 2014; Lee et al., 2016). Both the HMM and MHMM indicate that players initially exhibit high engagement levels, which gradually decline over time, signaling potential churn. Properly identifying this decline is crucial for managing customer relationships and retention strategies. We assess the models' ability to predict gamer defection, defined as the absence of gameplay during the last 10 days of the holdout period. Figure 8, which displays the ROC curves for each model, demonstrates the superior performance of the MHMM in predicting churn relative to the HMM. Indeed, the area under the curve (AUC) of the MHMM is 5.3% higher than that of the HMM.

Figure 8: ROC curve for churn prediction: HMM vs. MHMM.



Customer value. We compare the inferences made by the two models (HMM and MHMM) with respect to customers' value. Since we do not observe revenue in our data, we compute the discounted expected number of active days. This metric is similar to the discounted expected residual transactions (DERT) developed as a proxy for customer value (Fader et al., 2005, 2010).[20] We use estimates from

[20]Specifically, we compute the discounted number of active days over 180 days after the calibration window.

the MHMM and categorize customers into two groups: those with three states, same as the HMM, and those with one or two states, for whom the HMM may have unidentified parameters. Figure 9a shows the average DERT according to each model for these two groups of customers.

Figure 9: HMM vs. MHMM.
Discounted expected number of active days (DERT) by predicted number of states according to the MHMM.

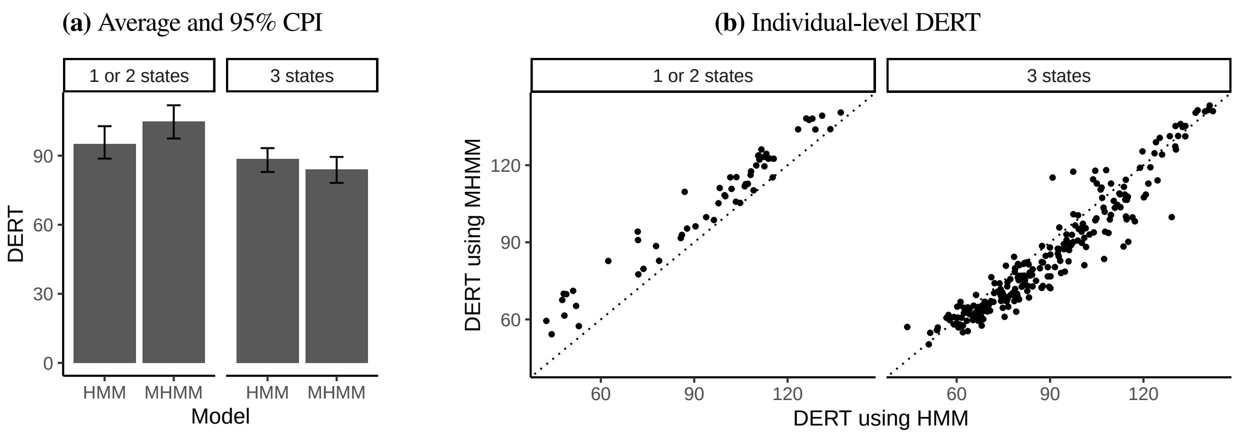


Figure 9a reveals that, compared to the MHMM, the HMM underestimates the value of customers that likely have 1 or 2 states by an average of 14.5% across customers (95% CPI: [2.24%, 32.8%]) while overestimating the value for those that are likely to have 3 states by 7.6% (95% CPI: [15.24%, 32.8%]). The same pattern is displayed at the individual level (Figure 9b), where the HMM consistently underestimates the value of customers with fewer states.

In sum, in this online gaming application, the MHMM captures richer dynamics and provides superior out-of-sample predictions for daily playing behavior and better churn and customer lifetime value assessment compared to the HMM.

DISCUSSION

HMMs have proven to be an effective modeling approach for describing customer dynamic behavior, resulting in a diverse range of marketing applications, from eye-tracking to B2B contexts. While most applications accommodate flexible heterogeneity in customer behavior (*parametric heterogeneity*), they often assume uniformity in the degree of dynamics by imposing the same number of hidden states on all customers (*structural homogeneity*).

In this work, we first analyze through a series of simulations the implications of such an assumption. We simulate a population with two segments having different numbers of states and estimate a typical HMM that allows for parametric heterogeneity but does not account for structural heterogeneity. Our findings show that a few customers with a high number of states can influence model selection, leading to the choice of complex models that do not accurately represent the majority's behavior. Consequently, population estimates may not effectively summarize customer behavior and should be interpreted cautiously to avoid misleading conclusions. Furthermore, we observe biased estimates even for customers with a correctly specified number of states due to the influence of those with misspecified and unidentified dynamics on the overall parameter estimates. We show that these biases may result in wrong inferences about the effectiveness of marketing actions and subsequent targeting decisions.

To address these issues, we recommend using approaches that flexibly account for structural heterogeneity in the number of states and, at the same time, allow for parametric heterogeneity, such as a mixture of HMMs. We demonstrate the value of these approaches using simulation exercises and two empirical applications. The pharmaceutical application highlights the different characterization of dynamics, and the impact on targeting of marketing actions. The online gaming application further highlights the importance of properly disentangling parametric heterogeneity from structural heterogeneity and its implications for customer lifetime value and churn predictions.

However, we caution that using approaches like the MHMM could become expensive in terms of the number of parameters to estimate if one allows for a large number of segments with a higher number of states. That being said, in most applications seen in the marketing literature, the number of estimated HMM states has been relatively small. Alternative approaches to mixtures of HMMs could be to use a Hierarchical Dirichlet Process or a Reversible Jump MCMC approach to let the data select the number of hidden states as part of the inference procedure. However, the application of such approaches with an individual-level selection of mass points can be computationally challenging (see, e.g., Gopalakrishnan et al., 2017). We leave alternative modeling approaches for future research.

Overall, despite the deserved importance and care given in the marketing literature and practice to capturing individual-level preference heterogeneity, the application of HMMs in marketing commonly assumes heterogeneous parameters but homogeneity in the number of states for all consumers. We demonstrate the risks of assuming such homogeneity and propose an effective solution. We hope this research will pave the road for further exploring structural heterogeneity in HMMs and other dynamic models.

REFERENCES

Allenby, G. M., Arora, N., and Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research*, 35(3):384–389.

Alós-Ferrer, C. and Garagnani, M. (2023). Part-time bayesians: incentives and behavioral heterogeneity in belief updating. *Management Science*, 69(9):5523–5542.

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).

Chen, P. P., Guitart, A., del Río, A. F., and Periánez, A. (2018). Customer lifetime value in video games using deep learning and parametric models. In *2018 IEEE international conference on big data (big data)*, pages 2134–2140. IEEE.

Daljord, Ø., Misra, S., and Nair, H. S. (2016). Homogeneous contracts for heterogeneous agents: Aligning sales force composition and compensation. *Journal of Marketing Research*, 53(2):161–182.

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.

Fader, P. S., Hardie, B. G., and Lee, K. L. (2005). “counting your customers” the easy way: An alternative to the pareto/nbd model. *Marketing Science*, 24(2):275–284.

Fader, P. S., Hardie, B. G., and Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29(6):1086–1108.

Ferecatu, A. and De Bruyn, A. (2022). Understanding managers’ trade-offs between exploration and exploitation. *Marketing Science*, 41(1):139–165.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2020). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24:997–1016.

Gopalakrishnan, A., Bradlow, E. T., and Fader, P. S. (2017). A cross-cohort changepoint model for customer-base analysis. *Marketing Science*, 36(2):195–213.

Gopalakrishnan, A., Jiang, Z., Nevskaya, Y., and Thomadsen, R. (2021). Can non-tiered customer loyalty programs be profitable? *Marketing Science*, 40(3):508–526.

Gu, Z. and Yang, S. (2010). Quantity-discount-dependent consumer preferences and competitive nonlinear pricing. *Journal of Marketing Research*, 47(6):1100–1113.

Hadiji, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K., and Bauckhage, C. (2014). Predicting player churn in the wild. In *2014 IEEE Conference on Computational Intelligence and Games*, pages 1–8. IEEE.

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. et al. (1981). Heterogeneity and state dependence. *Studies in labor markets*, 31:91–140.

Helske, S. and Helske, J. (2019). Mixture hidden markov models for sequence data: The seqhmm package in r. *Journal of Statistical Software*, 88:1–32.

Huang, Y., Jasin, S., and Manchanda, P. (2019). "Level up": Leveraging skill and engagement to maximize player game-play in online video games. *Information Systems Research*, 30(3):927–947.

Hui, S. K. (2017). Understanding repeat playing behavior in casual games using a bayesian data augmentation approach. *Quantitative Marketing and Economics*, 15:29–55.

Kamakura, W. A., Kim, B.-D., and Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, 15(2):152–172.

Kappe, E., Blank, A. S., and DeSarbo, W. S. (2018). A random coefficients mixture hidden Markov model for marketing research. *International Journal of Research in Marketing*, 35(3):415–431.

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.

Lee, C., Ofek, E., and Steenburgh, T. J. (2018). Personal and social usage: The origins of active customers and ways to keep them engaged. *Management Science*, 64(6):2473–2495.

Lee, S.-K., Hong, S.-J., Yang, S.-I., and Lee, H. (2016). Predicting churn in mobile free-to-play games. In *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1046–1048. IEEE.

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.

Liberali, G. and Ferecatu, A. (2022). Morphing for consumer dynamics: Bandits meet hidden markov models. *Marketing Science*, 41(4):769–794.

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*, 50(1):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.

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.

Misra, S. and Nair, H. S. (2011). A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, 9:211–257.

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. (2019). A hidden markov model to detect on-shelf out-of-stocks using point-of-sale data. *Manufacturing & Service Operations Management*, 21(4):932–948.

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.

Nahm, I. Y., Ahearne, M. J., Lee, N., and Tirunillai, S. (2022). Managing positive and negative trends in sales call outcomes: The role of momentum. *Journal of Marketing Research*, 59(6):1120–1140.

Naumzik, C., Feuerriegel, S., and Weinmann, M. (2022). I will survive: Predicting business failures from customer ratings. *Marketing Science*, 41(1):188–207.

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.

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. *Journal of the American Statistical Association*, 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*, 59(5):1009–1026.

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

Smith, A., Naik, P. A., and Tsai, C.-L. (2006). Markov-switching model selection using kullback–leibler divergence. *Journal of Econometrics*, 134(2):553–577.

Stüttgen, P., Boatwright, P., and Monroe, R. T. (2012). A Satisficing Choice Model. *Marketing Science*, 31(6):878–899.

Van de Pol, F. and Langeheine, R. (1990). Mixed markov latent class models. *Sociological methodology*, pages 213–247.

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.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27:1413–1432.

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.

Yang, S. and Allenby, G. M. (2000). A model for observation, structural, and household heterogeneity in panel data. *Marketing Letters*, 11:137–149.

Zhang, J. Z., Chang, C.-W., and Neslin, S. A. (2022). How physical stores enhance customer value: The importance of product inspection depth. *Journal of Marketing*, 86(2):166–185.

Zhang, J. Z., Netzer, O., and Ansari, A. (2014). Dynamic Targeted Pricing in B2B Relationships. *Marketing Science*, 33(3):317–337.

Zhang, J. Z., Watson, G. F., Palmatier, R. W., and Dant, R. P. (2016). Dynamic Relationship Marketing. *Journal of Marketing*, 80(5):53–75.

Zhang, X., Kumar, V., and Cosguner, K. (2017). Dynamically managing a profitable email marketing program. *Journal of Marketing Research*, 54(6):851–866.

Zucchini, W. and MacDonald, I. (2009). Hidden Markov Models for Time Series. *Chapman and Hall/CRC*.

WEB APPENDIX

Appendix contents

A HMM model specification **App - 3**
A.1 State dependent probabilities . . . . . App - 3
A.2 Transition probabilities . . . . . App - 3
A.3 Parametric heterogeneity . . . . . App - 4
A.4 Priors . . . . . App - 4

B Full simulation analysis **App - 6**
B.1 Model selection criteria . . . . . App - 8
B.2 Baseline condition . . . . . App - 9
B.3 Experimental conditions . . . . . App - 13
B.4 Random Coefficient Mixture HMM . . . . . App - 23
B.5 Mixture of Hidden Markov Models (MHMM) . . . . . App - 25
B.6 Simulation with marketing decisions . . . . . App - 31
B.7 MHMM performance for data generated following an HMM . . . . . App - 37

C Why an Unconstrained Enlarged HMM Does Not Recover the MHMM? **App - 39**

D Model Selection for hierarchical Bayesian Regression **App - 42**

E Aggregate HMM tables reported in the literature **App - 43**

F MHMM model specification **App - 44**
F.1 HMM components . . . . . App - 44
F.2 Parametric heterogeneity . . . . . App - 45
F.3 Priors . . . . . App - 45

G Details of Application 1: Pharmaceutical marketing **App - 46**
G.1 Selection of number of states and Gaussian mixture components . . . . . App - 46
G.2 Model estimates . . . . . App - 46
G.3 Targeting . . . . . App - 51

H Details of Application 2: Online gaming **App - 51**

H.1 Data . . . . App - 51
H.2 HMM and MHMM specifications . . . . App - 52
H.3 Results . . . . App - 53
H.4 Parameter estimates for the 3-state HMM . . . . App - 54
H.5 Parameter estimates for the MHMM . . . . App - 56

A HMM model specification

We detail the parameter transformations and priors for the general model specification in the main manuscript.

A.1 State dependent probabilities

We define the increasing vector of probabilities $\mathbf{p}_i$ with $p_{i1} < p_{i2} < \ldots < p_{iS}$, by

$$p_{is} = \text{logit}^{-1}\left[\alpha_i + \delta_s\right] \tag{A.1}$$

with $\delta_1 < \delta_2 < \ldots < \delta_S$. To ensure such ordering, we parametrize δ_s by

$$\delta_s = \text{logit}\left[\sum_{\ell=1}^{s} u_\ell\right] \tag{A.2}$$

with $\mathbf{u} = \{u_1, \ldots, u_S\}$ a vector in the $(S-1)$dimensional simplex $\left(\mathbf{u} \in \mathbb{R}^S,\ u_s \geq 0,\ \sum_s u_s = 1\right)$.

For the specification with time-varying covariates X_{it} in the state dependent probabilities (Sections 4.1 and 4.2), we include these covariates by

$$p_{ist} = \text{logit}^{-1}\left[\alpha_i + \delta_s + X_{it}' \cdot \boldsymbol{\beta}_{is}\right].$$

A.2 Transition probabilities

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

$$q_{iss'} = \begin{cases} \frac{\exp(\gamma_{iss'})}{\sum_{\ell=1}^{S-1} \exp(\gamma_{is\ell})+1} & \text{if } s' \in \{1, \ldots, S-1\} \\ \frac{1}{\sum_{\ell=1}^{S-1} \exp(\gamma_{is\ell})+1} & \text{if } s' = S \end{cases} \tag{A.3}$$

For the specification with time-varying covariates X_{it} in the transition matrix (Sections 2.3 and 4.1), we use an ordered logistic specification where transition probabilities $q_{its\ell}(X_{it}) = P(z_{it+1} = \ell | z_{it} = s, X_{it})$ are specified by:

$$\begin{aligned} q_{its1} &= \frac{\exp\left(\gamma_{is1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)} \\ q_{its\ell} &= \frac{\exp\left(\gamma_{isl} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is\ell} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)} - \frac{\exp\left(\gamma_{is\ell-1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{is\ell-1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)} \qquad \ell = 2, \ldots, S-1 \\ q_{itsS} &= 1 - \frac{\exp\left(\gamma_{isS-1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)}{1 + \exp\left(\gamma_{isS-1} - X_{it}' \cdot \boldsymbol{\rho}_{is}\right)} \end{aligned}$$

and $\gamma_{is1} < \gamma_{is2} < \ldots < \gamma_{isS-1}$. We model such ordering directly in Stan using the native ordered vector parameters.

A.3 Parametric heterogeneity

We model the heterogeneity in state-dependent intercepts with a zero-mean Gaussian distribution (the mean is identified through $\boldsymbol{\delta}$) by

$$\alpha_i = \mathcal{N}(0, \sigma_\alpha).$$

We denote $\boldsymbol{\theta}_i = (\gamma_{i11}, \ldots, \gamma_{iSS-1})$. We defined a Gaussian distribution to account for unobserved heterogeneity,[1]

$$\boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \Sigma_\theta)$$

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

For ease of computation. We re-parameterize $\boldsymbol{\tau} = \sqrt{\text{diag}(\Sigma_\theta)}$ the vector of standard deviations, Ω the correlation matrix for covariance Σ, and L_Ω its corresponding Cholesky decomposition. We also define $\mathbf{z}_i$ the standardized individual level parameters, such that

$$\boldsymbol{\theta}_i = \boldsymbol{\mu}_\theta + \text{diag}(\boldsymbol{\tau}) \cdot L_\Omega \cdot \mathbf{z}_i$$

with $\mathbf{z}_i \sim \mathcal{N}(0, I)$. We parametrize σ_α by

$$\sigma_\alpha = 2.5 \cdot \tan(\widetilde{\sigma}_\alpha).$$

Similarly, we parametrize each component ℓ of $\boldsymbol{\tau}$ using the vector $\widetilde{\boldsymbol{\tau}}$ such that

$$\tau_\ell = 2.5 \cdot \tan(\widetilde{\tau}_\ell).$$

A.4 Priors

A.4.1 Initial probabilities

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

$$\boldsymbol{\pi} \sim \text{Dirichlet}(1).$$

A.4.2 State dependent probabilities probabilities

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

$$u \sim \text{Dirichlet}(1).$$

We also use uniform priors for $\widetilde{\sigma}_\alpha$

$$\widetilde{\sigma}_\alpha \sim U(0, \pi/2).$$

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

[1] For the specification with covariates, since the threshold parameters are ordered, we model heterogeneity on the unconstrained parameters and correct for the transformation using the absolute value of the determinant of the Jacobian transformation.

A.4.3 Transition probabilities

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

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

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

Here, π also 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 of the cholesky factor of a matrix that distributes according to an LKJ correlation distribution (LKJ_Corr$(\Omega|c) \propto$ det$(\Omega)^{c-1}$.

B Full simulation analysis

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, and explore different conditions that may affect the performance of the standard HMM in this context. To do so, we investigate the following factors:

1. Degree of structural heterogeneity. Corresponds to the maximum number of hidden states: 1 and 2 vs. 2 and 3.

2. Structural heterogeneity mixture. Corresponds to the percentage of customers with 2 states: 50%, 70%, 80%, 90%, 95%, 98%.

3. Non-structural parametric heterogeneity. Dispersion of state-dependent and switching probabilities across customers: No heterogeneity, low heterogeneity ($\sigma = 0.2 \cdot |\mu|$), high heterogeneity ($\sigma = 0.5 \cdot |\mu|$).

4. Degree of dynamics. Corresponds to the magnitude of the switching probabilities in the transition matrix that characterizes the stickiness of the underlying dynamics: low vs. high (0.85 vs. 0.70 in the diagonal of the switching matrix.)

5. Length of time series data for calibration. Corresponds to the number of observations per customer: short vs. long (25 vs. 40 periods)

Additionally, we evaluate further implications for:

6. Effect of structural heterogeneity on marketing decisions. We investigate the effect of marketing covariates on customer dynamics.

Rather than opting for a comprehensive full-factorial design, as not all combinations hold significant interest, we adopt a sequential approach by manipulating one condition at a time and comparing it to the base case, following an A/B testing framework. Moreover, to ensure the robustness of our findings, we employed a random data generation process to create 100 distinct datasets for each experimental condition. Subsequently, we conducted separate model estimations on each of these datasets.

The baseline condition corresponds to the following scenario:

- Degree of structural heterogeneity: Large
- Structural heterogeneity mixture: 50%
- Non-structural parametric heterogeneity: Low

- Degree of dynamics: Low
- Length of time series data: Short

Throughout, we assume that there are two customer segments. 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 time periods. 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 $\lambda \in [0, 100]$ 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 B.1 shows the corresponding initial, state-dependent, and transition probabilities used to simulate the behavior of each segment.

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

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

In the baseline condition, we simulate 500 customers during $T = 45$ time periods (25 periods for calibration, 10 for validation, and 10 for testing) with $\lambda = 50\%$. This implies that 50% of the customers belong to a segment with three states (Segment A) and 50% belongs to a segment with two states (Segment B).

We estimate an HMM with a Binomial state-dependent distribution as described in Section 1.1 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 corresponding posterior distributions (see Section 1.3 for details of the estimation procedure). 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, the researcher typically pools the data of all customers and estimates an HMM with heterogeneous parameters but with a common number of states across customers. For this simulation exercise, we specify initial state parameters as homogeneous while allowing state-dependent parameters and 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 as 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 Web Appendix A for further details of the model specification and prior distributions.

B.1 Model selection criteria

We rely on out-of-sample metrics (LL and RMSE) for model selection regarding the number of states. In-sample model selection metrics have been shown to provide mixed results, even when there is no structural heterogeneity, and have been questioned in the literature (Gelman et al., 2014; Vehtari et al., 2017). To see this, Table B.2 reports the results across 100 replications when the true underlying behavior corresponds to customers with 3 states.

Table B.2: Model selection performance when there is no structural heterogeneity (all customers have 3 states).

States	Proportion of reps with best				
	In sample			Validation	
	DIC	LMD	WAIC	LL	RMSE
2	0.05	1.00	0.68	0.00	0.00
3	0.10	0.00	0.05	0.83	1.00
4	0.85	0.00	0.27	0.17	0.00

The results in Table B.2 show that in-sample metrics (DIC, LMD, and WAIC) consistently failed to recover the true number of states. In contrast, Validation LL (expected log predictive density) and Validation RMSE provide better results. In consequence, we use the latter metrics throughout the simulation exercise.

We now proceed to report the results of the simulation exercise, starting with the baseline condition.

B.2 Baseline condition

We first note that, for the baseline condition, the model selection criteria (Validation LL and Validation RMSE) suggest that a 3-states HMM best represents customer behavior in all 100 replications compared with a 2-states HMM. Next, the parameter estimates across replications are illustrated in Figure B.1 (conditional probabilities P) and Figure B.2 (transition matrix Q).

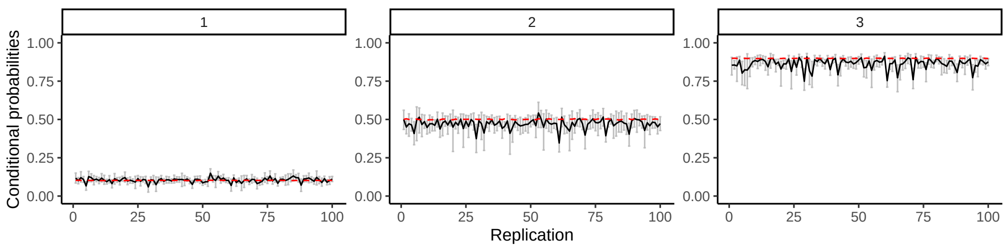


Figure B.1: Baseline condition.
Recovery of population conditional probabilities across 100 replications. True simulated values are in red.

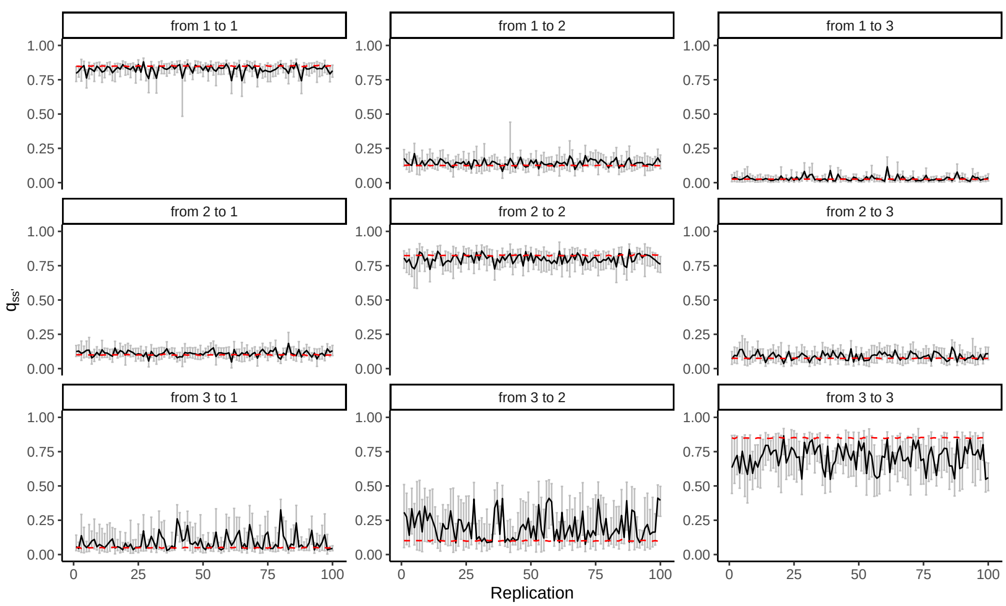


Figure B.2: Baseline condition. Recovery of population transition probabilities across 100 replications. True simulated values are in red.

In Figure B.1, we observe a consistent downward bias in the conditional probabilities of states 2 and 3. In Figure B.2, we observe that transition probabilities from states 1 and 2 are generally well recovered. However, the estimated last row of the transition matrix (corresponding to transitions from state 3) is more diffuse than the true underlying behavior, with significant overestimation of q_{32} and underestimation of q_{33}. In particular, the downward bias of q_{33} suggests a lower stickiness in the third state than the true simulated value, which implies a higher degree of dynamics.

To further investigate the bias in the transition matrix, we split the results for customers that have 3 states (Figure B.3a) and 2 states (Figure B.3b). Recall that all of them were estimated to have 3 states.

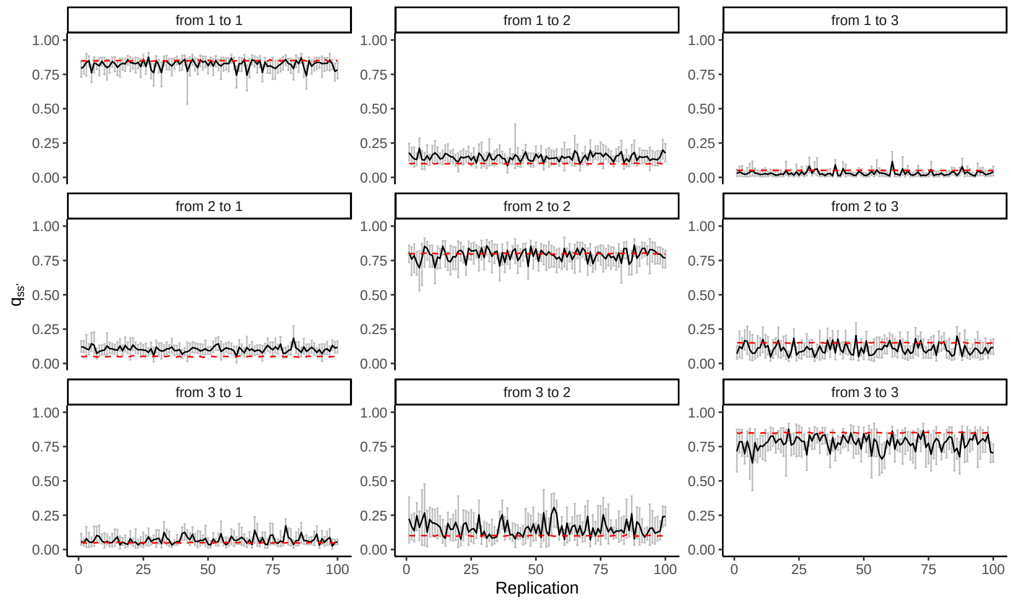
Figure B.3b shows that the last row for Segment B is not identified and that their values are heavily influenced by the pooling with 3 state customers. Interestingly, Figure B.3a shows that such pooling also affects 3-state customers.

In contrast, we note that when there is no structural heterogeneity; that is, when we estimate an HMM only for customers with 3 states (Segment A), the HMM does a good job in recovering the true parameters even though the model uses only half the data compared to the case of structural heterogeneity where the model includes all customers (see Figure B.4).

At the individual level, we can see this pattern more strongly. Figure B.5 shows a 2D histogram of the true vs. individual posterior mean q_{i33} across Segment A customers in all replications when there is structural heterogeneity (top) and where there is not (bottom).

Figure B.3: Baseline condition. Recovery
of population transition probabilities across 100 replications for Segment A. True simulated values are in red.

(a) Segment A (3 states)



(b) Segment B (2 states)

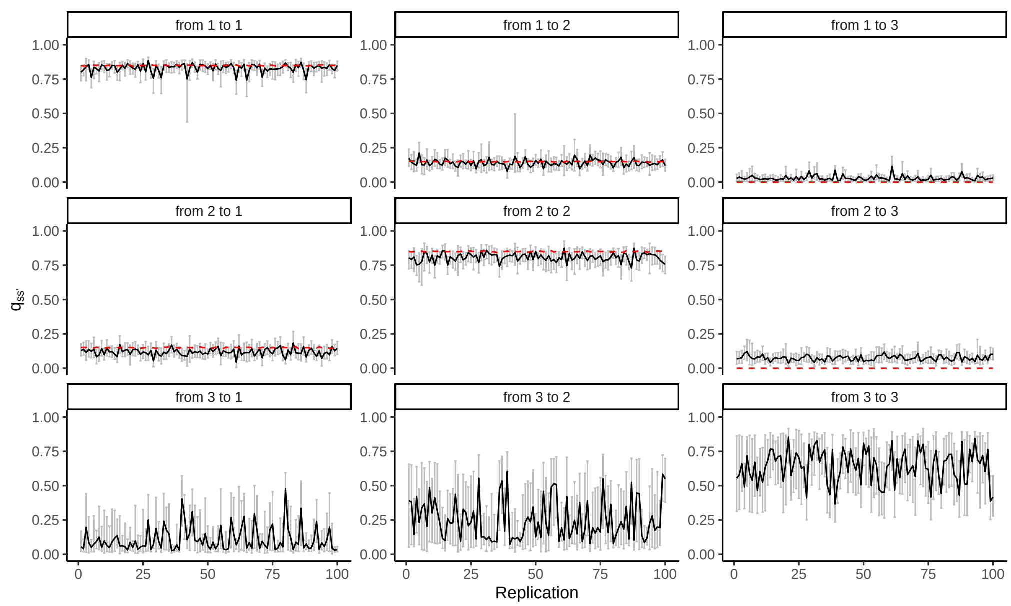


Figure B.6a reports the posterior mean and 95% confidence intervals of the individual transition probabilities for all customers sorted in decreasing estimated probability. For some transition probabilities, we observe large confidence intervals for many customers that prevent any statistical inference regarding their 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

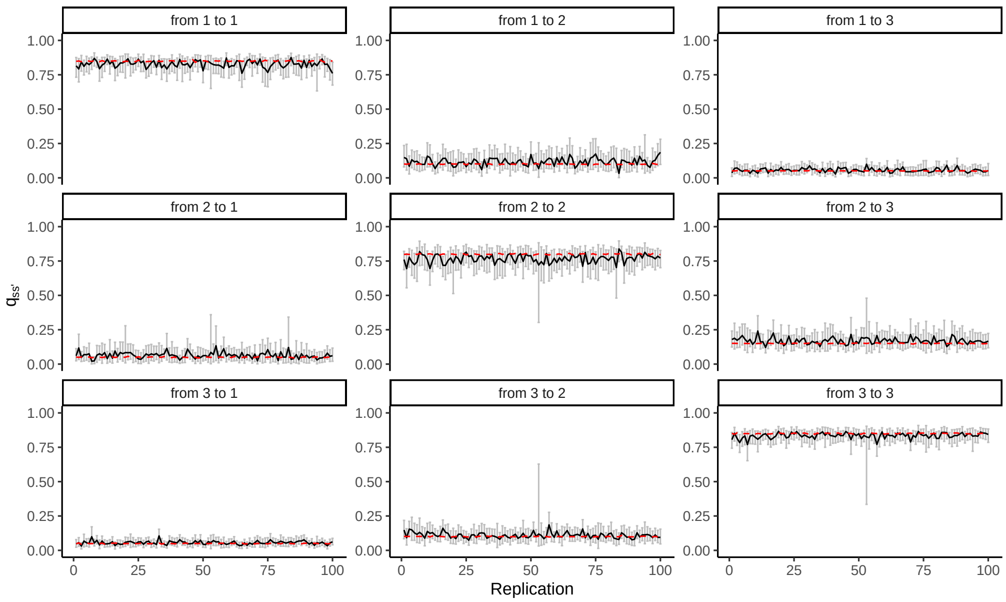


Figure B.4:
Baseline condition with no structural heterogeneity. Recovery of population transition probabilities across 100 replications for Segment A (3 states) when only Segment A data are used. True simulated values are in red.

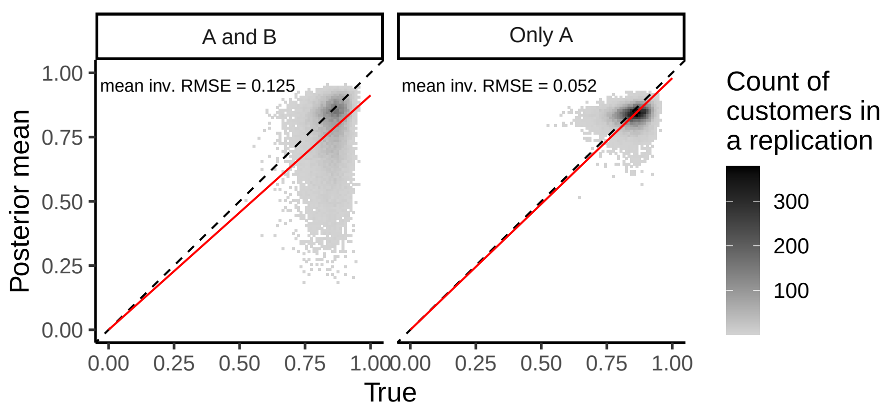
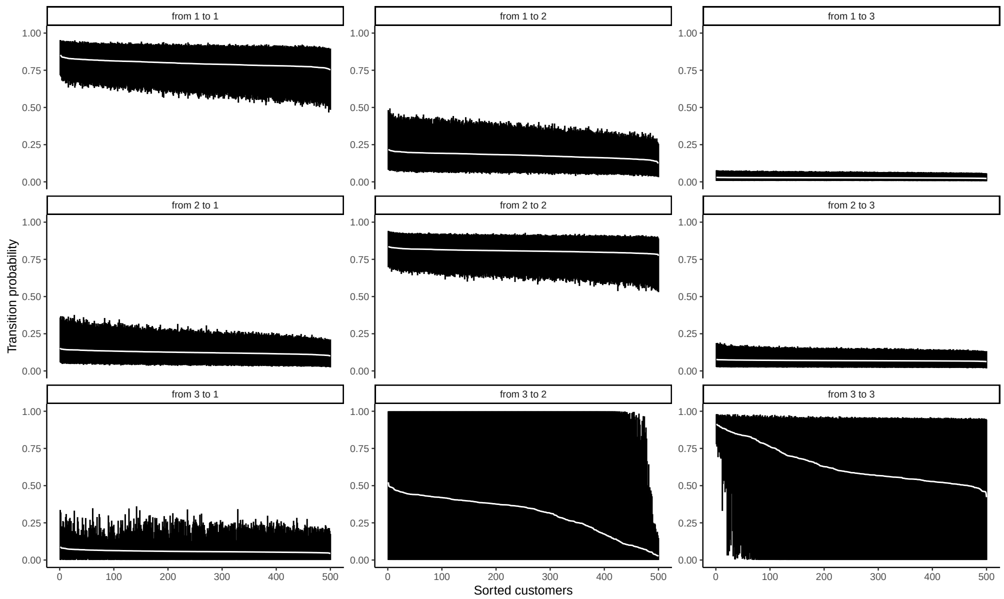


Figure B.5: Baseline condition.
True vs. posterior mean of the individual q_{i33}. The left panel corresponds to the case when both segments are used in the estimation, whereas the right panel corresponds to the case when only Segment A customers are used in the estimation. Each observation represents a customer in a replication. A darker color indicates a higher count.

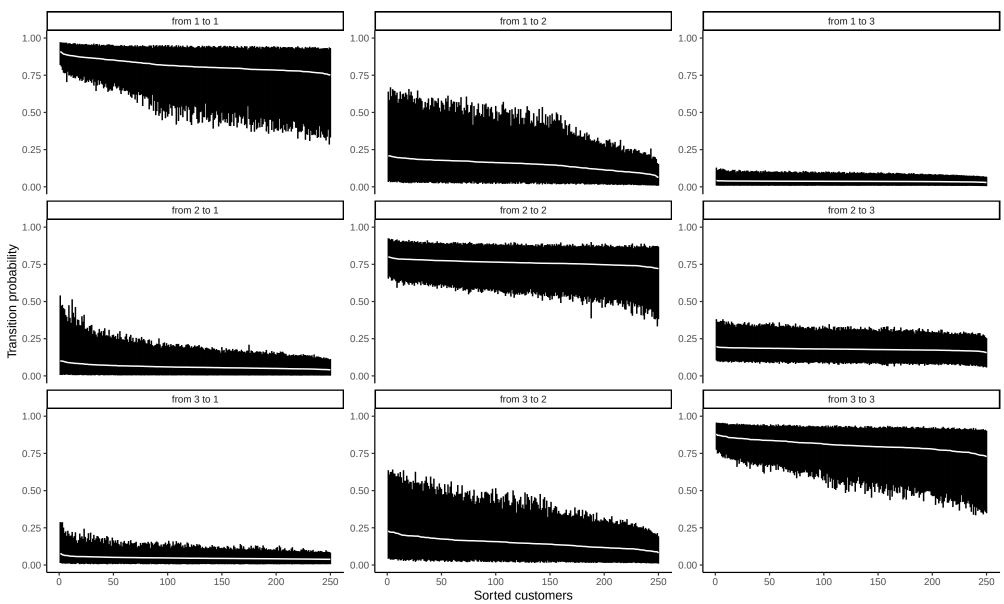
customers (see the corresponding last row of the transition matrix in Figure B.7b).

Figure B.6: Posterior mean and 95% confidence interval for individual transition probabilities. For each transition probability, customers are sorted in decreasing order 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)

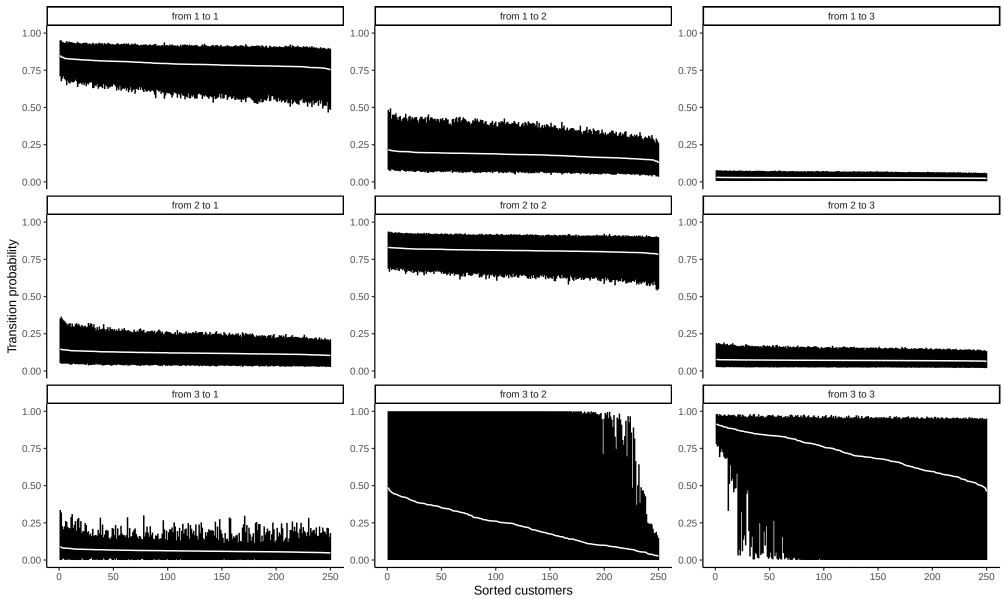


B.3 Experimental conditions

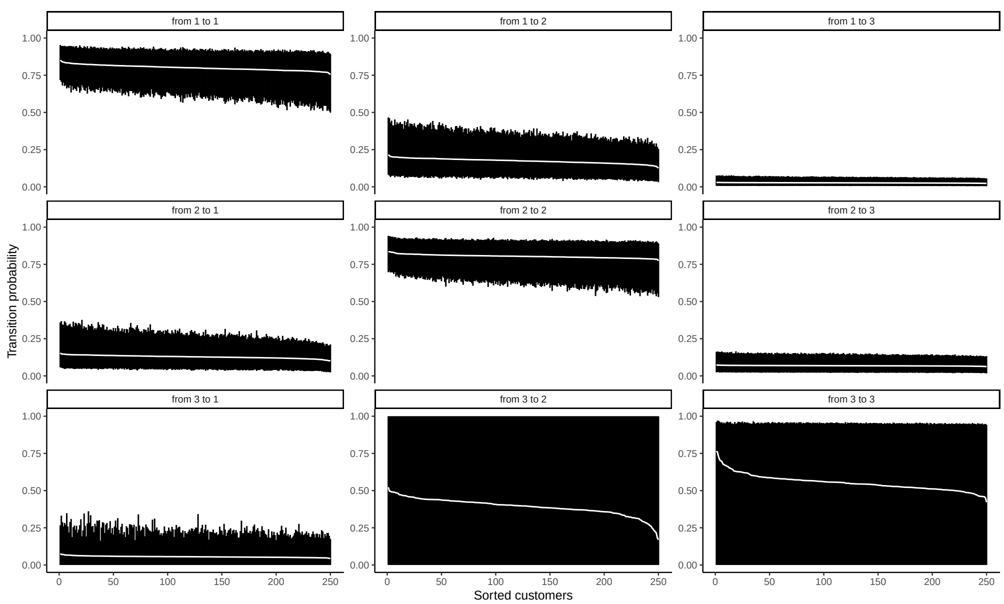
We now explore whether these results also extend to the other experimental conditions. To examine the recovery of the parameter estimates in a structured manner, we compute the individual-level RMSE

Figure B.7: Posterior mean and 95% confidence interval for individual transition probabilities. For each transition probability, customers are sorted in decreasing order 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)



of the parameters across replications for customers in Segment A (those with three states). We focus on the transition probability of the highest state across experimental conditions. At the end of this section, we report all parameter estimates for all experimental conditions.

B.3.1 Degree of structural heterogeneity

We compare the performance of the state-of-the-art HMM when there is low and high structural heterogeneity (baseline condition). In the case of low structural heterogeneity, one of the segments is static (has one state), and the other segment is dynamic with two latent states.

First, in the case of low structural heterogeneity, and similar to the baseline scenario, the model selection criteria consistently suggest that the more complex HMM specification with 2 states provides better results in all 100 replications.

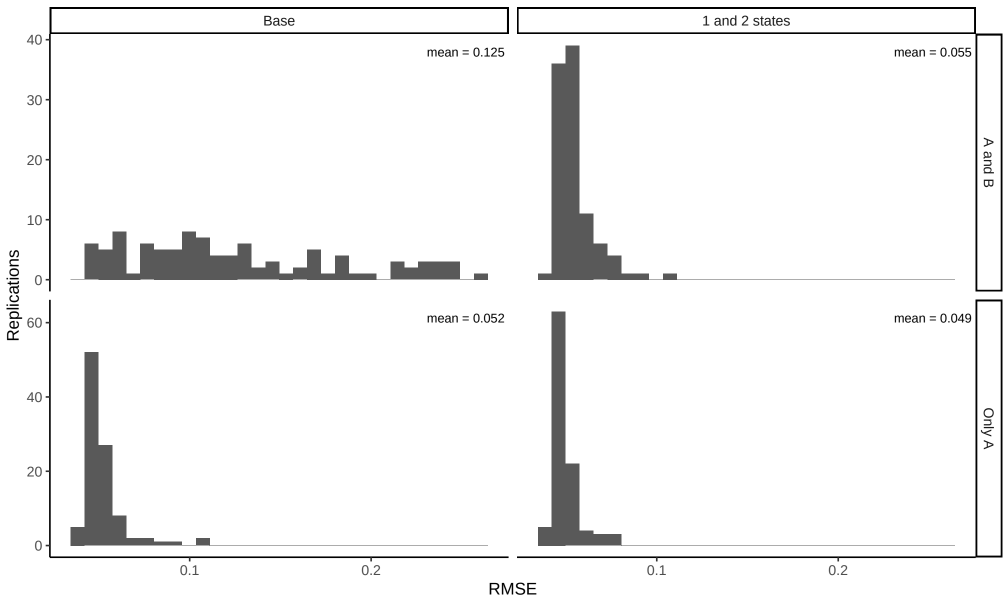


Figure B.8: Histogram of the individual RMSE between the true simulated and the estimated transition probability of the highest state q_{33} and q_{22} across 100 replications for Segment A customers. The top panel corresponds to the case when both segments are used in the estimation, whereas the bottom panel corresponds to the case when only Segment A customers are used in the estimation. The columns represent the baseline (high structural heterogeneity), and 1 and 2 states (low structural heterogeneity), respectively.

In Figure B.8, we report the recovery of the transition probability of the highest state q_{33} (for the high structural heterogeneity condition) and q_{22} (for the low structural heterogeneity condition) for Segment A customers. The higher panel shows the case when segments A and B are used in the estimation, whereas the bottom panel shows the case when only Segment A is used. The results show that having low structural heterogeneity provides better recovery for the corresponding parameter (Average RMSE = 0.055 vs. 0.125). The bottom panel serves as a validity check because, in both cases, only one segment is used to estimate the parameters. The results are similar between experimental conditions, although, as expected, the HMM performs better when there are two hidden states vs. three hidden states (Average RMSE = 0.049 vs. 0.052).

B.3.2 Structural heterogeneity mixture

We now analyze the impact of the mixture of customers on the resulting estimated HMM. In particular, we focus on determining the number of states that best characterizes customer dynamics. This corresponds to the hyperparameter that typical model selection criteria suggest.

In Table B.3, we present the percentages of instances where a 3-state HMM specification is recommended (vs. a 2-state HMM). The first column provides the true proportion of customers with 2 states, while the second and third columns indicate the percentages of cases where a 3-state HMM is suggested based on the Validation Log-Likelihood (Val. LL) or the Validation Root Mean Squared Error (Val. RMSE) criteria, respectively.

Table B.3: Prevalence of 3-state HMM specification suggestions for different mixtures of 2 and 3-state customers across 100 replications.

Proportion with 2 states	Replications with 3 states as best model Val. LL	Val. RMSE
50%	100%	100%
70%	100%	100%
80%	99%	100%
90%	97%	99%
95%	92%	80%
98%	44%	25%

Table B.3 shows that model selection criteria (Validation LL and Validation RMSE) favor a more complex model even when most customers have two states. Only when the number of 3-state customers is relatively small (2%), the criteria start suggesting a simpler HMM specification with two states.

We also investigate the impact of the mixture of 2 and 3-state customers on parameter recovery. To do so, we focus on the last row of the transition matrix with three states ($q_{3\cdot}$, that corresponds to the transition probabilities from the highest state) for Segment A customers. On that row, q_{33} represents the stickiness in the more intensive state, and it has been the focus of previous research (see e.g., Kumar et al., 2011).

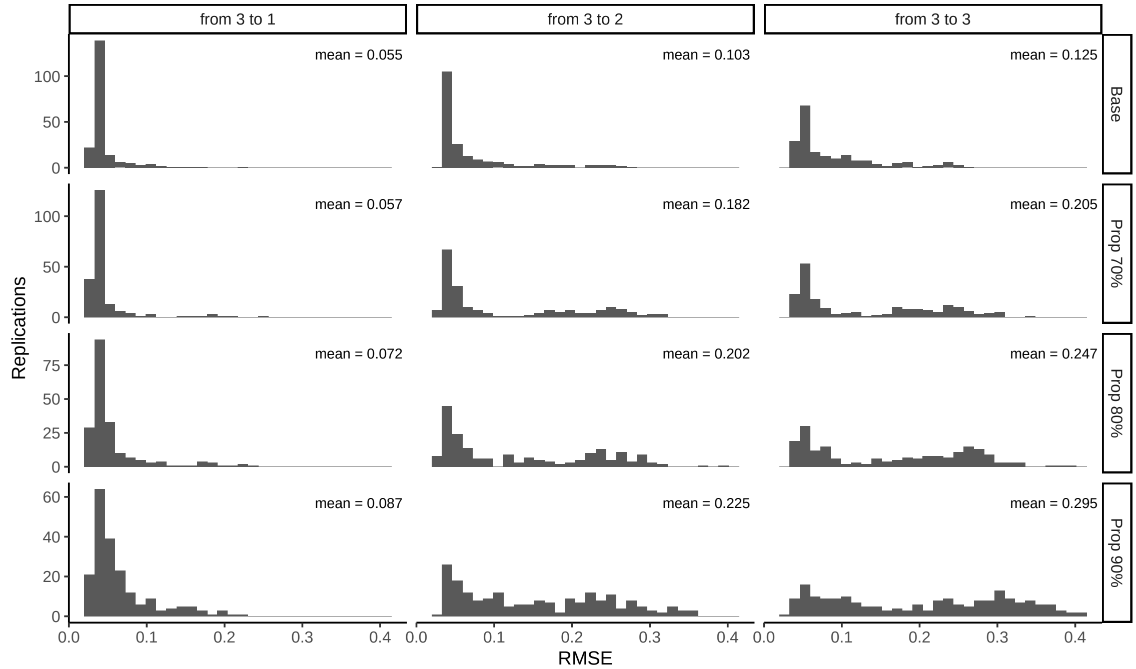


Figure B.9: Histogram of RMSE
between true simulated and estimated $q_{3\cdot}$ for Segment A customers across 100 replications. Considering different structural heterogeneity mixtures. Each row represents the RMSE distribution of the corresponding mixture.

As expected, Figure B.9 shows that even though, in most conditions, a 3-state HMM is selected, the recovery of $q_{3\cdot}$ worsens with fewer 3-state customers. Therefore, it appears that a consistent bias exists in estimating more complex model specifications and tends to persist regardless of the degree of structural heterogeneity.

B.3.3 Non-structural parametric heterogeneity

We now analyze the implications of having more or less dispersion in the conditional and transition probabilities of the HMM. In particular, we analyzed three conditions:

(i) No heterogeneity
(ii) Low heterogeneity ($\sigma = 0.2 \cdot \mu$, baseline scenario)
(iii) High heterogeneity ($\sigma = 0.5 \cdot \mu$)

To illustrate the effect of dispersion on the accuracy of parameter estimation, Figure B.10 presents a histogram of the individual RMSE values. This figure reveals significant inaccuracies in parameter recovery at all levels of parametric heterogeneity, with an increase in estimated bias with the increase in parametric heterogeneity.

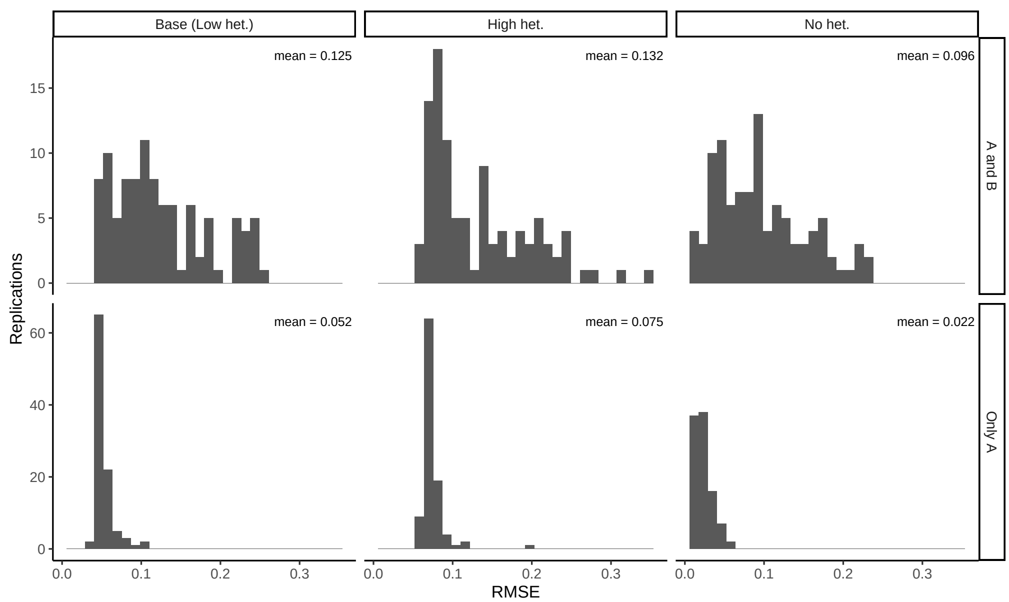


Figure B.10: Histogram of the individual RMSE of the estimated q_{33} vs. true q_{33} across 100 replications for Segment A customers. The top panel corresponds to the case when both segments are used in the estimation, whereas the bottom panel corresponds to the case when only Segment A customers are used in the estimation. The columns represent the baseline (low heterogeneity), high heterogeneity, and no heterogeneity conditions, respectively.

To shed light on this result, we show in Figure B.11 the individual level posterior mean estimates vs. the true simulated q_{33} probabilities across all replications and customers belonging to Segment A.

Figure B.11 offers evidence of a downward bias resulting from structural heterogeneity, further exacerbated by parametric heterogeneity.

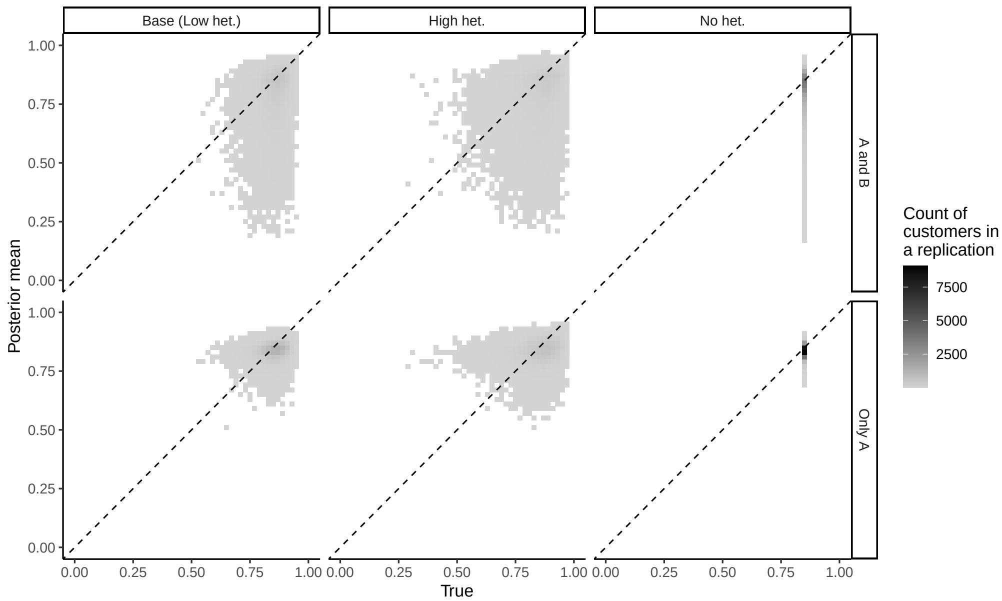


Figure B.11: Scatter plot of individual-level recovery of q_{33} across 100 replications for Segment A customers. The top panel corresponds to the case when both segments are used in the estimation whereas the bottom panel corresponds to the case when only Segment A customers are used in the estimation. True simulated values are in the x-axis and the estimated values are in the y-axis. The columns represent the baseline (low heterogeneity), high heterogeneity, and no heterogeneity conditions, respectively.

B.3.4 Degree of dynamics

As HMMs have proven effective in characterizing customer dynamics, we now examine the influence of dynamic behavior on parameter estimation, particularly within the context of structural heterogeneity. In this experimental condition, we maintain our focus on Segment A customers and specifically explore the dynamics associated with the highest state and the transition matrix parameter q_{33}. We manipulate the degree of dynamics when customers are in the third state, considering two conditions: $q_{33} = 0.70$ to represent low stickiness (high dynamics) and $q_{33} = 0.85$ as the baseline condition signifying high stickiness (low dynamics).

Figure B.12 provides a visual representation of the parameter recovery of q_{33} across different customer dynamics scenarios. In the absence of structural heterogeneity, as depicted in the bottom panel, a high stickiness parameter (low dynamics) exhibits a slightly better recovery than the high dynamics condition (Average RMSE= 0.052 vs. 0.057). In contrast, in the presence of structural heterogeneity, as shown in the top panel, a notable and substantial bias is observed in both the high and low stickiness conditions, with a somwhat higher bias in the high stickiness condition (Average RMSE= 0.125 vs. 0.099).

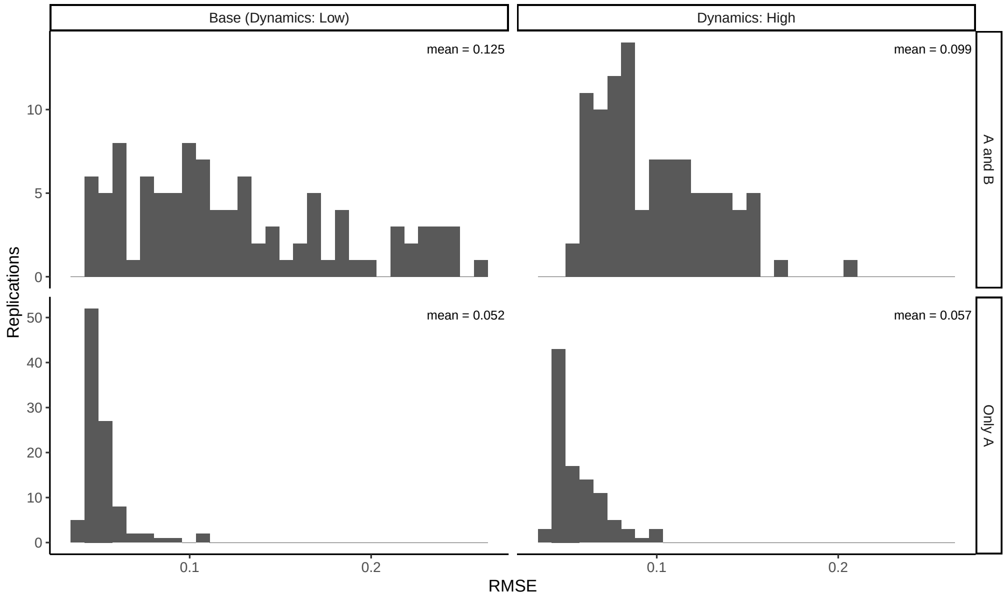


Figure B.12: Histogram of the individual RMSE between the true and the estimated q_{33} across 100 replications for Segment A customers. The top panel shows the bias when both segments are used in the estimation, whereas the bottom panel shows the bias when only Segment A customers are used in the estimation. The columns represent the high stickiness (baseline), and low stickiness conditions, respectively.

B.3.5 Length of time series data

We now turn our attention to the impact of having more information per customer. Can longer time series per customer mitigate the bias from not accounting for structural heterogeneity? To explore this, we introduce variations in the length of individual time series data used to calibrate the HMM, employing two distinct conditions: one with $T = 25$ periods, representing a baseline of short data, and another with $T = 40$ periods, representing an extended dataset. In this experimental scenario, our primary focus remains on Segment A customers.

Figure B.13 offers a graphical illustration of the bias in q33, comparing cases with relatively short and long individual time series. When structural heterogeneity is absent, as illustrated in the lower panel, a longer time series mitigates any minor downward bias (average RMSE=0.047 vs. 0.052, corresponding a reduction of 9.6%). Conversely, in scenarios marked by structural heterogeneity, as depicted in the upper panel, the availability of more individual-level information does not significantly alleviate the bias observed in parameter estimation (average RMSE=0.122 vs. 0.125, corresponding to a reduction of 2.4%).

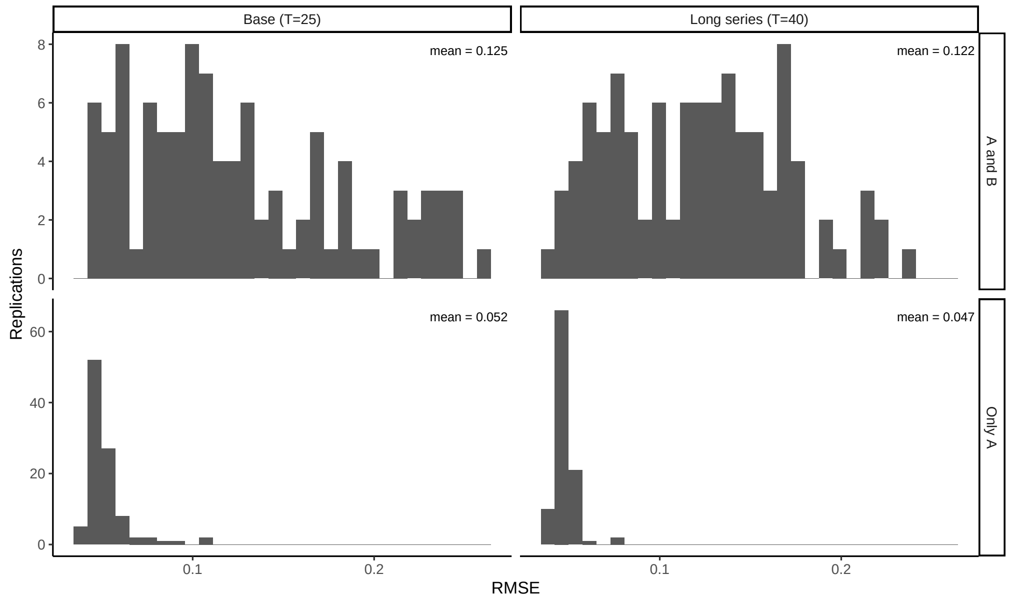


Figure B.13: Histogram of the individual RMSE between the estimated q33 across 100 replications for Segment A customers. The top panel shows the bias when both segments are used in the estimation, whereas the bottom panel shows the bias when only Segment A customers are used in the estimation. The columns represent the short series (baseline), and long series conditions, respectively.

B.3.6 Summary of all experimental conditions

We summarize the results for the main parameters across all analyzed experimental conditions in Table B.4 for the conditional probabilities and Table B.5 for the transition probabilities.

Table B.4:
Average of the individual RMSE between the true and the estimated parameters of the conditional probabilities of the HMM across 100 replications for Segment A customers. A and B indicates that all data were used to estimate the corresponding models, whereas only A indicates that only Segment A data were used to estimate the HMM.

	Estimation	Baseline	Str. Het.	Str. Het. Mixture		Param. dispersion		Dynamics	Seq. Length
State	Data	Base	1/2 states	70%	80%	No het.	High het.	High	Long series
1	A and B	0.036	0.021	0.033	0.031	0.020	0.083	0.045	0.026
	Only A	0.026	0.021	0.033	0.042	0.012	0.058	0.033	0.022
2	A and B	0.077	0.054	0.070	0.067	0.041	0.151	0.127	0.062
	Only A	0.063	0.053	0.070	0.080	0.030	0.127	0.069	0.057
3	A and B	0.040	.	0.039	0.038	0.024	0.091	0.099	0.025
	Only A	0.025	.	0.032	0.038	0.012	0.057	0.036	0.021

Table B.4 presents the results of recovering the conditional probabilities for Segment A over 100 replications. It is observed that recovery generally deteriorates in scenarios with structural heterogeneity (Segments A and B) as opposed to scenarios lacking structural heterogeneity (Segment A only). This observation holds even though the number of customers in the structurally homogeneous scenario (only Segment A) is significantly lower, with 50% fewer customers in the baseline and 20% fewer customers in the 80% mixture case. Moreover, the recovery of conditional probabilities appears more effective for the extreme states (lowest and highest) and less so for the intermediate state.

Table B.5: Average of the individual RMSE between the true and the estimated parameters of the transition probabilities of the HMM across 100 replications for Segment A customers. A and B indicates that all data were used to estimate the corresponding models, whereas only A indicates that only Segment A data were used to estimate the HMM.

State		Estimation	Baseline	Str. Het.	Str. Het. Mixture		Param. dispersion		Dynamics	Seq. Length
From	To	Data	Base	1/2 states	70%	80%	No het.	High het.	High	Long series
1	1	A and B	0.061	0.099	0.059	0.055	0.030	0.109	0.112	0.053
		Only A	0.061	0.049	0.093	0.097	0.031	0.094	0.069	0.056
	2	A and B	0.060	0.099	0.063	0.057	0.050	0.088	0.084	0.058
		Only A	0.049	0.049	0.075	0.071	0.029	0.072	0.064	0.044
	3	A and B	0.030	.	0.032	0.035	0.025	0.036	0.066	0.036
		Only A	0.021	.	0.028	0.035	0.014	0.031	0.041	0.020
2	1	A and B	0.067	0.055	0.082	0.084	0.060	0.101	0.110	0.053
		Only A	0.046	0.049	0.065	0.083	0.025	0.068	0.058	0.039
	2	A and B	0.069	0.055	0.064	0.063	0.049	0.105	0.116	0.094
		Only A	0.068	0.049	0.103	0.141	0.041	0.102	0.088	0.056
	3	A and B	0.072	.	0.094	0.108	0.069	0.075	0.066	0.088
		Only A	0.043	.	0.059	0.074	0.027	0.064	0.055	0.037
3	1	A and B	0.055	.	0.057	0.072	0.035	0.061	0.056	0.057
		Only A	0.038	.	0.038	0.042	0.013	0.054	0.049	0.036
	2	A and B	0.103	.	0.182	0.202	0.074	0.116	0.111	0.105
		Only A	0.042	.	0.045	0.065	0.019	0.060	0.063	0.039
	3	A and B	0.125	.	0.205	0.247	0.096	0.132	0.099	0.122
		Only A	0.052	.	0.055	0.079	0.022	0.075	0.057	0.047

Table B.5 presents the results of recovering the transition probabilities for Segment A across 100 replications. Similar to the conditional probabilities, in general, the recovery deteriorates in scenarios with structural heterogeneity (Segments A and B) as opposed to scenarios without structural heterogeneity (Segment A only). Furthermore, compared to the baseline, the bias consistently rises in tandem with increased structural heterogeneity across various conditions. Specifically, in instances where structural heterogeneity is present, the bias experiences its most significant increase (decrease) when the degree of dynamics (preference heterogeneity) is high (low, no preference heterogeneity). Notably, extending the time series length appears to have only a negligible impact on mitigating bias.

B.4 Random Coefficient Mixture HMM

We investigate an alternative potential solution for addressing the challenge posed by structural heterogeneity in the traditional HMM. In particular, we study the more flexible approach to modeling preference heterogeneity, wherein parameters are drawn from multimodal prior distributions. We extend the heterogeneity structure described in Web Appendix A.3 to capture multi-modality using a Mixture of Gaussians,

$$\boldsymbol{\theta}_i \sim \sum_{c=1}^{C} \psi_c \cdot \mathcal{N}(\boldsymbol{\mu}_{\theta,c}, \Sigma_{\theta,c})$$

where $\boldsymbol{\mu}_{\theta,c}$ and $\Sigma_{\theta,c}$ are the mean vector and covariance matrix respectively for component c, and ψ_c is the mixture probability such that $\psi_c \geq 0, \ \ \sum_c \psi_c = 1$.

This formulation adheres to a discrete-continuous specification of parameter heterogeneity and is known as the Random Coefficient Mixture HMM (RCM-HMM; Kappe et al., 2018).

In particular, we focus on two questions: (i) Does this specification reduce the bias from not accounting for structural heterogeneity? and (ii) Does this specification improve model predictions?

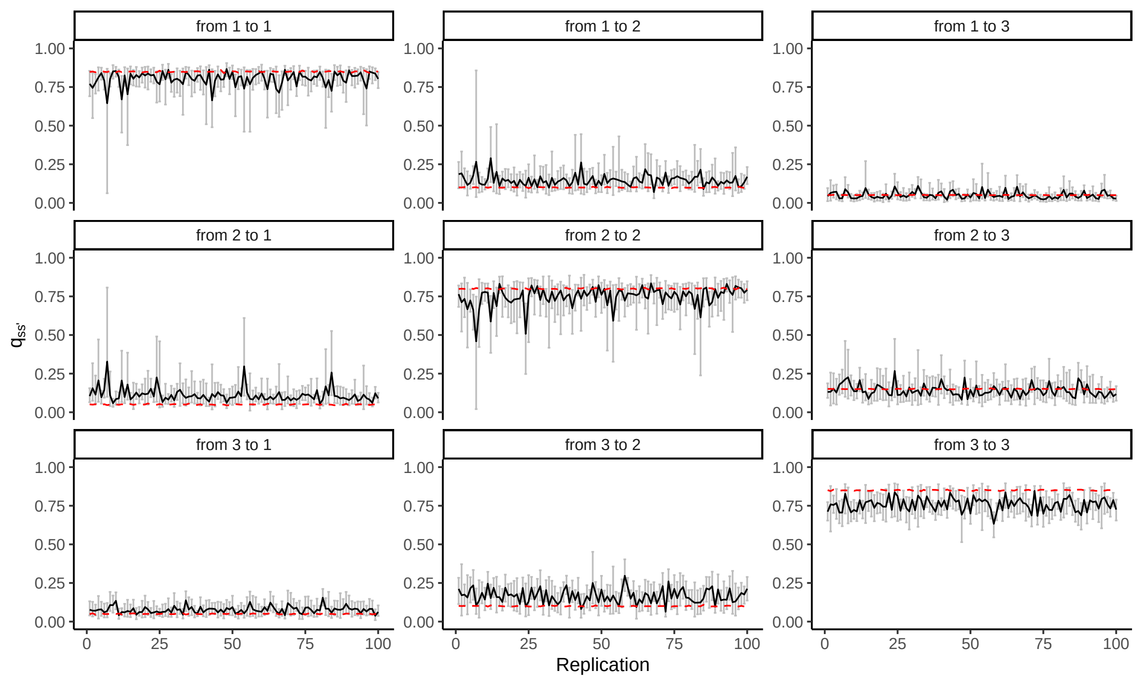
Table B.6: Average of the individual RMSE between the true and the estimated parameters of the transition probabilities of the HMM and RCM-HMM across 100 replications for Segment A customers.

State		Model	Baseline	Str. Het.	Str. Het. Mixture		Param. dispersion		Dynamics	Seq. Length
From	To		Base	1/2 states	70%	80%	No het.	High het.	High	Long series
1	1	HMM	0.061	0.099	0.059	0.055	0.030	0.109	0.112	0.053
		RCM-HMM	0.081	0.082	0.098	0.103	0.058	0.136	0.114	0.070
	2	HMM	0.060	0.099	0.063	0.057	0.050	0.088	0.084	0.058
		RCM-HMM	0.069	0.082	0.078	0.077	0.054	0.101	0.091	0.058
	3	HMM	0.030	.	0.032	0.035	0.025	0.036	0.066	0.036
		RCM-HMM	0.029	.	0.038	0.042	0.025	0.046	0.077	0.026
2	1	HMM	0.067	0.055	0.082	0.084	0.060	0.101	0.110	0.053
		RCM-HMM	0.080	0.117	0.130	0.181	0.076	0.141	0.176	0.057
	2	HMM	0.069	0.055	0.064	0.063	0.049	0.105	0.116	0.094
		RCM-HMM	0.097	0.117	0.152	0.210	0.080	0.178	0.215	0.074
	3	HMM	0.072	.	0.094	0.108	0.069	0.075	0.066	0.088
		RCM-HMM	0.070	.	0.079	0.078	0.061	0.080	0.085	0.065
3	1	HMM	0.055	.	0.057	0.072	0.035	0.061	0.056	0.057
		RCM-HMM	0.060	.	0.065	0.070	0.048	0.071	0.069	0.060
	2	HMM	0.103	.	0.182	0.202	0.074	0.116	0.111	0.105
		RCM-HMM	0.130	.	0.161	0.160	0.116	0.122	0.126	0.142
	3	HMM	0.125	.	0.205	0.247	0.096	0.132	0.099	0.122
		RCM-HMM	0.165	.	0.201	0.204	0.148	0.153	0.131	0.178

B.4.1 Parameter recovery

Figure B.14 shows the posterior mean of the average transition probabilities across individuals in segment A with truly 3 states over all replications. The RCM-HMM estimates show the same patterns displayed by the HMM. In other words, a discrete-continuous heterogeneity specification does not resolve the downward bias in q_{33} exhibited by the HMM for customers with truly 3 states in the presence of structural heterogeneity. The results at the individual level are consistent with this poor recovery shown by the RCM-HMM (see Table B.6 for the individual level RMSE of the transition probabilities for the HMM and RCM-HMM). The discrete-continuous heterogeneity specification not only does not resolve the issues exhibited by the HMM but also worsens them in certain scenarios. For example, when the continuous heterogeneity is high or customers are highly dynamic among states, the RCM-HMM exhibits poorer recovery of the transition probabilities relative to the HMM.

Figure B.14: Recovery of transition probabilities for customers in segment A across 100 replications using the RCM-HMM. The black line represents the posterior mean, and the gray bars the 95% posterior credible interval. True simulated values are in red



B.4.2 Model predictions

We now compare the out-of-sample predictions provided by the traditional HMM and the more flexible RCM-HMM.

Table B.7 shows that the traditional HMM provides better predictions than the more flexible RCM-HMM in most cases. For instance, for the baseline scenario, the Test-LL shows that in 91% of the cases the

HMM outperforms the RCM-HMM whereas the validation RMSE shows that in 93% of the replications the HMM wins. Similarly, we find that the MHMM (described in the "Mixture of HMMs (MHMM)" subsection) outperforms the RCM-HMM on predictive performance on all the replications (see Table B.7). Naturally, when the true data-generating process corresponds to the RCM-HMM (i.e., no structural heterogeneity; 2 segments with three states), the RCM-HMM outperforms both the HMM and the MHMM.

Table B.7: Comparison of RCM-HMM vs. HMM and MHMM across 100 replications by scenario. The values represent the proportion of instances where RCM-HMM performs best based on out-of-sample metrics.

Scenario		Prop. as best model vs. HMM		Prop. as best model vs. MHMM	
		Test-LL	Test-RMSE	Test-LL	Test-RMSE
DGP: Structural heterogeneity HMM (MHMM)					
Baseline		0.09	0.07	0.00	0.00
Str. Het. Mixture	λ=70%	0.42	0.33	0.25	0.21
	λ=80%	0.37	0.28	0.29	0.25
Structural het.	1/2 states	0.45	0.52	0.28	0.24
Param. dispersion	No het.	0.41	0.35	0.18	0.17
	High het.	0.32	0.39	0.16	0.11
Dynamics	High	0.46	0.52	0.07	0.08
Seq. Length	Long series	0.46	0.41	0.22	0.18
DGP: HMM					
No Struct. het.	1 segment	0.00	0.00	0.35	0.80
DGP: RCM-HMM					
No Struct. het.	2 segments	0.93	0.63	0.96	0.94

B.5 Mixture of Hidden Markov Models (MHMM)

We first explore how the MHMM performs under the different experimental conditions analyzed previously for the traditional HMM. We compare the out-of-sample fit and predictions by using the expected log predictive posterior (Test-LL) and the outcome root mean square error (Test-RMSE) on the test set, respectively.[2]

Table B.8 compares the predictive ability of HMM and MHMM across 100 replications considering the holdout period for different experimental conditions. Table B.8 shows that the proposed MHMM provides systematically better predictions across experimental conditions except when there is no struc-

[2]The 3-states HMM yields a better in-sample fit than the MHMM, which may seem counterintuitive given that the 3-states HMM is one of the components of the MHMM. While this outcome may seem surprising from a frequentist perspective, it is not unusual in Bayesian models that integrate over the posterior, as they account for both estimation uncertainty and model complexity since a more complex model needs to integrate the posterior over a higher-dimensional space (Chapter 22.3 in Gelman et al., 2020).

Table B.8: Best model by scenario:
proportion of replications per scenario that MHMM is chosen as the best on Test-LL and Test-RMSE vs. HMM.

Scenario		Prop. as best model	
		Test-LL	Test-RMSE
DGP: Structural heterogeneity HMM (MHMM)			
Baseline		0.99	0.99
Str. Het. Mixture	$\lambda = 70\%$	0.97	0.92
	$\lambda = 80\%$	0.93	0.85
	$\lambda = 90\%$	0.61	0.60
	$\lambda = 95\%$	0.36	0.52
	$\lambda = 98\%$	0.44	0.72
Structural het.	1/2 states	1.00	1.00
Param. dispersion	No het.	0.98	0.95
	High het.	0.99	0.98
Dynamics	High	1.00	1.00
Seq. Length	Long series	1.00	1.00
DGP: HMM			
No Struct. het.	1 segment	0.02	0.00
DGP: RCM-HMM			
No Struct. het.	2 segments	0.15	0.05

tural heterogeneity in the data (the vast majority of the individuals in the data have 3 states). In that case, as expected, the traditional HMM performs better.

B.5.1 Parameter recovery

We now analyze the recovery of the model for the baseline condition. The parameter estimates across replications are illustrated in Figure B.15 (conditional probabilities P for each mixture), Figure B.16 (transition matrix Q for 2-state customers), and Figure B.17 (transition matrix Q for 3-state customers). Figure B.15 shows that conditional probabilities are well recovered for both Segments. Similarly, Figure B.16 and Figure B.17 show that the transition matrices are well recovered for each segment, respectively.

Now, one of the main drawbacks of estimating an HMM in the presence of structural heterogeneity is that even the estimates of customers with the same number of states as the estimated HMM may be biased. To properly compare the two models, we contrast the parameter recovery of the 3-states MHMM with the HMM on Segment A customers. We show state-dependent probabilities in Table B.9 and transition probabilities in Table B.10 across scenarios. We find that the MHMM provides a considerably lower RMSE than the HMM in the third row of the transition matrix in the Baseline scenario. We find that this result is robust across most scenarios. We only find the HMM performing better when the proportion of

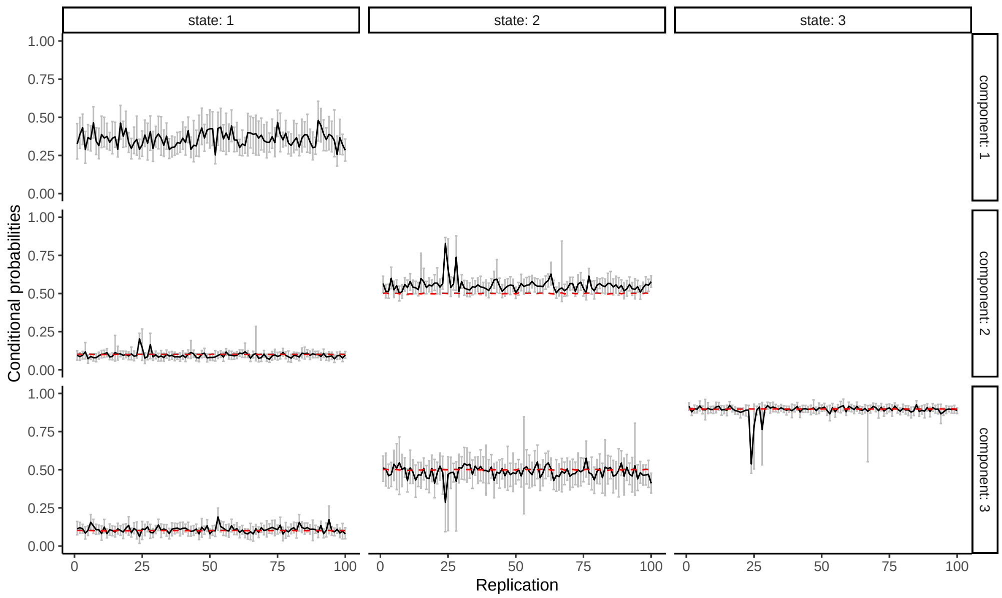


Figure B.15: MHMM state-dependent, Baseline scenario. Recovery of population state-dependent probabilities for each MHMM component across 100 replications. Population average is across all observations regardless of individual component membership. True simulated values are in red.

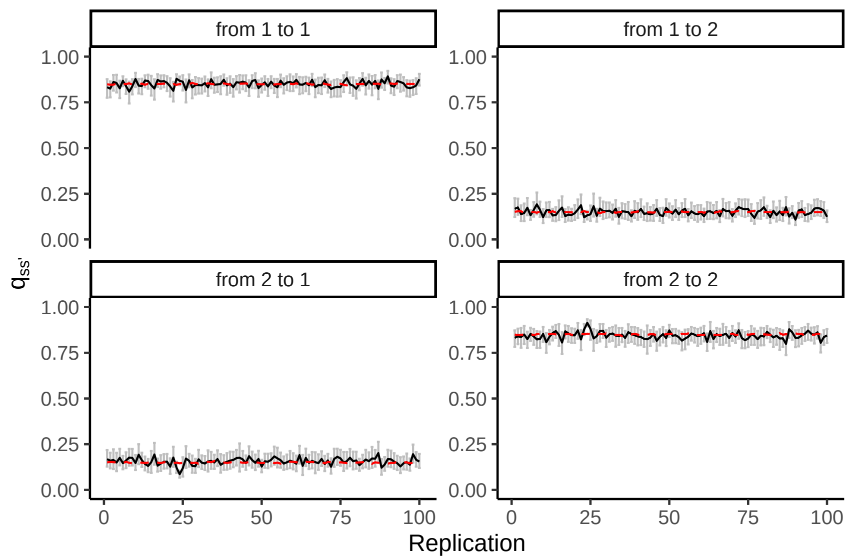


Figure B.16: MHMM 2-states transitions, Baseline scenario. Recovery of population transition probabilities for the 2-state MHMM component across 100 replications. Population average is across all observations regardless of individual component membership. True simulated values are in red.

customers with 3 states is low. However, this is not translated to worse predictive performance as shown in Table B.8 since the 2-states component of the MHMM may likely dominate the model in such instances.

Following the analysis in the main text, we corroborate that parameters are identified. Figures B.18 and B.19 show the posterior uncertainty of the transition matrices for the 2-state and 3-state components, for one replication of the *Baseline* scenario. Naturally, uncertainty is higher than in the case with no

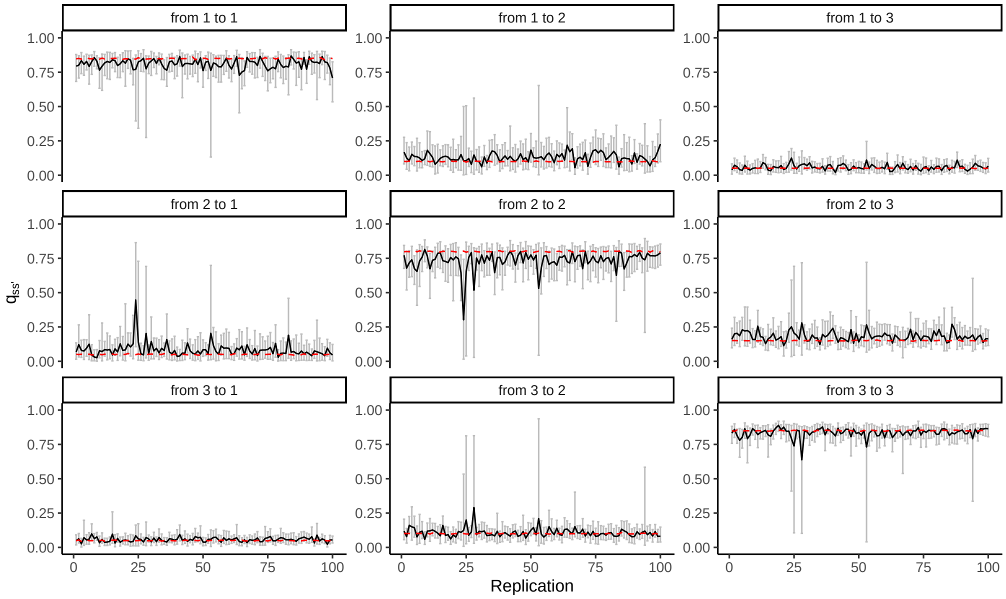


Figure B.17: MHMM 3-states transitions, Baseline scenario. Recovery of population transition probabilities for the 3-state MHMM component across 100 replications. Population average is across all observations regardless of individual component membership. True simulated values are in red.

Table B.9: Average of the individual RMSE between the true and the estimated state-dependent probabilities of the HMM vs. 3-state component of the MHMM for customers in Segment A across 100 replications.

	Estimated	Baseline	Str. Het.	Str. Het. Mixture		Param. dispersion		Dynamics	Seq. Length
State	Model	Base	1/2 states	70%	80%	No het.	High het.	High	Long series
1	HMM	0.036	0.021	0.033	0.031	0.020	0.083	0.045	0.026
	MHMM	0.028	0.022	0.046	0.060	0.014	0.062	0.042	0.023
2	HMM	0.077	0.054	0.070	0.067	0.041	0.151	0.127	0.062
	MHMM	0.065	0.054	0.085	0.105	0.037	0.133	0.080	0.057
3	HMM	0.040	.	0.039	0.038	0.024	0.091	0.099	0.025
	MHMM	0.029	.	0.066	0.097	0.023	0.064	0.060	0.021

structural heterogeneity, as there is additional uncertainty associated with each customer belonging to each component. Nevertheless, transitions are identified, especially on the high state for the 3-state component, which, in contrast with the HMM, their intervals do not cover the entire [0,1] range as in the HMM case shown in Figure 5a.

Table B.10: Average of the individual RMSE between the true and the estimated transition probabilities of the HMM vs. 3-state component of the MHMM for customers in Segment A across 100 replications.

State		Estimated	Baseline	Str. Het.	Str. Het. Mixture		Param. dispersion		Dynamics	Seq. Length
From	To	Model	Base	1/2 states	70%	80%	No het.	High het.	High	Long series
1	1	HMM	0.061	0.099	0.059	0.055	0.030	0.109	0.112	0.053
		MHMM	0.070	0.056	0.110	0.108	0.040	0.109	0.089	0.058
	2	HMM	0.060	0.099	0.063	0.057	0.050	0.088	0.084	0.058
		MHMM	0.056	0.056	0.081	0.074	0.035	0.082	0.066	0.046
	3	HMM	0.030	.	0.032	0.035	0.025	0.036	0.066	0.036
		MHMM	0.024	.	0.038	0.044	0.017	0.035	0.050	0.021
2	1	HMM	0.067	0.055	0.082	0.084	0.060	0.101	0.110	0.053
		MHMM	0.056	0.052	0.123	0.176	0.038	0.096	0.122	0.040
	2	HMM	0.069	0.055	0.064	0.063	0.049	0.105	0.116	0.094
		MHMM	0.088	0.052	0.179	0.248	0.068	0.140	0.180	0.059
	3	HMM	0.072	.	0.094	0.108	0.069	0.075	0.066	0.088
		MHMM	0.051	.	0.073	0.089	0.039	0.071	0.077	0.039
3	1	HMM	0.055	.	0.057	0.072	0.035	0.061	0.056	0.057
		MHMM	0.037	.	0.043	0.054	0.013	0.061	0.057	0.038
	2	HMM	0.103	.	0.182	0.202	0.074	0.116	0.111	0.105
		MHMM	0.044	.	0.062	0.106	0.025	0.068	0.063	0.038
	3	HMM	0.125	.	0.205	0.247	0.096	0.132	0.099	0.122
		MHMM	0.052	.	0.079	0.141	0.027	0.090	0.071	0.048

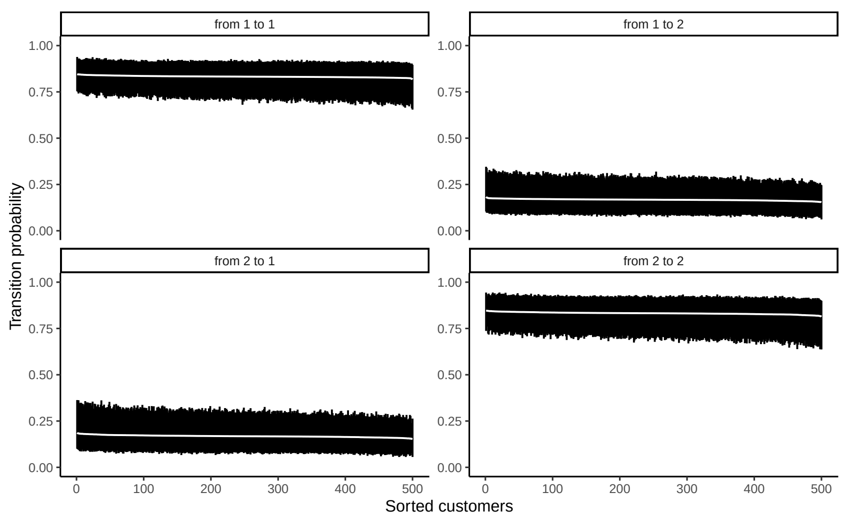


Figure B.18: Posterior mean and 95% confidence interval for individual transition probabilities for 2-state component. Customers from both segments are included. In each panel, customers are sorted decreasingly on their posterior mean. The shaded area represents the 95% intervals and the white line represents the posterior mean.

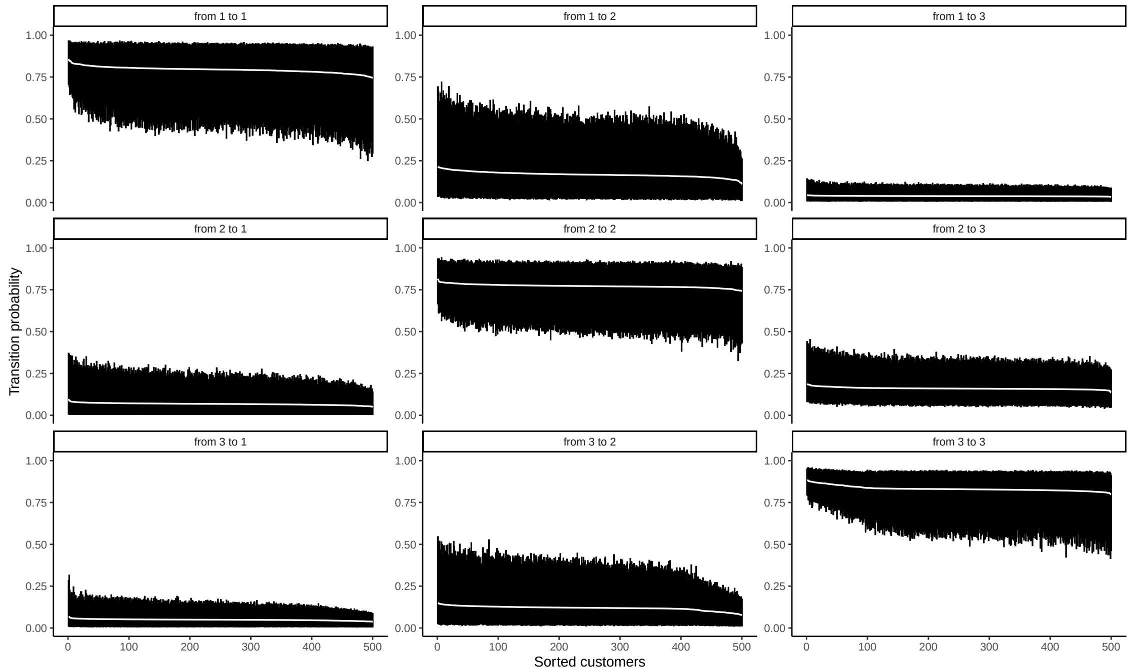


Figure B.19: Posterior mean and 95% confidence interval for individual transition probabilities for 3-state component. Customers from both segments are included. In each panel, customers are sorted decreasingly on their posterior mean. The shaded area represents the 95% intervals and the white line represents the posterior mean.

B.5.2 Structural heterogeneity recovery

We now analyze the MHMM recovery of the number of states at the individual level. To do so, we first compute the individual posterior component-membership probabilities as follows,

$$p(NS_i = S_k|Y_{i,1:T}, \theta_i, \Phi) = \frac{\omega_k \cdot \cdot p(Y_{i,1:T}|NS_i = S_m)}{\omega_1 \cdot p(Y_{i,1:T}|NS_i = 1) + \omega_2 \cdot p(Y_{i,1:T}|NS_i = 2) + \omega_3 \cdot p(Y_{i,1:T}|NS_i = 3)}$$
$$\propto \omega_k \cdot L_{i,\text{HMM}}(p_i^k, Q_i^k|Y_{i,1:T}, NS_i = S_k), \tag{B.4}$$

where $L_{i,\text{HMM}}(p_i^k, Q_i^k|Y_{i,1:T}, NS_i = S_k)$ is the individual likelihood of an HMM associated with component m introduced in Equation (6). Accordingly, we classify customers to the component with the highest individual posterior component probabilities, as follows:

$$\widetilde{S}_i(\theta, Y_{i,1:T}) = \arg\max_k \left\{\omega_k \cdot L_{i,\text{HMM}}(p_i^k, Q_i^k|Y_{i,1:T}, NS_i = S_k)\right\}.$$

We compute the proportion of customers classified in each component and integrate over the posterior of model parameters (since the classification is conditional on the model parameters). We compute such proportions for the entire population and for each segment separately:

$$pS_k = \frac{1}{n_a + n_B} \int_\theta \sum_i \mathbf{1}\{\widetilde{S}_i(\theta, y_{i,1:T}) = k\} \cdot p(\theta|\mathcal{D}) d\theta$$

$$pS_k(A) = \frac{1}{n_A} \int_\theta \sum_{i \in A} \mathbf{1}\{\widetilde{S}_i(\theta, y_{i,1:T}) = k\} \cdot p(\theta|\mathcal{D}) d\theta$$

$$pS_k(B) = \frac{1}{n_B} \int_\theta \sum_{i \in B} \mathbf{1}\{\widetilde{S}_i(\theta, y_{i,1:T}) = k\} \cdot p(\theta|\mathcal{D}) d\theta.$$

Table B.11 shows the mean and 95% confidence intervals for the memberships to each structural segment across replications in the case of the Baseline scenario. We also show the classification for customers simulated with 3 states (Segment A) and 2 states (Segment B) separately. We find that 63.8% of customers in Segment B and 70.67% in Segment A are correctly classified.

Table B.11: Inferred states. Proportion of customers on each MHMM component for the baseline scenario.

		Estimated (k)		
True		1	2	3
	pS_k	16.0% [15.7% , 16.4%]	42.4% [41.9% , 42.9%]	41.6% [41.2% , 42.0%]
By segment				
2	$pS_k(B)$	23.7% [23.0% , 24.5%]	63.8% [62.2% , 65.3%]	12.5% [11.2% , 13.8%]
3	$pS_k(A)$	8.37% [7.83% , 8.91%]	20.96% [19.49% , 22.42%]	70.67% [69.24% , 72.11%]

We also show the correctly classified proportions (diagonal of the confusion matrix) across scenarios in Table B.12. We find that the MHMM is more accurate at correctly classifying customers when the structural mixture is static compared to the dynamic case (*Structural het.: 1/2 states*). We also find that the model significantly improves with longer time series. Finally, as expected, the MHMM decreases its classification performance with higher parametric heterogeneity.

B.5.3 Model predictions

We now compare the predictive performances of the MHMM and HMM. Table B.13 reports the proportion of replications in which the MHMM performs better than the HMM, considering Test-LL and Test-RMSE metrics, across scenarios.

B.6 Simulation with marketing decisions

Our study further explores the impact of structural heterogeneity on marketing decisions by incorporating marketing covariates into our simulation model (Equation (4)).

Table B.12: Correctly classified customers per segment by scenario (recall)

Scenario	True/Estimated 1	2	3
Baseline	. .	63.8% [62.2% , 65.3%]	70.67% [69.24% , 72.11%]
Structural het.: 1/2 states	81.1% [80.0% , 82.2%]	90.5% [90.09% , 90.91%]	. .
Param. dispersion: No het.	. .	66.0% [63.59% , 68.4%]	73.26% [70.98% , 75.55%]
Param. dispersion: High het.	. .	51.4% [49.6% , 53.1%]	58.8% [57.05% , 60.5%]
Dynamics: High	. .	51.3% [48.4% , 54.2%]	64.76% [61.93% , 67.59%]
Seq. length: Long	. .	74.85% [73.67% , 76.03%]	83.04% [82.26% , 83.82%]

Table B.13: Proportion of replications
that MHMM is chosen as the best model compared to HMM on Test-LL and Test-RMSE across scenarios

Scenario		Prop. MHMM predicts better than HMM: Test-LL	Test-RMSE
Baseline		0.99	0.99
Str. Het. Mixture	$\lambda = 70\%$	0.97	0.92
	$\lambda = 80\%$	0.93	0.85
Structural het.	1/2 states	1.00	1.00
Param. dispersion	No het.	0.98	0.95
	High het.	0.99	0.98
Dynamics	High	1.00	1.00
Seq. Length	Long series	1.00	1.00

B.6.1 Design

To assess the influence of these covariates and establish a clear framework for our investigation, we consider the marketing variable to represent the execution of a promotional campaign for a particular customer within a defined timeframe. This involved conducting simulations with a Bernoulli-distributed covariate that takes the values of 0 or 1, which was integrated into the transitions of the HMM through an ordered logistic approach.[3]

We define the transition probabilities with the targeted promotion (X=1) and without the promotion

[3] Alternatively, since the covariate is a binary variable, one could also model transitions using two matrices, one for each covariate value.

(X=0); with the latter aligned with the *Baseline* scenario. Table B.14 displays the transition probabilities of each segment, with and without the promotion. Note that the effects are both state and segment specific. For customers in Segment A, the promotion strongly favors transitions to higher states if the customer is in state 1, mildly favors transitions to higher states if in state 2, and favors transitions to lower states if in state 3 when exposed to the promotion. Conversely, for customers in Segment B, the promotion mildly favors transitions to higher states in state 1 and lower states in state 2.

Table B.14: Simulated segment-average transition matrices with covariate effects.

Segment		Without promotion $X = 0$			With promotion $X = 1$		
	from/to	1	2	3	1	2	3
A	1	0.85	0.10	0.05	0.44	0.30	0.26
	2	0.05	0.80	0.15	0.02	0.66	0.32
	3	0.05	0.10	0.85	0.13	0.19	0.68
B	1	0.85	0.15	.	0.68	0.32	.
	2	0.15	0.85	.	0.32	0.68	.

To facilitate a meaningful comparison of these findings with those from previous analyses, we generated a long series of individual data (T=40), given that the inclusion of covariates necessitates the estimation of a greater variety of transition types compared to the baseline scenario.

B.6.2 Parameter estimates

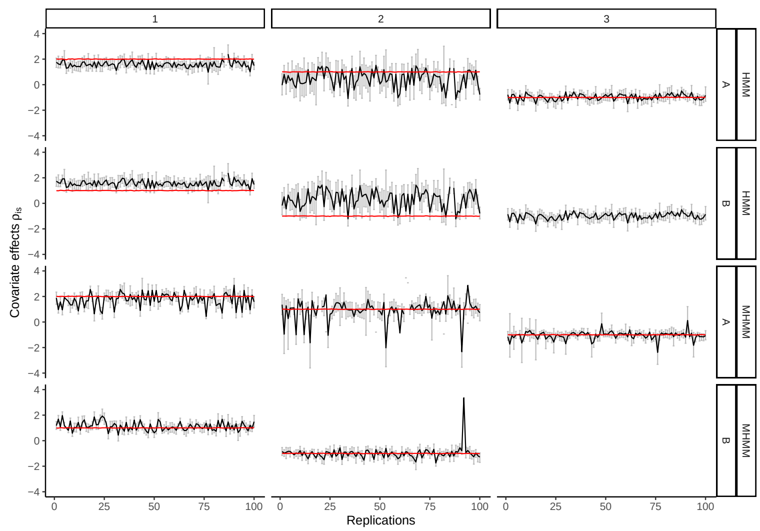
We present in Table B.15 the estimated transitions for each segment under the HMM, RCM-HMM and MHMM models across replications. The initial rows of Table B.15 present the average simulated transition probabilities for each customer segment under both covariate conditions (X=0,1) over 100 replications. The next rows illustrate the recovery of transition probabilities for the HMM, RCM-HMM and MHMM, separately for each segment. These probabilities are shown under two distinct conditions: when the covariate is set to 0 and when it is set to 1.

Table B.15: Transition matrices as a function of covariates of HMM, RCM-HMM and MHMM. For each model, we compute the segment-average posterior mean, averaged across all replications. In parentheses below, we show the 2.5% and 97.5% quantiles across replications of each replication posterior mean (i.e., the variability of segment-average posterior means across all 100 replications).

Model	Segment		X = 0			X = 1		
		from/to	1	2	3	1	2	3
True	A	1	0.85	0.10	0.05	0.44	0.30	0.26
		2	0.05	0.80	0.15	0.02	0.66	0.32
		3	0.05	0.10	0.85	0.13	0.19	0.68
	B	1	0.85	0.15	.	0.68	0.32	.
		2	0.15	0.85	.	0.32	0.68	.
HMM	A	1	0.78	0.14	0.08	0.44	0.31	0.25
			[0.71 , 0.84]	[0.07 , 0.22]	[0.04 , 0.13]	[0.33 , 0.53]	[0.16 , 0.42]	[0.15 , 0.38]
		2	0.10	0.74	0.17	0.10	0.65	0.25
			[0.05 , 0.18]	[0.55 , 0.82]	[0.09 , 0.33]	[0.02 , 0.26]	[0.45 , 0.75]	[0.06 , 0.37]
		3	0.06	0.11	0.83	0.15	0.17	0.68
			[0.04 , 0.09]	[0.06 , 0.19]	[0.72 , 0.88]	[0.12 , 0.20]	[0.10 , 0.25]	[0.60 , 0.74]
	B	1	0.79	0.15	0.06	0.46	0.35	0.19
			[0.71 , 0.88]	[0.09 , 0.22]	[0.01 , 0.10]	[0.34 , 0.61]	[0.26 , 0.43]	[0.05 , 0.33]
		2	0.11	0.77	0.12	0.12	0.68	0.19
			[0.05 , 0.20]	[0.64 , 0.84]	[0.04 , 0.20]	[0.02 , 0.32]	[0.55 , 0.77]	[0.03 , 0.33]
		3	0.07	0.20	0.73	0.16	0.27	0.57
			[0.04 , 0.10]	[0.07 , 0.63]	[0.31 , 0.88]	[0.13 , 0.22]	[0.12 , 0.63]	[0.23 , 0.72]
RCM-HMM	A	1	0.79	0.14	0.07	0.45	0.31	0.24
			[0.73 , 0.86]	[0.07 , 0.21]	[0.03 , 0.12]	[0.35 , 0.53]	[0.19 , 0.43]	[0.11 , 0.36]
		2	0.09	0.73	0.17	0.10	0.65	0.25
			[0.05 , 0.15]	[0.58 , 0.82]	[0.10 , 0.30]	[0.02 , 0.26]	[0.49 , 0.74]	[0.10 , 0.39]
		3	0.07	0.11	0.82	0.16	0.17	0.67
			[0.05 , 0.10]	[0.06 , 0.20]	[0.70 , 0.89]	[0.12 , 0.21]	[0.12 , 0.24]	[0.59 , 0.73]
	B	1	0.80	0.15	0.05	0.48	0.34	0.17
			[0.73 , 0.87]	[0.09 , 0.21]	[0.01 , 0.10]	[0.38 , 0.62]	[0.26 , 0.43]	[0.04 , 0.30]
		2	0.11	0.78	0.11	0.14	0.69	0.17
			[0.05 , 0.17]	[0.70 , 0.84]	[0.05 , 0.18]	[0.02 , 0.31]	[0.60 , 0.77]	[0.04 , 0.32]
		3	0.08	0.22	0.69	0.18	0.28	0.54
			[0.05 , 0.19]	[0.07 , 0.64]	[0.27 , 0.87]	[0.13 , 0.35]	[0.15 , 0.63]	[0.22 , 0.71]
MHMM	A	1	0.82	0.11	0.06	0.47	0.30	0.23
			[0.74 , 0.89]	[0.07 , 0.18]	[0.03 , 0.10]	[0.34 , 0.73]	[0.16 , 0.41]	[0.09 , 0.32]
		2	0.09	0.68	0.22	0.07	0.55	0.38
			[0.03 , 0.25]	[0.27 , 0.83]	[0.11 , 0.63]	[0.006 , 0.35]	[0.08 , 0.70]	[0.25 , 0.90]
		3	0.07	0.14	0.79	0.18	0.21	0.62
			[0.04 , 0.21]	[0.07 , 0.43]	[0.40 , 0.88]	[0.09 , 0.45]	[0.13 , 0.36]	[0.24 , 0.73]
	B	1	0.86	0.14	.	0.66	0.34	.
			[0.82 , 0.91]	[0.09 , 0.18]	.	[0.52 , 0.74]	[0.26 , 0.48]	.
		2	0.14	0.86	.	0.32	0.68	.
			[0.09 , 0.19]	[0.81 , 0.91]	.	[0.21 , 0.40]	[0.60 , 0.79]	.

For the HMM, the estimated model tends to overestimate the dynamics of customer transitions. That is, it predicts that customers are more likely to switch states than they actually are. This tendency is particularly pronounced for customers in Segment B. Notably, this overestimation of dynamics is also observed among Segment A customers, despite the correct number of states being identified. In this case, the stickiness of state 1 when the covariate equals 0 is underestimated (0.78 vs. 0.85). The RCM-HMM estimates are similar to those of the HMM. In contrast, the MHMM recovers these transition probabilities more accurately on average. We analyze the recovery of the covariate effect ρ_s by model and segment in Figure B.20.

Figure B.20: Recovery of covariate effects ρ_{is} for the HMM and MHMM across 100 replications. The top two panels show segments A and B for the HMM, and the bottom two panels show segments A and B for the MHMM. Each column represents a state. The black line represents the posterior mean, and the gray bars the 95% posterior credible interval on each replication. True simulated values are in red.

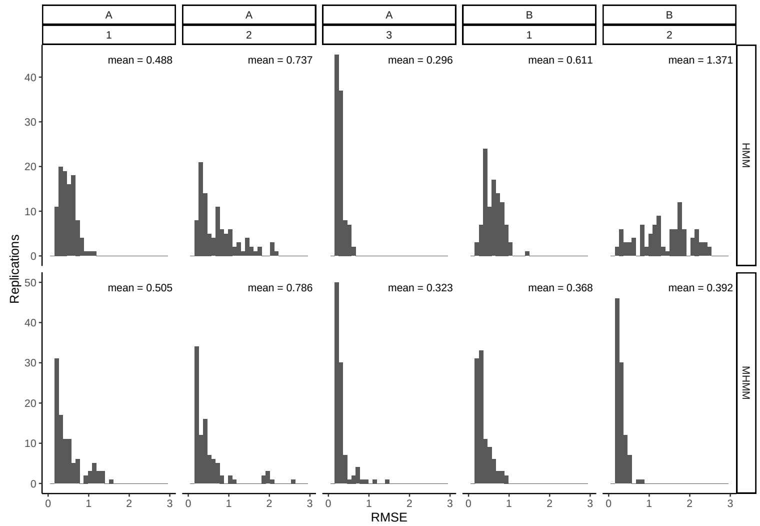


For customers in Segment B (see the second row of Figure B.20), we observe that the HMM incorrectly estimates the impact that the covariates have on the corresponding dynamics. In contrast, the MHMM infers the covariate effects on these transition probabilities substantially better. For customers in Segment A, both models recover the effects similarly. These results also replicate at the individual level (see individual level RMSE's of covariate effects ρ_{is} in Figure B.21). These differences may translate into better decisions compared to those implied by the HMM when targeting based on these estimates in the presence of structural heterogeneity.

B.6.3 Targeting

For each model, we compute the individual-level expected lift (τ_i) of targeting the incentive to salesperson i using the corresponding estimated model parameters. That is, the lift is computed as the expected difference in outcomes between receiving the incentive and not receiving the incentive. Since marketing actions affect transitions between the latent states, these marketing actions can have an enduring effect on salespeople. We evaluate the differential impact of being targeted on the outcomes considering several periods after the intervention. Specifically, we compute the lift of an incentive ($X_{i\bar{T}} = 1$ vs. $X_{i\bar{T}} = 0$) implemented on period $\bar{T} = 40$ after calibration and evaluate its impact on the subsequent $T_e = 12$

Figure B.21: Covariates condition. Histogram of individual RMSE of covariate effects ρ_{is} by segment by model.



periods. That is, we define the salesperson level lift by

$$\tau_i = E\left(\sum_{t=\bar{T}}^{\bar{T}+T_e} Y_{it} \middle| X_{i\bar{T}} = 1\right) - E\left(\sum_{t=\bar{T}}^{\bar{T}+T_e} Y_{it} \middle| X_{i\bar{T}} = 0\right), \tag{B.5}$$

where each expectation is taken using the parameters of the model.[4]

Then, for each model, we define the policy by targeting salespeople based on their expected lift. Subsequently, we evaluate each policy by simulating outcomes from the true data-generating process and contrast their performance with the optimal policy derived from the true simulated parameters. This comparative analysis allows us to gauge how a policy, developed using a potentially misspecified model, deviates from optimality.

We explore two possible targeting policies:

- Positive Lift: We target all salespeople with a positive expected lift. Any deviations from the optimal allocation stem from the model's inability to accurately estimate the true lift, potentially leading to the inclusion of salespeople with a negative lift or the omission of those with a true positive lift.

- Top 10%: We target the top 10% of salespeople with the highest expected lift. To isolate the model's ability to rank salespeople from its ability to estimate lift levels, we include salespeople even if their expected lift is negative, as the previous scenario already addresses the model's ability to infer the lift's level.

[4] Each expectation corresponds to the posterior predictive expectation given the training data. In Equation (B.5), we omit such reference to simplify the notation.

B.7 MHMM performance for data generated following an HMM

So far, we have compared the performance of the HMM to MHMM when data was generated following the MHMM. We now turn to comparing the two models when data is generated following the HMM. We first look at a case of data generated from an HMM with continuous parametric heterogeneity, followed by a DGP of an HMM with heterogeneity captured by two discrete segments. Specifically, we simulated two segments, each comprising three states with different parameters. These segments, A and B, are simulated such that their segment-average transition matrices are

$$\mathbf{Q}^A = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.05 & 0.80 & 0.15 \\ 0.05 & 0.10 & 0.85 \end{bmatrix} \text{ and } \mathbf{Q}^B = \begin{bmatrix} 0.70 & 0.20 & 0.10 \\ 0.10 & 0.70 & 0.20 \\ 0.10 & 0.20 & 0.70 \end{bmatrix}, \tag{B.6}$$

while all other design parameters are the same as the *Baseline* scenario. Similarly to the other simulation conditions, we generate 100 replications for this scenario.

We estimate the HMM with 3 states and the MHMM with 1, 2, and 3 states for each replication. Table B.16 shows, for the *Baseline* and *Structural homogeneity* scenarios, the proportion of replications that each model performs the best in the test holdout periods on expected log predictive density (Test-LL) and root mean square error (Test-RMSE). We find that when there is no structural heterogeneity, as expected, the HMM outperforms the MHMM.

Table B.16: Proportion of replications
that the MHMM is chosen as the best model compared to HMM on Test-LL and Test-RMSE across scenarios

Scenario		Prop. as best model	
		Test-LL	Test-RMSE
Baseline	50% 2 states - 50% 3 states	0.99	0.99
Structural homogeneity	Two segments, both with 3 states	0.26	0.15

Table B.17: Proportion of replications
that the RCM-HMM is chosen as the best model compared to HMM on Test-LL and Test-RMSE across scenarios

Scenario		Prop. as best model	
		Test-LL	Test-RMSE
Baseline	50% 2 states - 50% 3 states	0.41	0.37
Structural homogeneity	Two segments, both with 3 states	0.38	0.36

Finally, we also contrast in Table B.17 the HMM with the RCM-HMM model that accounts for parametric discrete-continuous heterogeneity but not structural heterogeneity. We find that this model does not improve the HMM in the majority of replications.

The estimation of individual transition probabilities follows a similar trend. Unlike the baseline scenario, the HMM performs better (especially in q_{33}) at inferring transitions compared to the MHMM when there is no structural heterogeneity, even when discrete segments of parametric heterogeneity are present (see Table B.18).

Table B.18: Average of the individual RMSE between the true and estimated transition probabilities of the HMM vs. 3-state component of the MHMM across 100 replications

State		Estimated	Baseline	Structural homogeneity
From	To	Model	Base	Two segments w/ 3 states
1	1	HMM	0.061	0.102
		MHMM	0.070	0.085
	2	HMM	0.060	0.074
		MHMM	0.056	0.074
	3	HMM	0.030	0.032
		MHMM	0.024	0.022
2	1	HMM	0.067	0.049
		MHMM	0.056	0.043
	2	HMM	0.069	0.085
		MHMM	0.088	0.068
	3	HMM	0.072	0.052
		MHMM	0.051	0.046
3	1	HMM	0.055	0.044
		MHMM	0.037	0.044
	2	HMM	0.103	0.066
		MHMM	0.044	0.063
	3	HMM	0.125	0.089
		MHMM	0.052	0.085

C Why an Unconstrained Enlarged HMM Does Not Recover the MHMM?

This appendix examines whether structural heterogeneity in the number of latent states can be recovered by estimating a single, larger unconstrained HMM, instead of explicitly imposing the mixture structure of the MHMM. While the MHMM can be represented as an enlarged HMM with a block-diagonal transition matrix, this representation crucially relies on zero restrictions on off-diagonal transitions. We assess whether these restrictions emerge naturally when estimating an unconstrained model.

In an unconstrained enlarged HMM, individuals always have a non-zero probability of transitioning across all states. When the true DGP follows an MHMM, this implies that transitions into states outside a customer's true structural component are allowed but rarely observed, rendering the associated transition probabilities weakly identified, affecting the identification of all states. We demonstrate this issue empirically. That is, using the simulated data from the *Baseline* scenario, we estimate an unconstrained five-state HMM, which, in principle, is sufficiently large to nest the true structures (two- and three-state HMMs).

Table C.1 reports the proportion of replications in which each model (five-state HMM vs. MHMM) is selected as best based on expected log predictive density and out-of-sample RMSE on the held-out test set. We observe that the unconstrained five-state HMM does not improve predictive performance relative to the MHMM (with additional structure).

Table C.1: Proportion of replications with each model as best on expected log predictive density on the test set (LL) and out-of-sample predictions (RMSE) of an unconstrained 5-states HMM v. MHMM.

			Prop. as best model	
Scenario	Model	States	LL	RMSE
Base	HMM	5	0.02	0.03
	MHMM	1, 2 and 3	0.98	0.97

Beyond prediction, we find that the unconstrained larger HMM model does not recover the true DGP. Tables C.2 and C.3 report posterior means and 95% credible intervals for the transition and state-dependent probabilities of the unconstrained five-state HMM. Table C.2 presents results for all customers, while Table C.3 focuses on customers belonging to the three-state component in the DGP. Despite the MHMM DGP, the estimated transition matrices do not exhibit a block-diagonal structure. Off-block transition probabilities remain non-zero, and their credible intervals do not concentrate near zero, indicating that the unconstrained model does not recover the MHMM structure even at the population level.

At the individual level, Figure C.1 reports posterior means and 95% credible intervals for individual-specific transition probabilities. Customers are sorted by posterior mean within each panel. We find that individual-level transition probabilities are not identified for most transitions among states.

These results show that estimating a larger unconstrained HMM does not recover the MHMM representation. Without explicitly imposing block-diagonal restrictions, the expanded model remains weakly identified and fails to isolate structural heterogeneity in the number of states. Consequently, structural constraints are essential, and simply enlarging the HMM state space (without explicitly imposing the structure) does not constitute a valid alternative to the MHMM.

Table C.2: Posterior mean (and 95% CPI) of population-level conditional and transition probabilities for 5-states HMM without constraints in the transition matrix. Baseline scenario, one replication.

(a) Conditional probabilities for 5-states HMM without constraints in the transition matrix.

	1	2	3	4	5
$\mathbf{p_s}$	0.111	0.380	0.513	0.686	0.901
	[0.079 0.141]	[0.123 0.535]	[0.283 0.848]	[0.478 0.920]	[0.826 0.978]

(b) Transition probabilities for 5-states HMM without constraints in the transition matrix.

		Population Mean				
		1	**2**	**3**	**4**	**5**
Q	**1**	0.770	0.111	0.065	0.039	0.014
		[0.123 0.854]	[0.001 0.556]	[0.000 0.194]	[0.000 0.180]	[0.002 0.038]
	2	0.199	0.338	0.209	0.189	0.064
		[0.004 0.660]	[0.005 0.813]	[0.002 0.730]	[0.002 0.616]	[0.006 0.225]
	3	0.151	0.215	0.330	0.228	0.076
		[0.003 0.481]	[0.002 0.722]	[0.004 0.814]	[0.003 0.732]	[0.007 0.268]
	4	0.092	0.112	0.164	0.532	0.100
		[0.003 0.363]	[0.002 0.482]	[0.003 0.608]	[0.020 0.843]	[0.011 0.310]
	5	0.078	0.119	0.201	0.184	0.417
		[0.005 0.344]	[0.003 0.489]	[0.004 0.503]	[0.004 0.562]	[0.014 0.823]

Table C.3: Posterior mean (and 95% CPI) of Segment A customers (3 states) of conditional and transition probabilities for 5-states HMM without constraints in the transition matrix. Baseline scenario, one replication.

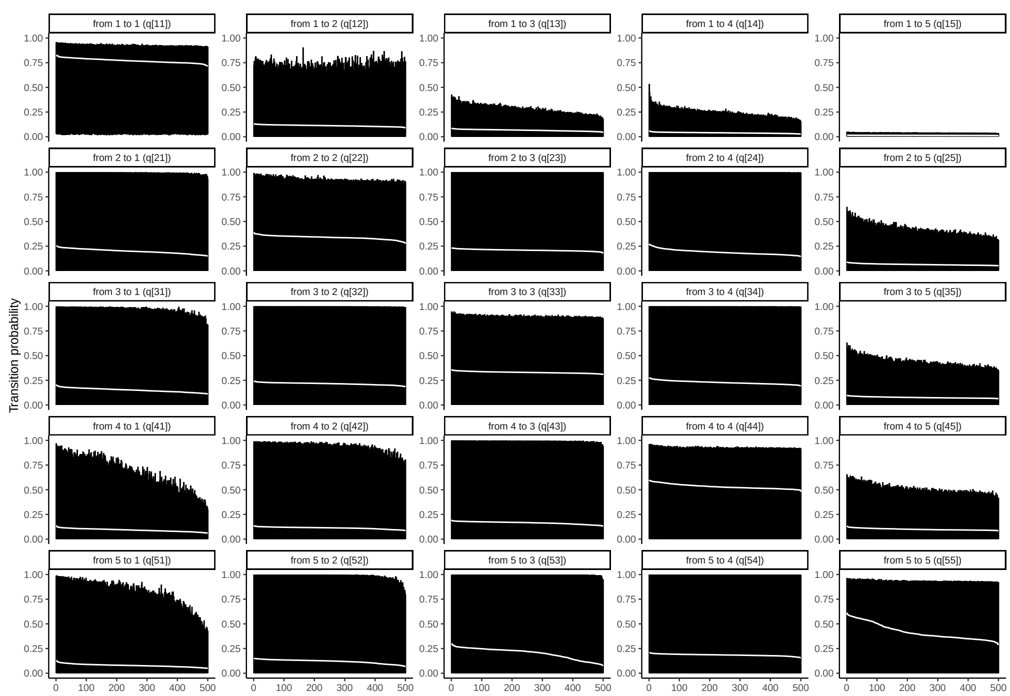
(a) Conditional probabilities for 5-states HMM without constraints in the transition matrix.

	1	2	3	4	5
p$_s$	0.130	0.421	0.556	0.720	0.917
	[0.093 0.173]	[0.147 0.590]	[0.328 0.869]	[0.520 0.931]	[0.865 0.982]

(b) Transition probabilities for 5-states HMM without constraints in the transition matrix.

		Population Mean				
		1	2	3	4	5
	1	0.766	0.113	0.067	0.040	0.014
		[0.119 0.852]	[0.001 0.556]	[0.000 0.201]	[0.000 0.184]	[0.002 0.038]
Q	**2**	0.190	0.328	0.209	0.204	0.069
		[0.005 0.650]	[0.005 0.804]	[0.002 0.719]	[0.002 0.616]	[0.005 0.248]
	3	0.142	0.209	0.332	0.237	0.080
		[0.003 0.456]	[0.002 0.724]	[0.004 0.809]	[0.003 0.726]	[0.007 0.298]
	4	0.084	0.107	0.157	0.547	0.105
		[0.004 0.315]	[0.002 0.458]	[0.003 0.594]	[0.020 0.847]	[0.011 0.343]
	5	0.073	0.107	0.165	0.185	0.470
		[0.005 0.313]	[0.003 0.454]	[0.004 0.481]	[0.004 0.610]	[0.013 0.837]

Figure C.1: Posterior mean and 95% confidence interval for individual transition probabilities. In each panel, customers are sorted in decreasing order of their posterior mean. The shaded area represents the 95% intervals, and the white line represents the posterior mean.



D Model Selection for hierarchical Bayesian Regression

Note that, beyond HMMs, model selection criteria may also favor a more complex model even if it only benefits 5% of data. We explore this issue in the context of a simple Bayesian Regression. Specifically, we simulated data using the following model $y_{it} = \beta_{oi} + \beta_{1i}x1_{it} + \beta_{2i}x2_{it} + \varepsilon_{it}$ with $\beta_{oi} = 1, \beta_{1i} = 1$, and $\beta_{2i} = 2$ for 5% of the customers and $\beta_{2i} = 0$ for 95% of the customers. $x1 \sim N(0, 1)$ and $x2 \sim N(0, 1)$.

We simulated 300 customers during 20 periods. We use the first 15 periods for calibration, and we leave the last 5 periods for validation purposes. We estimated two hierarchical Bayesian regression models, allowing heterogeneity in all regression coefficients. The first model (full model) estimated all parameters, whereas the second model (restricted model) estimated β_o and β_1. Restricting β_2 to be zero. The results in the table below show that both model selection criteria favor the more complex model (Full Model), even though in the DGP β_2 is effective for only 5% of the customers.

Table D.1: Model selection for Bayesian Regression

	Validation LL (higher is better)	Validation RMSE (lower is better)
Full Model	-2126.037	0.998
Restricted Model	-2217.692	1.060

This analysis confirms that the issue that model selection criteria tend to favor more complex models goes beyond HMMs as illustrated with relatively simple Bayesian regression models.

Note that in this regression analysis, the selection of the full model may not hurt the estimation procedure as it will identify that most customers are not sensitive to $x2$. This is not the case in HMMs because consumers who do not visit a third state have an unidentified transition from that state, and shrinkage causes these consumers to affect even those who truly have three states.

E Aggregate HMM tables reported in the literature

Figure E.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

F MHMM model specification

We detail the parameter transformations and priors for the general specification in Section 3.2.

F.1 HMM components

Consider the k'th HMM component with S_k states. We parametrize this component following the specification in Web Appendix A where we adapt the majority of the parameters to be component specific. That is, we rewrite equations (A.1) and (A.2) such that,

$$p_{isk} = \text{logit}^{-1}\left[\alpha_i + \delta_{sk}\right]$$
$$\delta_{sk} = \text{logit}\left[\sum_{\ell=1}^{s} u_{\ell k}\right],$$

with $\mathbf{u}_k$ a vector in the $(S_k - 1)-$dimensional simplex $\left(\mathbf{u} \in \mathbb{R}^{S_k},\ u_{sk} \geq 0,\ \sum_{s=1}^{S_k} u_{sk} = 1\right)$. Note that we hold α_i constant across components to help the model identify the states of different components across customers.

For the specification with time-varying covariates X_{it} in the state dependent probabilities, we include these covariates by

$$p_{iskt} = \text{logit}^{-1}\left[\alpha_i + \delta_{sk} + X_{it}' \cdot \boldsymbol{\beta}_{isk}\right].$$

We parametrize the transition probabilities for the HMM component with S_k states, $q_{iss'k}$, in Eq. (3) by

$$q_{iss'k} = P(z_{it+1} = s' | z_{it} = s, S_i = S_k) = \begin{cases} \frac{\exp(\gamma_{iss'k})}{\sum_{\ell=1}^{S_k-1} \exp(\gamma_{is\ell k})+1} & \text{if } s' \in \{1, \ldots, S_k - 1\} \\ \frac{1}{\sum_{\ell=1}^{S_k-1} \exp(\gamma_{is\ell k})+1} & \text{if } s' = S_k \end{cases} \tag{F.7}$$

For the specification with time-varying covariates X_{it} in the transition matrix (Sections 2.3 and 4.1), we use an ordered logistic specification where transition probabilities $q_{its\ell k}(X_{it}) = P(z_{it+1} = \ell | z_{it} = s, X_{it})$ are specified by:

$$q_{its1k} = \frac{\exp\left(\gamma_{is1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}{1 + \exp\left(\gamma_{is1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}$$
$$q_{its\ell k} = \frac{\exp\left(\gamma_{islk} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}{1 + \exp\left(\gamma_{is\ell k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)} - \frac{\exp\left(\gamma_{is\ell-1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}{1 + \exp\left(\gamma_{is\ell-1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)} \qquad \ell = 2, \ldots, S_k - 1$$
$$q_{itsS_kk} = 1 - \frac{\exp\left(\gamma_{isS_k-1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}{1 + \exp\left(\gamma_{isS_k-1k} - X_{it}' \cdot \boldsymbol{\rho}_{isk}\right)}$$

and $\gamma_{is1k} < \gamma_{is2k} < \ldots < \gamma_{isS_k-1k}$. We model such ordering directly in Stan using the native ordered vector parameters.

F.2 Parametric heterogeneity

We follow the exact specification of Web Appendix A.3, where each parameter is component specific, except for the α_i and its correspoonding variance parameter across customers σ_α.

F.3 Priors

For the parameters governing the HMM components, we follow identical priors to those outlined in Web Appendix A.4.

We put Dirichlet priors on the mixture component weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$,

$$\boldsymbol{\omega} \sim \text{Dirichlet}(a_1, \ldots, a_K).$$

We use values of $(a_1, \ldots, a_K)$ to enforce regularization on the model to favor HMM components with fewer states. We use values of a of 10%-11% of the number of customers ($a_1 = a_2 = a_3 = 50$ for our simulations and $a_1 = a_2 = a_3 = 33$ in our pharmaceutical application). For our gaming application, since there are only two components, we reparametrize $\omega = \text{logit}(\widetilde{\omega})$ and use Gaussian priors on $\widetilde{\omega}$ ($\widetilde{\omega} \sim \mathcal{N}(0, 1.8)$) which generate flat priors for ω.

G Details of Application 1: Pharmaceutical marketing

G.1 Selection of number of states and Gaussian mixture components

Table G.1: Model selection for the HMM

Number of states	Number of Mixture components	Validation	
		Val-LL	Val-RMSE
1	1	-1,749.160	2.101
	2	-1,757.173	2.293
	3	-1,770.010	2.316
	4	-1,762.955	2.254
2	1	-1,586.337	1.941
	2	-1,584.141	1.945
	3	-1,584.679	1.963
	4	-1,592.132	1.964
3	**1**	**-1,548.948**	**1.923**
	2	-1,600.697	2.063
	3	-1,563.950	1.947
	4	-1,579.325	1.944
4	1	-1,561.553	1.972
	2	-1,554.646	1.960
	3	-1,555.381	1.980
	4	-1,555.344	1.958

G.2 Model estimates

G.2.1 Raw estimates of the HMM

Table G.2: Raw parameter estimates for the HMM with 3 states. Each column corresponds to a state. For each parameter, the table shows the posterior mean with its corresponding 95% CPI in brackets underneath.

Parameter		States		
		L	M	H
Initial state				
Probabilities	π_s	0.83 $[0.73, 0.91]$	0.14 $[0.06, 0.24]$	0.03 $[0.00, 0.06]$
State dependent				
Intercept	δ_s	-7.86 $[-11.25, -5.88]$	-3.33 $[-3.63, -3.07]$	-1.93 $[-2.04, -1.82]$
	τ_α		0.65 $[0.58, 0.73]$	
Detailing	$\mu_{\beta,1s}$	0.36 $[-0.64, 1.99]$	0.01 $[-0.09, 0.12]$	0.01 $[-0.04, 0.05]$
	$\tau_{\beta,1s}$	0.38 $[0.07, 0.90]$	0.11 $[0.03, 0.25]$	0.11 $[0.05, 0.19]$
Sampling	$\mu_{\beta,2s}$	0.27 $[-0.84, 1.80]$	0.18 $[0.08, 0.29]$	0.03 $[-0.02, 0.08]$
	$\tau_{\beta,2s}$	0.40 $[0.04, 1.05]$	0.13 $[0.04, 0.26]$	0.16 $[0.09, 0.22]$
Transitions				
Threshold 1	$\mu_{\gamma,s1}$	0.33 $[0.03, 0.59]$	-3.23 $[-4.62, -2.40]$	-2.78 $[-3.44, -2.14]$
	$\tau_{\gamma,s1}$	0.18 $[0.02, 0.49]$	0.55 $[0.13, 1.19]$	1.85 $[1.38, 2.47]$
Threshold 2	$\mu_{\gamma,s2}$	0.75 $[0.32, 1.39]$	1.16 $[0.84, 1.52]$	0.88 $[0.46, 1.16]$
	$\tau_{\gamma,s2}$	0.41 $[0.02, 1.52]$	0.30 $[0.08, 0.56]$	0.14 $[0.02, 0.56]$
Detailing	$\mu_{\rho,1s}$	0.17 $[-0.12, 0.42]$	0.07 $[-0.24, 0.38]$	-0.02 $[-0.32, 0.23]$
	$\tau_{\rho,1s}$	0.10 $[0.02, 0.27]$	0.23 $[0.04, 0.70]$	0.30 $[0.04, 0.80]$
Sampling	$\mu_{\rho,2s}$	0.23 $[-0.01, 0.46]$	0.20 $[-0.10, 0.50]$	0.00 $[-0.24, 0.28]$
	$\tau_{\rho,2s}$	0.16 $[0.03, 0.43]$	0.20 $[0.04, 0.59]$	0.16 $[0.02, 0.45]$

G.2.2 Raw estimates of the MHMM

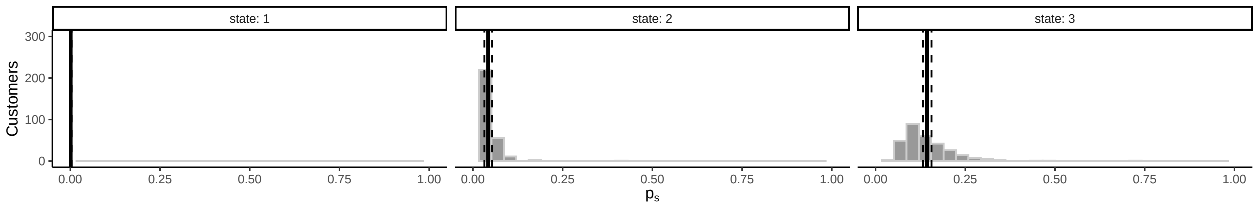
Table G.3: Raw parameter estimates for the MHMM. Each column corresponds to a state in a HMM component. For each parameter, the table shows the posterior mean with its corresponding 95% CPI in brackets underneath.

Parameter		HMM component (number of states)					
		1 state	2 states		3 states		
		M	L	H	L	M	H
Structural mixture							
Probabilities	ω_m	0.19	0.33		0.47		
		[0.16, 0.24]	[0.27, 0.40]		[0.41, 0.54]		
Initial state							
Probabilities	$\pi_{m,s}$	.	0.92	0.08	0.91	0.07	0.02
		.	[0.82, 0.99]	[0.01, 0.18]	[0.82, 0.98]	[0.01, 0.16]	[0.00, 0.06]
State dependent							
Intercept	$\delta_{m,s}$	-2.89	-7.43	-2.07	-8.09	-3.26	-1.81
		[−3.23, −2.54]	[−10.06, −5.95]	[−2.29, −1.81]	[−10.82, −6.37]	[−3.52, −2.94]	[−2.02, −1.52]
	$\tau_{\alpha,m}$	0.61	0.55		0.69		
		[0.37, 0.89]	[0.38, 0.78]		[0.54, 0.83]		
Detailing	$\mu_{\beta,1s}$	0.07	0.35	0.05	-0.52	-0.02	-0.03
		[−0.10, 0.27]	[−0.64, 1.54]	[−0.05, 0.15]	[−1.87, 0.93]	[−0.14, 0.11]	[−0.11, 0.05]
	$\tau_{\beta,1s}$	0.17	0.27	0.09	0.35	0.10	0.15
		[0.01, 0.38]	[0.04, 1.00]	[0.01, 0.27]	[0.09, 1.53]	[0.03, 0.27]	[0.02, 0.31]
Sampling	$\mu_{\beta,2s}$	0.03	0.39	-0.03	0.52	0.17	0.08
		[−0.11, 0.20]	[−1.76, 1.81]	[−0.12, 0.07]	[−1.00, 1.91]	[0.05, 0.30]	[0.00, 0.15]
	$\tau_{\beta,2s}$	0.18	0.75	0.08	0.48	0.12	0.18
		[0.06, 0.33]	[0.15, 1.57]	[0.01, 0.19]	[0.05, 1.53]	[0.01, 0.34]	[0.09, 0.28]
Transitions							
Threshold 1	$\mu_{\gamma,s1}$	.	0.37	-1.41	0.26	-3.08	-3.01
		.	[0.00, 0.74]	[−2.44, −0.57]	[−0.09, 0.60]	[−3.95, −2.05]	[−3.95, −2.21]
	$\tau_{\gamma,s1}$	.	0.18	2.06	0.12	0.35	1.41
		.	[0.04, 0.49]	[1.20, 3.40]	[0.02, 0.27]	[0.07, 0.75]	[0.60, 2.37]
Threshold 2	$\mu_{\gamma,s2}$	.	.	.	1.22	1.30	0.95
		.	.	.	[0.70, 1.86]	[0.89, 1.58]	[0.57, 1.30]
	$\tau_{\gamma,s2}$	.	.	.	0.22	0.18	0.27
		.	.	.	[0.03, 0.58]	[0.03, 0.47]	[0.06, 0.65]
Detailing	$\mu_{\rho,1s}$	.	0.26	0.04	0.07	0.06	-0.09
		.	[−0.05, 0.59]	[−0.50, 0.58]	[−0.30, 0.45]	[−0.26, 0.39]	[−0.56, 0.23]
	$\tau_{\rho,1s}$	.	0.16	0.31	0.22	0.11	0.19
		.	[0.03, 0.38]	[0.04, 0.81]	[0.01, 0.72]	[0.03, 0.28]	[0.03, 1.00]
Sampling	$\mu_{\rho,2s}$	.	0.07	-0.08	0.37	0.29	0.10
		.	[−0.19, 0.34]	[−0.54, 0.42]	[0.07, 0.70]	[−0.05, 0.60]	[−0.29, 0.44]
	$\tau_{\rho,2s}$	.	0.13	0.12	0.14	0.29	0.09
		.	[0.04, 0.43]	[0.02, 0.44]	[0.01, 0.54]	[0.06, 0.73]	[0.02, 0.21]

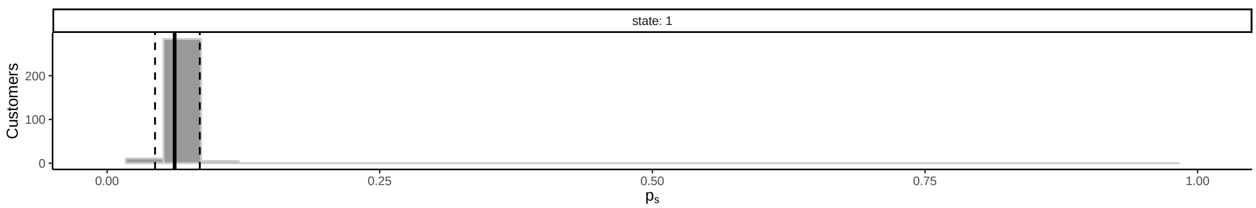
G.2.3 Heterogeneity in conditional prescription probabilities

Figure G.1: Histogram of individual level posterior mean of state dependent probabilities for the HMM and each component of the MHMM. All probabilities are computed with covariates at mean level. In solid black line is the population posterior mean, and in dashed lines the population mean 95% CPI.

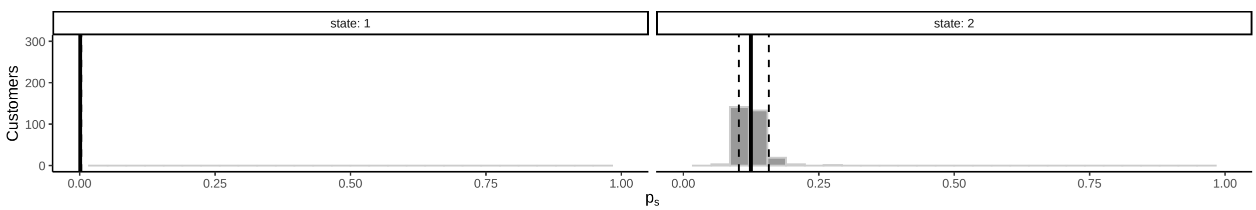
(a) HMM



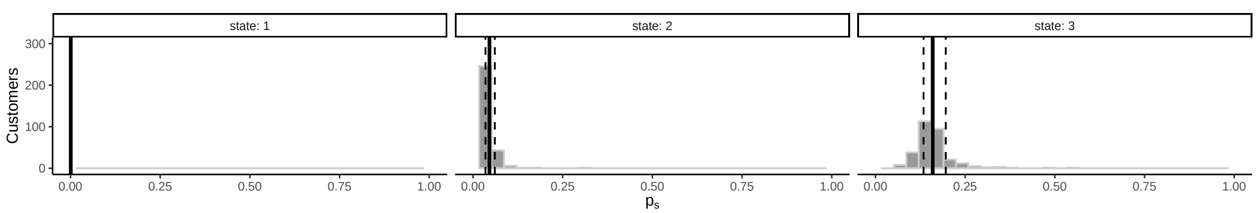
(b) MHMM: 1 state component



(c) MHMM: 2 states component



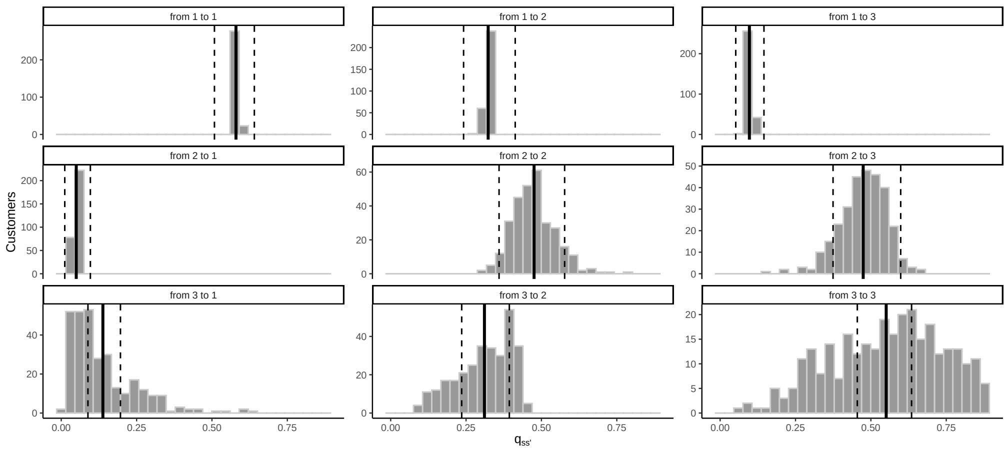
(d) MHMM: 3 states component



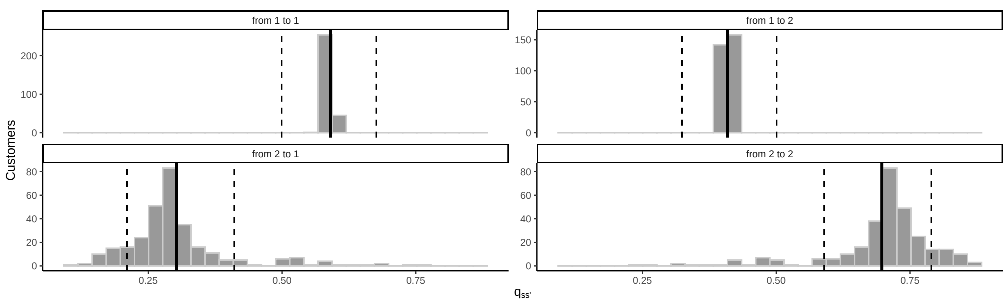
G.2.4 Heterogeneity in dynamics

Figure G.2: Histogram of individual level posterior mean of transition probabilities for the HMM and each component of the MHMM. All probabilities are computed with covariates at mean level. In solid black line is the population posterior mean, and in dashed lines the population mean 95% CPI.

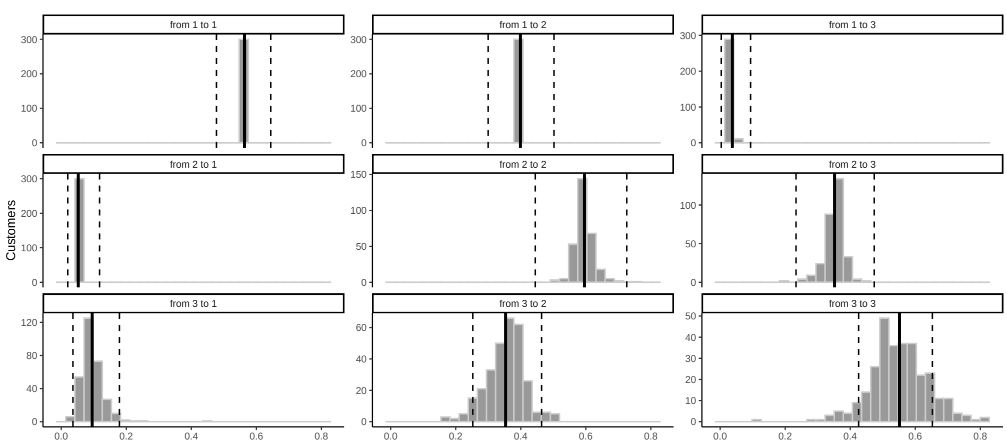
(a) HMM



(b) MHMM: 2 states component



(c) MHMM: 3 states component



G.3 Targeting

G.3.1 Targeted customers by HMM and MHMM

Table G.4: Targeted customers using HMM v. MHMM

Marketing action	Targeted (HMM)	Targeted (MHMM) No	Targeted (MHMM) Yes
Detailing only	No	192	33
	Yes	33	42
Sampling only	No	206	19
	Yes	19	56

Table G.5: Percentage of customers with 3-states by targeted group using HMM v. MHMM

Marketing action	Targeted group	Prop. 3-state customers HMM	Prop. 3-state customers MHMM
Detailing	No	59.1%	67.6%
	Yes	56.0%	30.7%
Sampling	No	51.6%	45.8%
	Yes	78.7%	96.0%

H Details of Application 2: Online gaming

H.1 Data

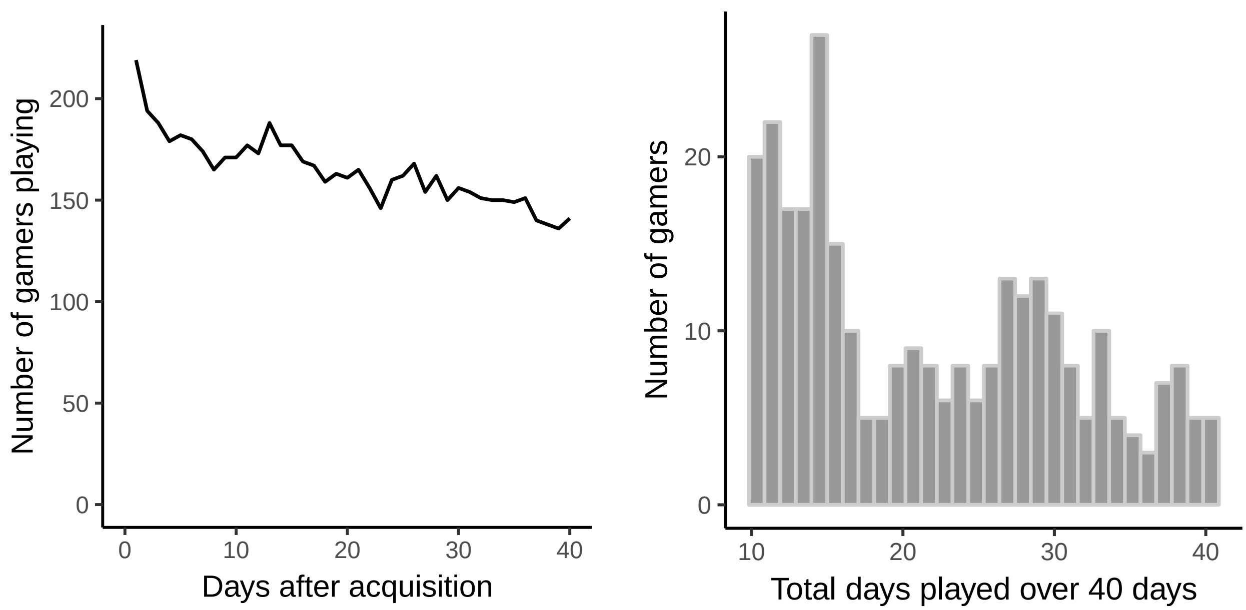
Our data comprise online game behavior in a role playing online game from a random sample of 300 gamers over a 60-day period after each gamer was acquired. Our full observation period is between April 1, 2008 and December 31, 2008. We observe when each player starts the game for the first time and follow them for 60 days. 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 represent 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 out-of-sample predictive ability of the HMM and the proposed MHMM. Consistent with common metrics in game user analysis (Huang et al., 2019), our aim is to predict whether a gamer plays on any given day.

Figure H.1 summarizes the playing behavior for the selected sample of 300 gamers. Figure H.1a shows the commonly observed overall decline in playing behavior over the course of the gamer's life post-acquisition. Figure H.1b 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 H.1: Summary of daily gaming behavior

(a) Evolution of playing behavior

(b) Heterogeneity in total days played



H.2 HMM and MHMM specifications

Following the description in Section 1.1.1 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 A.2).
- Conditional gaming behavior: We specify a Binomial distribution with $N = 1$. Specifically

$$P(Y_{it} = y_{it}|Z_{it} = s) = m_{it|s}(y_{it}) = \text{Binomial}(y_{it}|N = 1, p_{its}) \tag{H.8}$$

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 covariates that affect the gamer behavior when in 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 the "Mixture of HMMs (MHMM)" subsection 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.[5] The parameters for the HMM and the MHMM are estimated as described in the "Model estimation" subsection.

H.3 Results

H.3.1 Model selection for the HMM

We first determine the best specification for the HMM/RCM-HMM. To infer the number of states that best represents the data, we estimated the HMM for varying number of states (and mixture components for the RCM-HMM). Based on the expected log predictive density (Val-LL), we selected a 3-state HMM[6] (see Table H.1).

[5] We estimate an MHMM with up to three states as our recommended HMM has three states.

[6] Even though the criteria suggest different models, we favor the more parsimonious 3-state solution.

Table H.1: Selection of the number of hidden states and mixture components for the HMM and RCM-HMM.

Number of states	Number of Mixture components	Val-LL	Val-RMSE
1	1	-1828.26	23.735
1	2	-1826.55	23.892
1	3	-1832.33	23.872
1	4	-1837.52	23.743
2	1	-1424.01	10.983
2	2	-1434.19	10.793
2	3	-1425.93	10.822
2	4	-1430.66	10.791
3	1	**-1390.34**	9.300
3	2	-1421.88	10.352
3	3	-1425.52	10.458
3	4	-1416.66	9.101
4	1	-1390.93	9.591
4	2	-1394.28	**9.089**
4	3	-1392.17	9.666
4	4	-1402.28	9.450

Note: The best model in each column is in bold.

H.4 Parameter estimates for the 3-state HMM

In Table H.2 we report the population estimates of the 3-state HMM whereas in Figure H.2 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 H.2b). Note that the heterogeneity in the conditional behavior, p_s, when a gamer is in the second state is substantial (see also Figure H.2a). Second, the results for π_s in Table H.2a suggest that most gamers start in the high state. Finally, regarding gamers' dynamics, Table H.2c shows a large uncertainty in the population transition matrix, Q, particularly for state 2. In addition, Figure H.2b 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 H.2d 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 medium state but more likely to play if she is in the high state. In the low state, the effect is not statistically significant.

Table H.2: 3-state HMM parameter estimates

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

	1	**2**	**3**
π_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

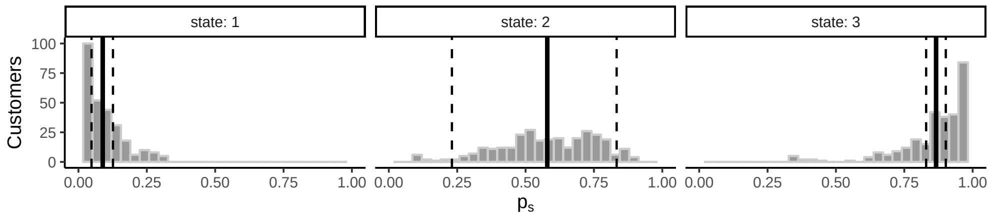
		1	**2**	**3**
Q	**1**	0.881	0.034	0.086
		[0.803 0.926]	[0.000 0.166]	[0.013 0.126]
	2	0.064	0.519	0.417
		[0.002 0.169]	[0.362 0.669]	[0.264 0.595]
	3	0.095	0.118	0.788
		[0.002 0.148]	[0.065 0.206]	[0.750 0.820]

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

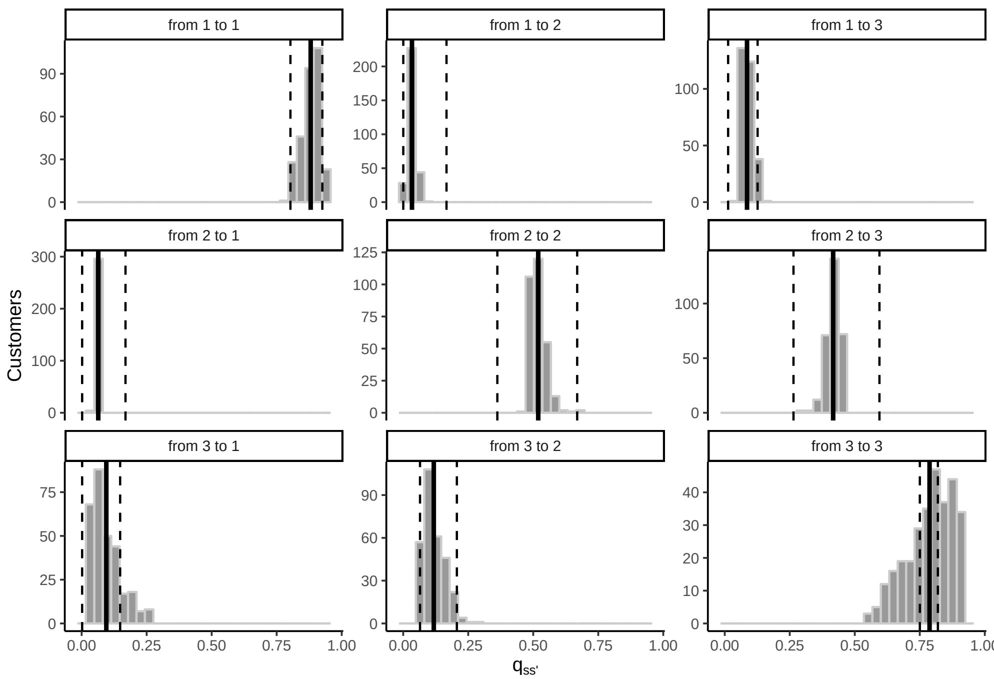
	1	**2**	**3**
$\mathbf{b_s}$	0.502	-4.529	3.501
	[−0.114 1.540]	[−9.320 − 2.200]	[2.065 7.506]

Figure H.2: 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



H.5 Parameter estimates for the MHMM

Table H.3 reports the parameter estimates for the proposed MHMM summarizing the posterior means and 95% confidence intervals for the population mean. Figure H.3 shows the histogram of the individual conditional probabilities and Figure H.4 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 H.3a 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 level of playing behavior overall with probability of playing, $p = 0.395$ (see Table H.3b).
2-state gamers. These dynamic gamers have two distinct states, characterized by a somewhat low level of playing behavior when they are in the low state, with a playing probability of $p_1 = 0.190$, and a very high level of playing behavior when they are in the high state, with a probability of playing of $p_2 = 0.969$ (see Table H.3d). All these gamers start in the high state (see Table H.3c) and stay in that state with high probability, $q_{22} = 0.913$. However, if they move to the low state, they also stay there with a relatively high probability, $q_{11} = 0.787$ (see Table H.3e).
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 of $p_1 = 0.025$, a somewhat 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 H.3g). Most of these gamers (60.8%) start in the high state (see Table H.3f) and stay in that state with a 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, although no gamer starts in the inactive state, when a gamer transitions to the inactive state she stays there with a 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 H.3 and H.4 confirm the characterization implied by the population estimates.

Thus, the dynamics in player behavior implied by the two approaches (HMM and MHMM) are quite different. The difference that probably stands 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 suggest that gamers are transitioning among non-playing, low-playing, or intensive-playing states, 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.

Table H.3: MHMM parameter estimates

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

	1 state	2 states	3 states
ω_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
π_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
π_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

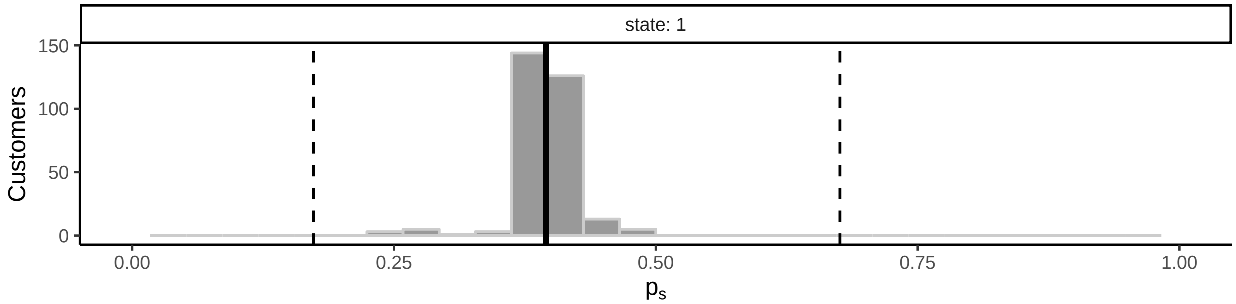
	1	2	3
$\mathbf{p_s}$	0.025	0.254	0.941
	[0.000 0.069]	[0.181 0.360]	[0.892 1.000]

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

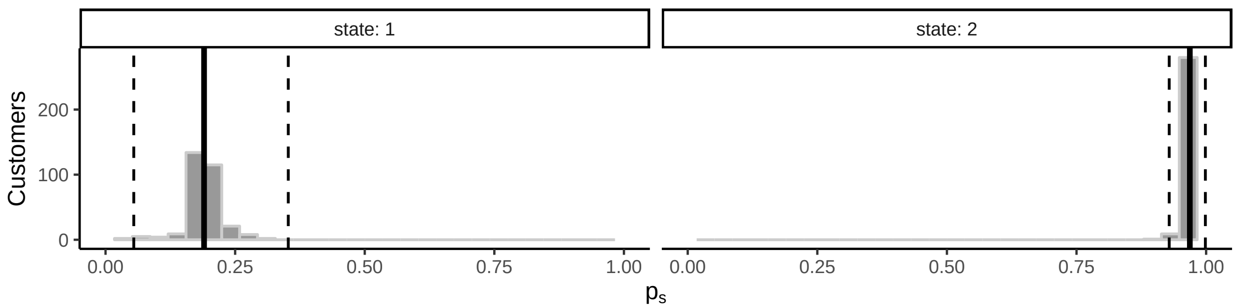
		1	2	3
Q	**1**	0.902	0.042	0.056
		[0.724 0.963]	[0.000 0.227]	[0.017 0.100]
	2	0.083	0.731	0.186
		[0.034 0.161]	[0.602 0.843]	[0.102 0.279]
	3	0.002	0.208	0.790

Figure H.3: 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

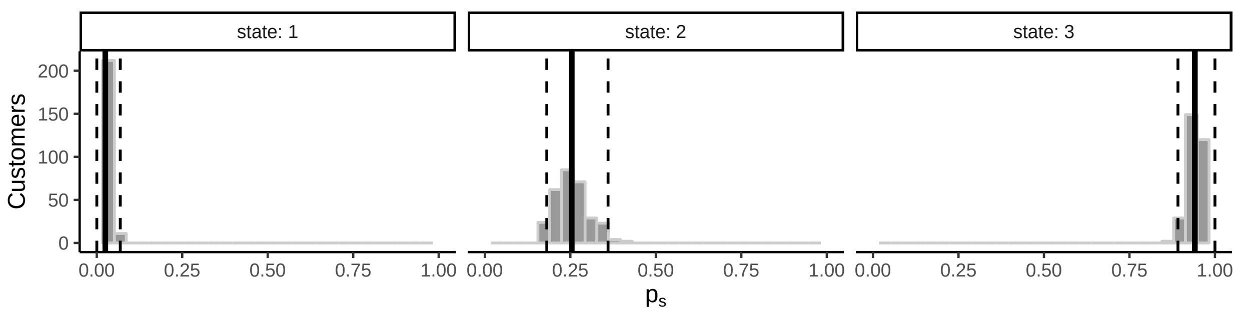
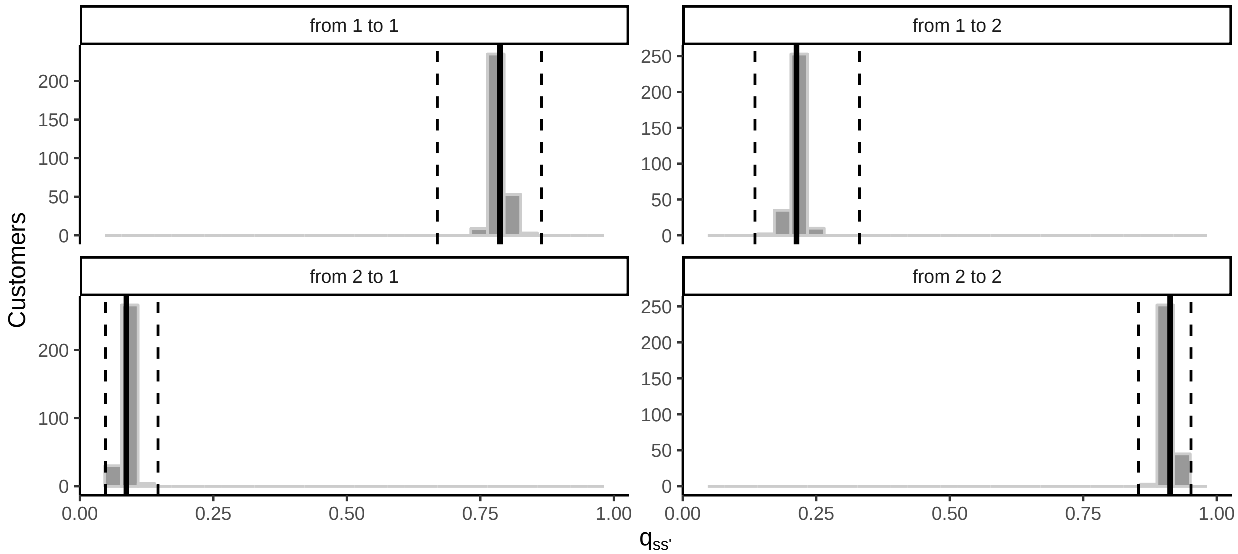


Figure H.4: Histogram of individual level posterior mean of transition probabilities for each component of the MHMM. In solid black line is the population posterior mean, and in dashed lines the population mean 95% CPI.

(a) 2 states component



(b) 3 states component