UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

HETEROGENEIDAD DE ESTADOS EN HIDDEN MARKOV MODELS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTÍON DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

NICOLÁS ALFONSO PADILLA PÉREZ

PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:
RICARDO MONTOYA MOREIRA
MARCEL GOIC FIGUEROA
RICHARD WEBER HAAS

Este trabajo ha sido parcialmente financiado por CONICYT

SANTIAGO DE CHILE
2014

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE: Ingeniero Civil Industrial y grado de
Magíster en Gestión de Operaciones
POR: Nicolás Padilla Pérez
FECHA: Abril 2014
PROFESOR GUÍA: Ricardo Montoya Moreira

HETEROGENEIDAD DE ESTADOS EN HIDDEN MARKOV MODELS

Hidden Markov models (HMM) han sido ampliamente usados para modelar comportamientos dinámicos tales como atención del consumidor, navegación en internet, relación con el cliente, elección de productos y prescripción de medicamentos por parte de los médicos. Usualmente, cuando se estima un HMM simultáneamente para todos los clientes, los parámetros del modelo son estimados asumiendo el mismo número de estados ocultos para cada cliente. Esta tesis busca estudiar la validez de este supuesto identificando si existe un potencial sesgo en la estimación cuando existe heterogeneidad en el número de estados.

Para estudiar el potencial sesgo se realiza un extenso ejercicio de simulación de Monte Carlo. En particular se estudia: a) si existe o no sesgo en la estimación de parámetros, b) qué factores aumentan o disminuyen el sesgo, y c) qué métodos pueden ser usados para estimar correctamente el modelo cuando existe heterogeneidad en el número de estados. En el ejercicio de simulación, se generan datos utilizando un HMM con dos estados para el 50% de clientes y un HMM con tres estados para el 50% restante. Luego, se utiliza un procedimiento MCMC jerárquico Bayesiano para estimar los parámetros de un HMM con igual número de estados para todos los clientes.

En cuanto a la existencia de sesgo, los resultados muestran que los parámetros a nivel individual son recuperados correctamente, sin embargo los parámetros a nivel agregado correspondientes a la distribución de heterogeneidad de los parametros individuales deben ser reportados cuidadosamente. Esta dificultad es generada por la mezcla de dos segmentos de clientes con distinto comportamiento.

En cuanto los factores que afectan el sesgo, los resultados muestran que: 1) cuando la proporción de clientes con dos estados aumenta, el sesgo de los resultados agregados también aumenta; 2) cuando se incorpora heterogeneidad en las probabilidades condicionales, se generan estados duplicados para los clientes con 2 estados y los estados no representan lo mismo para todos los clientes, incrementando el sesgo a nivel agregado; y 3) cuando el intercepto de las probabilidades condicionales es heterogéneo, incorporar variables exógenas puede ayudar a identificar los estados igualmente para todos los clientes.

Para reducir los problemas mencionados se proponen dos enfoques. Primero, usar una mezcla de Gaussianas como distribución a priori para capturar heterogeneidad multimodal, y segundo usar un modelo de clase latente con HMMs de distintos número de estados para cada clase. El primer modelo ayuda en representar de mejor forma los resultados agregados. Sin embargo, el modelo no evita que existan estados duplicados para los clientes con menos estados. El segundo modelo captura la heterogeneidad en el número de estados, identificando correctamente el comportamiento a nivel agregado y evitando estados duplicados para clientes con dos estados.

Finalmente, esta tesis muestra que en la mayoría de los casos estudiados, el supuesto de un número fijo de estados no genera sesgo a nivel individual cuando se incorpora heterogeneidad. Esto ayuda a mejorar la estimación, sin embargo se deben tomar precauciones al realizar conclusiones usando los resultados agregados.

ABSTRACT

HETEROGENEITY IN THE NUMBER OF STATES OF HIDDEN MARKOV MODELS

Hidden Markov Models (HMM) have been widely used in marketing to model dynamic consumer behavior such as consumer attention, web search behavior, customer relationships, choice selection and medical prescription behavior. When estimating simultaneously the same HMM for all customers, researchers usually assume a common number of hidden states for all customers. In this thesis, we analyze the potential bias of such assumption when there is heterogeneity on the number of states across customers.

To analyze the potential bias, we perform a comprehensive Monte Carlo simulation exercise. Specifically, we study: a) the potential bias on parameter estimates; b) under which conditions the bias increases or decreases; and c) which methods can be used to estimate correctly the model when there is heterogeneity on the number of states. In the simulation exercise, we generate data using a HMM with two states for 50% of customers and a HMM with 3 states for the other 50% of customers. Next, we use a hierarchical Bayesian MCMC procedure to estimate the parameters of a HMM with the same number of states for all customers.

First, the results show that parameters are correctly recovered at individual level. However, aggregate parameters from heterogeneity distribution have to be reported with precautions. This issue is generated by averaging the mixture of two segments of customers with different behavior.

Second, regarding the factors that affect the bias, we show that: 1) when the proportion of customers with two states increases, bias on aggregate results increases as well; 2) when heterogeneity in conditional probabilities is introduced in the model, duplicated states are estimated for customers with 2 states, and states are not identified as the same for all customers, which increases the bias on aggregate results; and 3) when the intercept of conditional probabilities is heterogeneous, introducing covariates to the model helps in identifying states across customers.

Third, we proposed two models to account for heterogeneity in the number of states: 1) using a mixture of Gaussians as prior distributions of individual level parameters, to capture multimodal heterogeneity; and 2) a latent class model with a HMM of different number of states on each class. The first model provides a better interpretation of aggregate results. However, this model does not avoid estimating duplicated states for customers with two states. The second model captures the heterogeneity on the number of states, identifying correctly the behavior at the aggregate level and avoiding the estimation of duplicated states for customers with two states.

Finally, in this thesis we show that in most cases, the assumption of a common number of states does not generate a bias at the individual level, given that accounting for heterogeneity improves the estimation results. However, conclusions from aggregate level results have to be made with precautions.

A mi futura compañera de la vida, mi familia y mi amigos...

Agradecimientos

Quisiera reflejar en estas palabras lo agradecido que estoy de cada una de las personas que me ayudaron de alguna u otra manera dentro de mi proceso académico y personal. Probablemente se me olvide más de alguno, pero eso no significa que no haya sido parte importante de mi vida universitaria.

Primero que todo quiero agradecer a Jacqueline, que ha sido mi fiel compañía y apoyo los ultimos 5 años. Sin su cariño y soporte incondicional, mi vida personal y académica hubiese sido mucho más difícil. Muchas gracias por el amor que me has entregado estos 5 años, y por acompañarme en esta aventura en Nueva York.

A mis padres Luis y Begoña, que me apoyaron en todas mis decisiones a lo largo de la carrera, y me alentaron a tomar desafíos difíciles y fuera de lo tradicional, para que pudiera encontrar el camino para hacer lo que me gusta. Muchas gracias por el apoyo y cariño durante todo lo que recuerdo de vida. A mis hermanos Felipe y Begoña, que siempre me alentaron durante la carrera y fueron complices de todas nuestras maldades durante nuestra infancia. Gracias también a la Yaya, mi abuela, que siempre se interesó por preguntarme cómo estaba y felicitarme por los logros que iba alcanzando en la vida.

A mi profesor guía, Ricardo Montoya, quién no solo fue un guía académico, sino también personal. Muchas gracias por toda la ayuda durante esta tesis, y principalmente por el apoyo y guía que recibí de su parte para proseguir mi carrera académica en Estados Unidos. Sus consejos no sólo me ayudaron a desarrollar un mucho mejor trabajo, sino que me permitieron darme cuenta de querer seguir una vida académica.

A mis amigos David y Francisco, que siempre, acompañados de nuestras tradicionales cervezas, me apoyaron y me aconsejaron durante estos 8 años. A mis amigos de la universidad, Bárbara, Miguel, Pablo y Óscar, por sus bromas respecto a mi eterna vida universitaria; a mis amigos de plan común: Gabriel, Gaspar, Sebastián, Yerko, Nicolás y Bruno; a mis amigos(as) del CERET y del Centro de Finanzas; y a mi cuñado Pablo con quién al fin celebramos la Décima.

Muchas gracias a todos por ser parte de mi vida durante esta etapa.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Hidden Markov models (HMM) have been successfully applied to model the changes in consumer behavior across different and unobserved states. HMM allows to characterize an underlying process generating an observable sequence. The underlying process is modeled using a Markov chain, and the observable sequence distribution is conditional on the state of the Markov chain.

In marketing, Bayesian models as well as latent class models have been used to account for unobserved customer heterogeneity whereas hidden Markov models have been used to account for dynamics.

HMM applications in marketing include customer attention [Liechty et al., 2003], web search behavior [Montgomery et al., 2004], customer relationship [Netzer et al., 2008], response to marketing activities [Montoya et al., 2010], learning in behavioral games [Ansari et al., 2012], and churn and usage [Ascarza and Hardie, 2013].

One key point that has been neglected in the literature is the heterogeneity in the dynamic structure. That is, researchers assume that customer transitions take place among a fixed unique number of states, i.e., all customers have same number of states in the Markov structure. Criteria such as BIC, AIC or DIC have been used to identify the appropriate structure in consumer dynamics but always assuming a common number of states for all consumers. This simplification helps the parameter estimation procedure and their convergence. In contrast, it can lead to spurious parameter estimation, biasing the parameter estimates and the implied results.

In this thesis, we address the following questions: (1) Is there a bias when assuming the same number of HMM states for all customers? (2) What factors increase/decrease the bias? (3) What alternative methods can be applied to estimate a HMM with heterogeneity in the number of states?

1.2 Main objective

The main objective of this thesis is to study the effects of assuming homogeneity in the number of states in a HMM.

1.3 Specific objectives

The specific objectives of this work are to:

1. Characterize the bias on the estimation of the parameters of a HMM when the data have heterogeneity in the number of states.
2. Determine the factors that improve or deteriorate the estimation results when the data have heterogeneity in the number of states.
3. Develop a method that estimates a HMM capturing heterogeneity in the number of states.

1.4 Structure of this thesis

This work is organized as follows. In Chapter 2 we describe the theoretical framework used in this thesis. We also provide a literature review of HMM applications in marketing and current approaches to select the number of states. In Chapter 3 we describe the general model used in this thesis. In Chapter 4 we describe the simulation experiments we perform to analyze the potential bias on the estimation when the data have heterogeneity in the number of states. In Chapter 5, we develop two models to incorporate the heterogeneity not captured by the standard model: a HMM with a mixture of Gaussians as prior distribution and a latent class of hidden Markov models. In Chapter 6 we apply the proposed models to empirical data. Finally, in Chapter 7 we summarize the conclusions of this work and give further research directions.

Chapter 2

Theoretical Framework

2.1 Hidden Markov Models

2.1.1 Motivation

A Hidden Markov Model (HMM) allows to study the dynamic behavior of systems, in which, unlike the traditional Markov Chains, the state of the system at each point of time is unknown. Instead, random variables, with distributions that depend on the unobserved hidden state, can be observed. HMM was first introduced in a series of papers by Baum and his colleagues in the late 1960s, and it was treated as probabilistic functions of a Markov chain before the name Hidden Markov Model became popular.

The most known applications of HMM are in speech recognition [Rabiner, 1989], cryptanalysis [Green et al., 2005] and genetics [Eddy, 1998]. Also there is a growing use of HMM in marketing for its flexibility to model dynamics and underlying motivations for customer behavior.

Figure 2.1 shows a HMM with states $\{x_t\}_{t=1}^{\infty}$ and observable random variables $\{y_t\}_{t=1}^{\infty}$. x_t are hidden states that are not necessary something in concrete, i.e., not necessary measurable values. Examples of hidden states are degrees of customer relationships [Netzer et al., 2008], goals in online searching behavior [Montgomery et al., 2004], willingness to prescribe pharmaceutical drugs [Montoya et al., 2010], among others. y_t, in contrast, are observable data, measurable values such as purchases, visits to a website, usage of a service, among others.

The essence of HMM is that the hidden state affects how the data are generated. Formally, the distribution of the observable variables depends on the hidden state. In addition, the system evolves dynamically between different states, and is modeled using a Markov chain.

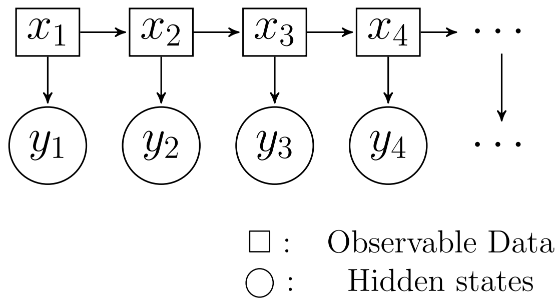


Figure 2.1: Discrete-time discrete-space HMM

2.1.2 Model specification

As the hidden state follows a Markov chain, a HMM consists of three components: initial probabilities, transition probabilities and the conditional distribution of the observable variable.

Let S a set of all possible hidden states, $\{X_t\}_{t=1}^{\infty} \in S$ the hidden state variable and $\{Y_t\}_{t=1}^{\infty}$ the observable variable. To identify a HMM, these three components must be estimated.

1. Initial probabilities

$$\pi_s = \mathbb{P}\left(X_1 = s\right)$$

   where $\pi \in \mathbb{R}^{|S|}$ such as $\pi_s \geq 0$ and $\sum_{s \in S} \pi_s = 1$.

2. Transition probabilities

$$q_{ss't} = \mathbb{P}\left(X_t = s' | X_{t-1} = s\right)$$

   where $q_{ss't} \in \mathbb{R}$ such as $q_{ss't} \geq 0$ and $\sum_{s' \in S} q_{ss't} = 1$.

3. Conditional probabilities

$$f_{st}\left(y\right) = \begin{cases} \mathbb{P}\left(Y_t = y | X_t = s\right) & \text{if } Y_t \text{ is discrete} \\ p\left(y | X_t = s\right) & \text{if } Y_t \text{ is continuous} \end{cases}$$

   where $p\left(y|X_t = s\right)$ is the conditional density of Y_t if Y_t is continuous.

Estimating $\{\pi_s, q_{ss't}, f_{st}\left(y\right)\}$ is necessary to identify the model but usually $\{\pi_s, q_{ss't}, f_{st}\left(y\right)\}$ is parameterized in a vector of parameters.

2.1.3 HMM applications in marketing

The flexibility of HMM to define what a hidden state is for each problem has allowed researchers in marketing to model several applications of customer behavior. In most cases, it is not needed to specify what the hidden state is, but instead, to understand that those states exist, and they influence the decisions that customers make. For example, Netzer et al. [2008] define the hidden state as the strength of the relationship between customer and firm. Most HMM applications consider a set of customers with an observable behavior, such as customer expenditure [Kumar et al., 2011; Ascarza and Hardie, 2013], prescriptions [Montoya et al., 2010], donations [Netzer et al., 2008], online browsing logs [Montgomery et al., 2004], among others.

Exogenous effects can also be incorporated in the model, such as marketing actions. Models defined by Netzer et al. [2008] and Montoya et al. [2010] incorporate exogenous factors (covariates) that influence both transitions between states and the conditional behavior. The transition matrices are modeled using an ordered multinomial logit model that allows for covariate effects [Greene, 1997]. For example, in a physician' prescription context, the model described by Montoya et al. [2010] captures not only short term effects of marketing actions (detailing and sampling) but also the long term effect, by introducing those covariates on the transition matrix of the hidden Markov model.

Multiple observed variables can be modeled using a single HMM as well. Ascarza and Hardie [2013] model jointly the usage and churn of customers in contractual settings (e.g. a warehouse with membership needed to purchase goods). In this work, the observable variable models the usage of the service (e.q. purchases of each customer) and the hidden state is the "commitment" of the customer to the firm. Then, every four periods if the customer is in the lower state of "commitment" then the customer quits the membership. Customer usage is modeled using a Poisson process with parameter λ depending on the hidden state and customer. The hidden state is modeled explicitly using a Multinomial-Dirichlet model[1], where the multinomial process models the state dynamics, and the Dirichlet distribution models the rows of the transition matrix.

Endogeneity in marketing actions can also be introduced in a HMM. Kumar et al. [2011] describe a HMM with endogeneity to model the expenditures of customers in a B2B context with marketing actions, and uses it to suggest that higher states are not stable, i.e., customers in higher state do not remain in that state as long as customers in other states. Kumar et al. [2011] models the marketing actions as function of marketing actions and revenues of previous periods to account for non-random allocation of marketing efforts.

Usually the conditional relationship on the observable variable distribution is parametrical, i.e., the functional form of the conditional distribution is the same for all states, but parameters are state dependent. However, different structures for the observable distribution can me modeled as well. Ansari et al. [2012] uses a HMM to model the strategies in behavioral games. The attraction to a specific strategy is computed either using a Reinforcement Learning model or a Belief Learning model depending on the hidden state.

Most of the HMM applications consider a discrete time Markov chain. Nevertheless, Montgomery et al. [2004] uses a continuous time HMM to model the behavior in online browsing on an e-commerce web site. The observable data are the path of pages that each customer follows on the site. On the other hand, the hidden state is the "goal" of the customer when he is browsing in the site (browse orientation or purchase orientation). The time each customer stays on a specific orientation state is modeled as continuous and follows a exponential distribution.

[1] Dirichlet distribution is a conjugate prior of the multinomial process [Rossi et al., 2005].

2.1.4 Selecting the number of states

There is not a single correct method of selecting the number of states of a HMM. The number of states is a parameter that changes the dimension of the parameters of the HMM, therefore selecting the number of states is often seen as a model selection procedure as mentioned in Scott [2002] and Netzer et al. [2008] among others.

Penalized likelihood criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) use log-likelihood function penalizing with the number of parameters for AIC and both number of parameters and observations for BIC [McLachlan and Peel, 2004]. Netzer et al. [2008] also uses a Markov Switching Criterion (MSC) which is an adapted version of the Kullback-Leibler divergence described in Smith et al. [2006] to penalize for higher number of states.

Hypothesis test approaches such as Likelihood Ratio Test uses a number of states n as null hypothesis (H_0) versus a number of states m, usually $m = n + 1$ (H_1), and moving forward until do not reject H_0 [McLachlan and Peel, 2004].

Scott [2002] proposed a posterior probability of the number of states computed using the results of a Monte Carlo simulation. Given a set of possible number of states $\{1, \ldots, M\}$, and L draws of parameter θ^n of each HMM with number of states n ($\Theta = \{\theta^n\}_{n=1}^{M}$), the posterior distribution of the number of states can be computed as follows:

$$\begin{aligned}
\mathbb{P}(n|Y) &= \int \mathbb{P}(n|Y)\, p(\Theta|Y)\, d\Theta \\
&\approx \frac{1}{L}\sum_{l=1}^{L} \mathbb{P}(n|Y, \Theta_l) \\
&\approx \frac{1}{L}\sum_{l=1}^{L} \left[\frac{p(Y|n, \theta_l^n)\, \mathbb{P}(n)}{\sum_{m=1}^{M} p\left(Y|m, \theta_l^m\right) \mathbb{P}(m)} \right]
\end{aligned} \tag{2.1}$$

Embedded algorithms also have been applied to estimate both number of states and parameters of HMM simultaneously. Robert et al. [2002] uses a reversible jump Markov chain Monte Carlo (RJMCMC) to estimate the number of states and parameters simultaneously. RJMCMC was introduced by Green [1995] as a Bayesian model selection procedure that estimates the parameters as well as the model that fits best the data. RJMCMC is a complex method that changes the dimension space of parameters between draws. Robert et al. [2002] uses birth, death, split and join moves to increase and decrease the number of states.

The papers cited before use a fixed number of states for the single HMM or for all HMMs estimated, and use the procedures we described before to estimate that number. When individual level HMMs are estimated, (Netzer et al. [2008] defines a HMM for each customer, and Montoya et al. [2010] defines a HMM for each physician) the number of states is usually common across customers. Accounting for heterogeneity in the number of states of HMM has not been clearly treated. Gunter and Bunke [2003] use a method with heterogeneity on the number of states in the context of handwriting recognition, but use an ad hoc procedures where the number of states is flexible of each HMM and correspond to the average length of the corresponding sequence of feature vectors times a constant f. This method is specially designed for the handwriting recognition problem.

Those papers do not use a more general framework for testing for presence of heterogeneity on the number of states in their data, or methods to capture it.

Chapter 3

Model Specification

3.1 Heterogeneous Hidden Markov Model

The specification of the hidden Markov model used in this thesis is based on the work of Netzer et al. [2008] and Montoya et al. [2010]. We use three subscripts in this model: i denotes an individual customer ($i = 1, ..., K$), t denotes a time period $t = 1, ..., T$ and s denotes a hidden state ($s = 1, ..., S$). Let Y_{it} be the observed behavior for customer i at period t, $z_{it} \in S$ the state of customer i at period t, and X_{it} the vector of C covariates for customer i at period t.

As shown in Section 2.1.2 a HMM includes three components: (1) initial probabilities (π_i), (2) transition probabilities (Q_{it}) and (3) conditional probabilities of the observed purchase behavior (M_{it}).

3.1.1 Initial probabilities

Let $h_{is} = \mathbb{P}(z_{i1} = s)$ be the probability that customer i is on state s at period 1, and

$$\Pi_i = [h_{i1} \dots h_{iS}] \tag{3.1}$$

the initial hidden state probability of customer i. These probabilities can be incorporated as a parameter of the model and be estimated with the rest of the parameters. However, in this work, we consider Π_i as fixed values that are not estimated.

3.1.2 Transition matrices

The transition matrix Q_{it} is modeled at the individual level as a function of customer level parameters and covariates. Individual level parameters are incorporated to estimate the intrinsic propensity of a given customer to transition from one state to another.

Let $q_{iss't}$ be the probability that customer i switches from state s at period t to state s' at period $t+1$, i.e.:

$$q_{iss't} = \mathbb{P}\left(z_{it+1} = s' | z_{it} = s, X_{it}\right) \tag{3.2}$$

where $q_{iss't} \geq 0$ and $\sum\limits_{s' \in S} q_{iss't} = 1$.

Equation 3.2 implies that covariates at period t, X_{it}, affect transitions from t to $t+1$. Accordingly, transition matrices Q_{it} can be written as:

$$Q_{it} = \begin{bmatrix} q_{i11t} & q_{i12t} & \cdots & q_{i1St} \\ \vdots & \ddots & & \vdots \\ q_{iS1t} & q_{in2t} & \cdots & q_{iSSt} \end{bmatrix} \tag{3.3}$$

To parametrize the transition probabilities and incorporate covariate effects in the transition matrices, we use the ordered *logit* model [Netzer et al., 2008; Montoya et al., 2010]. To compute the s'th row of Q_{it}, let consider a latent variable u_{it} that accounts for the propensity of moving from state s. We define $u_{it} = \boldsymbol{\rho}_{is}^T X_{it} + \epsilon_{it}$, where $\boldsymbol{\rho}_{is}$ is the vector of marketing effect parameters on the transition matrix and ϵ is an error term i.i.d., with extreme value distribution. u_{it} controls the customer propensity to move to higher states. Let $\{\hat{\tau}_{iss'}\}_{s' \in \{1 \ldots S-1\}}$ be the thresholds parameters for state s. If $\hat{\tau}_{iss'} \leq u_{it} \leq \hat{\tau}_{iss+1'}$ then customer i moves from state s to state s'.

The transition probabilities $q_{iss't}$ can be written as:

$$\begin{aligned}
q_{is1t} &= \frac{\exp\left(\hat{\tau}_{is1} - \boldsymbol{\rho}_{is}^T X_{it}\right)}{1 + \exp\left(\hat{\tau}_{is1} - \boldsymbol{\rho}_{is}^T X_{it}\right)}, \\
q_{iss't} &= \frac{\exp\left(\hat{\tau}_{kss'} - \boldsymbol{\rho}_{ks}^T X_{kt}\right)}{1 + \exp\left(\hat{\tau}_{iss'} - \boldsymbol{\rho}_{is}^T X_{it}\right)} - \frac{\exp\left(\hat{\tau}_{iss'-1} - \boldsymbol{\rho}_{is}^T X_{it}\right)}{1 + \exp\left(\hat{\tau}_{iss'-1} - \boldsymbol{\rho}_{is}^T X_{it}\right)}, \qquad & s' = 2, \ldots, S-1 \\
q_{isSt} &= 1 - \frac{\exp\left(\hat{\tau}_{isS-1} - \boldsymbol{\rho}_{is}^T X_{it}\right)}{1 + \exp\left(\hat{\tau}_{isS-1} - \boldsymbol{\rho}_{is}^T X_{it}\right)},
\end{aligned} \tag{3.4}$$

To ensure a proper ordering of the states, we impose an increasing ordered parametrization of the threshold parameters as described in Equations 3.5, using unbounded parameters $\tau_{iss'}$.

$$\begin{aligned}
\hat{\tau}_{is1} &= \tau_{is1} & s &= 1, \ldots, S \\
\hat{\tau}_{iss'} &= \hat{\tau}_{iss'-1} + \exp\left(\tau_{iss'}\right) & s &= 1, \ldots, S;\ s' = 2 \ldots S-1
\end{aligned} \tag{3.5}$$

These equations imply that $\hat{\tau}_{is1} < \hat{\tau}_{is2} < \ldots < \hat{\tau}_{isS-1}$.

3.1.3 Conditional probabilities

The conditional probabilities matrix M_{it} capture the distribution of the observed behavior Y_{it}.

Let m_{ist} be the probability of the observed behavior given that customer i is on state s at period t, as:

$$m_{ist} = \mathbb{P}\left(Y_{it} = y_{it} | z_{it} = s\right) \tag{3.6}$$

Following Montoya et al. [2010], we model the conditional distribution of Y_{it} as a Binomial distribution, with parameters N_{it} as the number of Bernoulli variables, and p_{ist} as the Bernoulli's probability of success. Parameters p_{ist} imply that this distribution is conditional on the hidden state s. In a marketing context, Y_{it} may represent the purchase behavior of a customer i at period t, N_{it} represents the total category purchases made by customer i at period t and p_{ist} represents the market share of the product for that customer at that period. Formally, this can be written by:

$$\begin{aligned} m_{ist} &= \mathbb{P}\left(Y_{it} = y_{it} | z_{it} = s\right) \\ &= \binom{N_{it}}{y_{it}} p_{ist}^{y_{it}} \left(1 - p_{ist}\right)^{N_{it} - y_{it}} \end{aligned} \tag{3.7}$$

Probability p_{ist} is modeled using a *logistic* regression model with intercept $\hat{\alpha}_s^0$ and vector $\boldsymbol{\alpha}_{is}$ which captures the effects of covariates X_{it}.

$$p_{ist} = \frac{\exp\left(\hat{\alpha}_s^0 + \boldsymbol{\alpha}_{is}^T X_{it}\right)}{1 + \exp\left(\hat{\alpha}_s^0 + \boldsymbol{\alpha}_{is}^T X_{it}\right)} \tag{3.8}$$

To avoid the label switching problem on hidden states we impose an increasing ordered parametrization of $\hat{\alpha}_s^0$, using unbounded parameters α_s^0.

$$\begin{aligned} &\hat{\alpha}_1^0 = \alpha_1^0 \\ &\hat{\alpha}_s^0 = \hat{\alpha}_{s-1}^0 + \exp\left(\alpha_s^0\right) \qquad\qquad s = 2, \ldots, S \end{aligned} \tag{3.9}$$

These equations imply that $\hat{\alpha}_1^0 < \hat{\alpha}_2^0 < \ldots < \hat{\alpha}_n^0$.

Finally, conditional state probabilities M_{it} are written using the notation for HMM models [MacDonald and Zucchini, 1997] as a diagonal matrix with m_{ist} as the s'th diagonal component.

$$M_{it} = \begin{bmatrix} m_{i1t} & 0 & \dots & \dots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & m_{ist} & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \dots & \dots & 0 & m_{iSt} \end{bmatrix} \tag{3.10}$$

3.1.4 Parameters of the model

In sum, the parameters of the three components of the HMM model can be grouped in two sets: individual level parameters and population parameters. Let θ_i be the individual level parameters of customer i and Φ the population parameters of the model. Then:

$$\theta_i = \left\{\tau_{is1}, \ldots, \tau_{isS-1}, \boldsymbol{\rho}_{is}, \boldsymbol{\alpha}_{is}\right\}_{s=1}^S \tag{3.11}$$

$$\Phi = \left\{\alpha_s^0\right\}_{s=1}^S \tag{3.12}$$

We define n_θ as the length of vectors θ_i and n_Φ as the length of vector Φ.

3.1.5 Likelihood function

Using the set of parameters $\left\{\{\theta_i\}_{i\in K}, \Phi\right\}$, and T observations for each customer ($Y = \left\{\{Y_{it}\}_{t\in T}\right\}_{i\in K}$) , the likelihood function for individual i can be written as [MacDonald and Zucchini, 1997]:

$$
\begin{aligned}
L_i\left(\theta_i, \Phi | \{Y_{it}\}_{t\in T}\right) &= \mathbb{P}\left(Y_{i1}, Y_{i2}, \ldots, Y_{iT} | \theta_i, \Phi\right) \\
&= \Pi_i M_{i1} \prod_{t\in T} Q_{it} M_{it} \mathbf{1}^T
\end{aligned}
\tag{3.13}
$$

where $\mathbf{1}$ is a $1 \times S$ vector of ones.

Using this equation, likelihood function for all customers can be written as:

$$
\begin{aligned}
L\left(\{\theta_i\}_{i\in K}, \Phi | Y\right) &= \prod_{i\in K} L_i\left(\theta_i, \Phi | \{Y_{it}\}_{t\in T}\right) \\
&= \prod_{i\in K} \Pi_i M_{i1} \prod_{t\in T} Q_{it} M_{it} \mathbf{1}^T
\end{aligned}
\tag{3.14}
$$

3.2 Variations of the model

Several simplifications and variations of this general model are used throughout this thesis. This section aims to explain the differences among these modified versions and the general form of the HMM described in Section 3.1.

3.2.1 Covariates

The presence of covariates in both the transition matrix (Q_{it}) and the conditional behavior (M_{it}) affects the parametrization of both variables. Covariates are incorporated in the transition matrix using an ordered *logit* parametrization to ensure that the effect of a covariate either increases the propensity to transition to higher states or increases the propensity to transition to lower states. When covariates are not included in the model, the transition matrix is constant for all periods, i.e., $Q_{it} = Q_i \, \forall t \in T$. In such a case, we parametrize the rows of Q_i using a multinomial *logit* model with unconstrained parameters $\tau_{iss'}$ as follows:

$$
\begin{aligned}
q_{iss'} &= \frac{\exp\left(\tau_{iss'}\right)}{1 + \sum\limits_{l=1}^{S-1} \exp\left(\tau_{isl}\right)}, && s, s' = 1, \ldots, S \\
q_{isn} &= \frac{1}{1 + \sum\limits_{l=1}^{S-1} \exp\left(\tau_{isl}\right)}, && s = 1, \ldots, S
\end{aligned}
\tag{3.15}
$$

Additionally, we use a simpler parametrization of M_{it} including only intercepts of the *logistic* regression to compute the probabilities p_{ist} as follows:

$$
p_{ist} = \frac{\exp\left(\hat{\alpha}_s^0\right)}{1 + \exp\left(\hat{\alpha}_s^0\right)} \tag{3.16}
$$

To avoid the label switching problem, we define parameters $\hat{\alpha}_s^0$ as shown on Equation 3.9 using unbounded parameters α_s^0.

3.2.2 Heterogeneity

The variations of the models used in this thesis also include accounting for heterogeneity on different components of the model.

The components where we alternatively account for heterogeneity are:

- Covariate effects on the transition matrices ($\boldsymbol{\rho}_s$)
- Covariate effects on the conditional probabilities ($\boldsymbol{\alpha}_s$)
- Conditional probabilities on the intercept (α_s^0)

Each model used in this thesis describes which components are heterogeneous.

Chapter 4

Simulation Exercise

To assess the effects of the assumption of homogeneity in number of states in a HMM, we design a Monte Carlo simulation exercise.

The objective of this Monte Carlo simulation is to measure the potential bias on the estimation of the parameters of a Hidden Markov Model when customers are heterogeneous in their dynamics and one assumes homogeneity in the number of states.

In order to achieve this objective, the data are generated using two hidden Markov models with different number of states for two segments of customers. Each segment is simulated independently using a single hidden Markov model. Then, we join the data of both segments of customers and estimate a unique HMM with a fixed number of states for all customers. We assume that segment 1 (K_2) has two latent states whereas segment 2 (K_3) has three latent states.

The simulation exercise is structured in two steps. First, we simulate the state transitions and the observable behavior following a hidden Markov model. Second, we use these data to estimate the parameters of a HMM, selecting the model with the number of states that maximizes penalized fit. Finally, to estimate the potential bias on the parameter estimates we use the results of the Monte Carlo simulations to measure the differences between the estimated model with a fixed number of states and the simulated model.

Several factors are manipulated to analyze their effect on the the potential bias when homogeneity in the number of states is assumed. In this thesis we modify: (1) the mixture of customers with 2 and 3 states, (2) the presence of full heterogeneity on the intercept parameters of the conditional probabilities, and (3) the presence of covariates on both transition matrices and conditional probabilities.

In Table 4.1 we summarize the simulation experiments that were performed. Experiment 1.1 is the basic model with no covariates and no heterogeneity in the conditional probabilities, and 50% customers on K_2 and 50% customers on K_3. Experiments 2.x, 3.x and 4.x are modified versions of Experiment 1.1. Experiments 2.x have different mixture of customers. Experiment 3.1 incorporates heterogeneity on the intercept of the conditional probabilities. Finally, Experiments 4.x incorporate covariates, and those experiments differ on the presence of heterogeneity on the model components and whether the covariates are discrete or continuous random variables. For Experiment 4.5 we analyze two scenarios that differ in the magnitude of covariate effects.

Experiment	Mixture of Customers[a]	Covariate[b] X_{kt}	Heterogeneity: Intercept of P α_s^0	Heterogeneity: Effects on Q $\boldsymbol{\rho}_{ks}$	Heterogeneity: Effects on P $\boldsymbol{\alpha}_{ks}$
1.1	50% - 50%	–	–	–	–
2.1	50% - 50%	–	–	–	–
2.2	70% - 30%	–	–	–	–
2.3	80% - 20%	–	–	–	–
2.4	90% - 10%	–	–	–	–
2.5	95% - 5%	–	–	–	–
2.6	98% - 2%	–	–	–	–
2.7	100% - 0%	–	–	–	–
3.1	50% - 50%	–	$\surd$	–	–
4.1	50% - 50%	Continuous	–	–	–
4.2	50% - 50%	Discrete	–	–	–
4.3	50% - 50%	Continuous	–	–	$\surd$
4.4	50% - 50%	Continuous	–	$\surd$	–
4.5.1	50% - 50%	Continuous	$\surd$	–	–
4.5.2	50% - 50%	Continuous	$\surd$	–	–
4.6	50% - 50%	Continuous	$\surd$	$\surd$	$\surd$

Table 4.1: Simulation experiments

[a]Percentage of customers that belong to segments K_2 and K_3, respectively.

[b]If a experiment includes covariates, this column indicates whether these covariates are continuous or discrete.

4.1 Simulation Design

In this section we explain how we designed the Monte Carlo simulations to capture the potential bias on the estimation if homogeneous dynamics are assumed when the number of states of HMM are truly heterogeneous.

4.1.1 Data generation process

Using MATLAB, we simulate data for two segments of customers using two different HMM. The first segment, K_2, is simulated using a HMM with 2 states. The other segment, K_3, is simulated using a HMM with 3 states. We denote the set of all customers as $K = K_2 \cup K_3$.[1]

For all versions of the simulations performed, we consider that the states 1 and 2 are the same for customers in K_2 and K_3, and the higher state could be reachable only by customers in K_3.

[1]By abuse of notation, we use K_2, K_3 and K for referring both the set and subsets of customers and the number of customers in each set, respectively.

Table 4.2 shows the total number of customers, the number of periods T, the number of trials of the Binomial distribution of observable data N_{kt}, and covariate X_{kt}. On these experiments, we use a single covariate, but this can be easily extended to a vector of covariates (see Chapter 6).

Number of customers		$\|K\|$	300
Number of periods		$\|T\|$	20
Number of trials (Binomial distribution)		N_{kt}	100
Covariate			
	Continuous	X_{kt}	$\sim \mathcal{N}(0,1)$
	Discrete	X_{kt}	$\sim B(1,0.5)$

Table 4.2: Parameters used in the Monte Carlo simulations

Following the model specified in Section 3.1, customers transition between different states; each state s is represented by a different parameter p_s of the Binomial distribution of the observable behavior. States are ordered increasingly by the parameter p_s, i.e., a higher state means a state with higher probability p_s.

According to Section 3.1, we set parameters Π_k, Q_k, and p_{kst} as follows:

		Sets	
		K_2	K_3
Number of states	$\|S\|$	2	3
Initial Probabilities	Π_k	$\begin{bmatrix}1 & 0\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0\end{bmatrix}$
Transition Probabilities	Q_k	$\begin{bmatrix}0.7 & 0.3\\0.3 & 0.7\end{bmatrix}$	$\begin{bmatrix}0.7 & 0.2 & 0.1\\0.2 & 0.5 & 0.3\\0.1 & 0.2 & 0.7\end{bmatrix}$
Conditional Probabilities	p_{kst}	$\begin{bmatrix}0.1 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.1 & 0.5 & 0.9\end{bmatrix}$

Table 4.3: Simulation HMM parameters

As described in Section 3.1, parameters Q_k and p_{kst} are computed using parameters $\hat{\tau}_{kss'}$, $\boldsymbol{\rho}_{ks}$ and α_s^0. To obtain the values for Q_k and p_{kst} described on Table 4.3, $\hat{\tau}_{kss'}$, $\boldsymbol{\rho}_{ks}$ and α_s^0 must take specific values. Table 4.4 shows simulated values of transition matrix parameters for models with and without covariates separately[2].

[2] Given the different parametrizations described in Section 3.2.1.

		Sets	
		K_2	K_3
Transition Matrix Parameters without Covariates			
Intercept of transition probabilities	τ_{k11}	0.85	1.95
	τ_{k12}	.	0.69
	τ_{k21}	.	-0.41
	τ_{k22}	0.85	0.51
	τ_{k32}	.	0.69
	τ_{k33}	.	1.95
Transition Matrix Parameters with Covariates			
Transition threshold parameters	$\hat{\tau}_{k11}$	0.85	0.85
	τ_{k12}	.	0.30
	τ_{k21}	-0.85	-1.39
	τ_{k22}	.	0.80
	τ_{k31}	.	-2.20
	τ_{k32}	.	0.30
Covariate effects[a]	$\boldsymbol{\rho}_{k1}$	0.51	0.51
	$\boldsymbol{\rho}_{k2}$	0.48	0.48
	$\boldsymbol{\rho}_{k3}$	.	0.41

Table 4.4: The simulated values for parameters corresponding to the transition matrix

[a]The simulated effects of the covariates are different on experiments 4.x. These specific values are simulated values for experiment 4.1. The values simulated for each experiment are detailed in Appendix A.

In addition, Table 4.5 shows simulated values of conditional probabilities. Models with and without covariates are simulated using the same intercepts.

		Sets	
		K_2	K_3
Conditional Parameters			
Intercept	α_1^0	-2.20	-2.20
	α_2^0	0.79	0.79
	α_3^0	.	0.79
Covariates Effects[a]	$\boldsymbol{\alpha}_1$	0.26	0.26
	$\boldsymbol{\alpha}_2$	0.46	0.46
	$\boldsymbol{\alpha}_3$	.	0.63

Table 4.5: The simulated values for parameters corresponding to the conditional probabilities

[a] Covariate effects simulated parameters are different on experiments 4.x. These specific values are simulated values for experiment 4.1, however simulated values for each experiment are detailed in Appendix A.

When we account for heterogeneity in a specific component of the model the values of the parameters of that component described on Tables 4.4 and 4.5 represent the mean of the corresponding parameters.

We simulate individual parameters θ_k by drawing from a Multivariate Normal distribution with mean μ_θ and covariance matrix Σ_θ. As mentioned before, we set μ_θ values as described in Tables 4.4 and 4.5 for each component of the model. On the other hand, we generate the covariance matrix Σ_θ as follows[3]:

$$\Sigma_\theta = \frac{1}{50}\, diag(|\mu_\theta|)$$

Appendix A details the standard deviation of heterogeneous parameters for each experiment.

The simulation procedure is performed as follows:

Simulation procedure

For each segment K_n, with $n \in \{2, 3\}$:

Step 1: Initialize μ_θ and Σ_θ.

Step 2: Initialize Φ.

Step 3: For each $i \in K_n$ draw $\theta_i \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$

Step 4: For each $i \in K_n$ draw X_{it} from the corresponding distribution (see Table 4.2)

[3] We use $|x|$ for the vector of absolute values of x, therefore $|x|_i = |x_i|$.

Step 5: Compute Q_{it} from θ_i, Φ and X_{it}.

Step 6: Compute p_{ist} from θ_i, Φ and X_{it}.

Step 7: Compute z_{it} using a discrete time discrete space first order Markov chain with transition matrix Q_{it}.

Step 8: Draw $Y_{it}|z_{it} = s$ using a Binomial distribution with parameters N_{it} and p_{ist}.

4.1.2 Estimation procedure

The model is estimated using a Bayesian approach from simulated data Y. In particular, the model parameters are estimated using a Markov Chain Monte Carlo procedure assuming a fixed number of states for the HMM (2 and 3) choosing the one that fits the data best.

It is important to recall that in this estimation procedure we do not estimate the initial probabilities (Π_k) and the number of trials of the Binomial distribution (N_{kt}), therefore we consider those parameters as given.

In Appendix B we described the estimation model with priors and full conditionals distribution. We estimate the model using a Markov chain Monte Carlo technique implemented in MATLAB, specifically, a Gaussian random-walk Metropolis-Hastings algorithm (M-H), described in Appendix C. We use the modified version of M-H introduced by Atchade [2006], and described in Appendix D.

Finally, to select the best model we compare HMM with 2 and 3 states using log-marginal likelihood (LML) criterion, the deviance information criterion (DIC) and the Markov switching criterion (MSC).

4.2 Results

4.2.1 Basic Model: Experiment 1.1

For Experiment 1.1, we report in Table 4.6 the model comparison criteria. Criteria DIC and MSC identify the 3 states HMM as the best model.

Model	**LML**	**DIC**	**MSC**
2 states	-71280.89	142637.53	152197.16
3 states	**-30750.53**	**61699.86**	**72471.99**

Table 4.6: Experiment 1.1: Model Comparison

We report in Table E.1 on Appendix E the parameter estimates of Experiment 1.1. Standard

Monte Carlo simulation reports consider the comparison of simulated and estimated parameters, and concluding if those parameters are correctly recovered. However, on these simulation experiments the set of simulated and estimated parameters have different dimensions.

The estimated model has parameters according to a HMM with 3 states. Therefore, for customers in segment 1 with two states, the simulated parameters and the estimated parameters have different interpretations. Furthermore, the population parameters represent the aggregate behavior of all customers (including those customers with two true hidden states). However, the simulated experiment do not have population parameters because the simulated parameters for different segments have different dimensions depending on the segments of the customers. Thus, to analyze if the estimated model is biased we compare the transition matrices and the conditional probabilities computed using the simulated and the estimated parameters. To accomplish this, we compare separately customers with 2 and 3 real states, by averaging the individual parameters within segments 1 and 2.

Let θ_{kl} be the l'th draw of parameter θ for customer k. We compute $\mu^2_{\theta l}$ as the average of θ_{kl} with $k \in K_2$.

$$\mu^2_{\theta l} = \frac{1}{K_2} \sum_{k \in K_2} \theta_{kl}$$

Analogously, we define $\mu^3_{\theta l}$ as the average of the l'th draw of parameter θ across customers in K_3.

$$\mu^3_{\theta l} = \frac{1}{K_3} \sum_{k \in K_3} \theta_{kl}$$

In summary, we compute a different 'population mean' for customers with 2 and 3 states.

Next, for each draw l, we compute separately a transition matrix Q_l using θ^2_l and θ^3_l.

In addition, we compute the transition matrix using population mean μ_θ for each draw (which aggregates behavior of all customers in K), as it is common[4] to report transition matrices computed using the posterior mean of μ_θ.

Finally, we compute the mean and the 95% CI for those transition matrices, and we compare them with true values[5].

Given that conditional are homogeneous on Experiment 1.1 across customers, we report the posterior mean and 95% CI of p_{kst}.

We report in Table 4.7 the comparison of posterior mean transition matrix (with its corresponding 95% CI below) and simulated transition matrix using the methodology explained before.

[4] Montoya et al. [2010] and Netzer et al. [2008].

[5] Simulated transition matrix does not apply for the aggregate analysis of customers in K given the mixture of customers with 2 and 3 states.

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	**1**	0.70	0.29	0.01	0.70	0.30	NA
		[0.68 0.72]	[0.27 0.31]	[0.01 0.01]			
	2	0.31	0.67	0.01	0.30	0.70	NA
		[0.29 0.33]	[0.65 0.69]	[0.01 0.02]			
	3	0.16	0.16	0.68	NA	NA	NA
		[0.13 0.21]	[0.12 0.20]	[0.61 0.74]			
$\mathbf{K_3}$	**1**	0.70	0.21	0.08	0.70	0.20	0.10
		[0.68 0.72]	[0.19 0.23]	[0.07 0.10]			
	2	0.18	0.55	0.27	0.20	0.50	0.30
		[0.16 0.20]	[0.52 0.58]	[0.24 0.30]			
	3	0.11	0.21	0.68	0.10	0.20	0.70
		[0.10 0.13]	[0.19 0.22]	[0.66 0.70]			
K	**1**	0.72	0.24	0.04			
		[0.70 0.73]	[0.23 0.26]	[0.03 0.05]			
	2	0.24	0.65	0.10		NA	
		[0.22 0.26]	[0.63 0.68]	[0.08 0.13]			
	3	0.13	0.19	0.68			
		[0.12 0.15]	[0.16 0.21]	[0.65 0.71]			

Table 4.7: Experiment 1.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table 4.8: Experiment 1.1 conditional probabilities posterior mean and 95% CI

Confidence intervals for p_{kst} shown on Table 4.8 indicate that the model can capture correctly the conditional behavior on each state, despite the heterogeneity in the number of states. This result is important to fix the states for all customers (K_2 and K_3), i.e. a specific state means the same in terms of the conditional behavior for all customers (a probability of 0.1, 0.5 and 0.9 for each state). Quantiles for Q on K_2 and K_3 also show that the states are correctly identified (the states represent the same for all customers). For customers in K_2, the transitions between states 1 and 2 are correctly recovered and the third state becomes unreachable from those states (the probability of moving to state 3 from the others is close to zero). For customers in K_3, the transition matrix is correctly recovered. These results suggest a good recovery of the parameters at the individual level.

We report in Table 4.9 the individual-level recovery of transition probabilities. The parameter is recovered if the posterior interval of q_{ij} contains the respective simulated transition probability. For customers in K_2 is only countable the 2×2 upper-left sub-matrix given that those customers only move through the first two states, i.e. the total number of transition probabilities that need to be recovered to estimate correctly the transition behavior is 4. For customers in K_3 on the other hand, all 9 transition probabilities need to be recovered.

This table shows that 96.7% of customers in K_2 have all 4 transition probabilities estimated correctly. For K_3, 75.3% of customers have all 9 transition probabilities estimated correctly; and

Set of customers	Customers #	%	q_{ij} recovered #	%
K_2	145	96.7 %	4	100.0 %
	5	3.3 %	3	75.0 %
K_3	113	75.3 %	9	100.0 %
	16	10.7 %	8	88.9 %
	7	4.7 %	7	77.8 %
	9	6.0 %	6	66.7 %
	2	1.3 %	5	55.6 %
	3	2.0 %	4	44.4 %

Table 4.9: Experiment 1: Number and percentage of transition probabilities q_{ij} recovered for sets K_2 and K_3

using the first three rows, 90.7% of customers in K_3 have at least 75% of 9 transition probabilities correctly identified. This result shows a good recovery at the individual level.

At the aggregate level instead, results are different. Using the results of the first two rows on Table 4.7, on average, customers on K_2 move between states 1 and 2, and the probability of moving to the third state is close to zero regardless of which state the customer is at. On the other hand, customers on K_3 have a 27% probability to move from state 2 to state 3. These results can not be captured by observing only the aggregate level. In fact, the third row shows that the probability of moving from state 2 to state 3 is 10%, but this probability lies outside of both confidence intervals of K_2 and K_3 (see Figure 4.1).

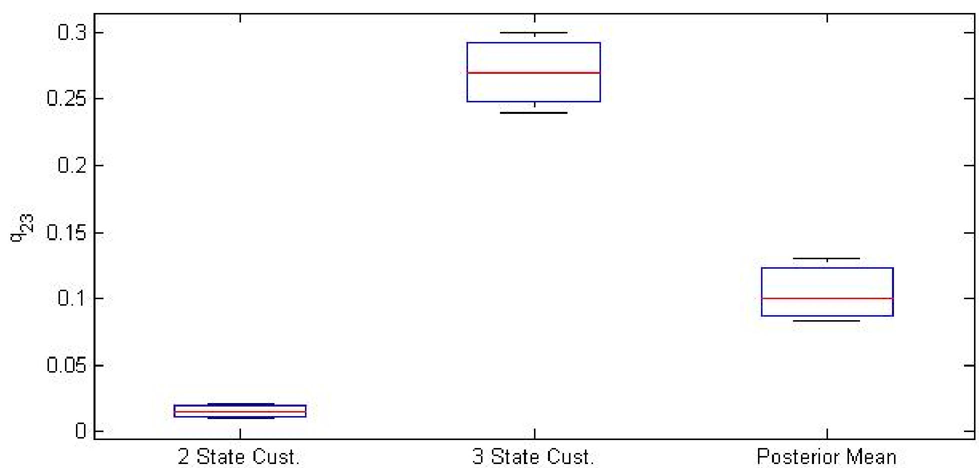
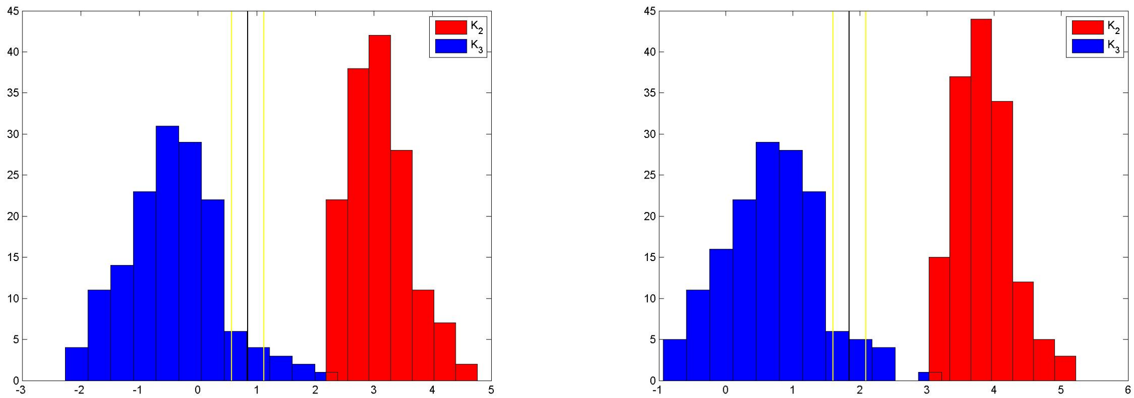


Figure 4.1: 95% confidence intervals of $q_{23}(\cdot)$ using θ_l^2, θ_l^3 and population mean μ_θ

Given the model described in Section 3.1 parameters τ_{k21} and τ_{k22} generate the second row of the transition matrix Q, i.e., the probabilities to move (or not) from state 2.



(a) Heterogeneity in τ_{k21} (b) Heterogeneity in τ_{k22}

Figure 4.2: Heterogeneity in τ_{k21} and τ_{k22}.

We show in Figure 4.2 the distribution of τ_{k21} and τ_{k22} across customers. In red (blue), the histogram of the mean of τ_{k21} and τ_{k22} for each customer of K_2 (K_3) across draws, i.e., showing the heterogeneity within K_2. The black line represents posterior mean of population parameter μ_θ for components τ_{21} and τ_{22}. Yellow lines represents the 95% confidence interval of population parameter μ_θ for τ_{21} and τ_{22}.

These graphs show that the heterogeneity distribution on τ_{k21} and τ_{k22} are bimodal, and μ_θ does not represent either the set K_2 or K_3.

In conclusion, we showed that this model has an acceptable recovery of the parameters at the individual level. However, at the aggregate level, the heterogeneity in θ is misinterpreted if observing only the population parameter μ_θ and the transition matrix obtained from that parameter $Q(\mu_\theta)$, yielding results that do not represent the transition behavior of any customer (on K_2 nor K_3).

4.2.2 Mixture of customers: Experiments 2.x

In the previous section we showed that the probability of moving to the third state is biased at the aggregate level. In this section, we test if a smaller proportion of customer reaching the higher state could affect the bias of this probability at the aggregate level.

We report in Table 4.10 the model comparison criteria for each experiment. The model with 3 states is indicated by both DIC and MSC as the best model for all experiments but experiment 2.7 (100%-0%).

Experiment	**Proportion**	**States**	**LML**	**DIC**	**MSC**
2.1	50% 2 States - 50% 3 States	2	-71280.89	142637.53	152197.16
		3	**-30750.53**	**61699.86**	**72471.99**
2.2	70% 2 States - 30% 3 States	2	-58026.75	116101.98	125611.97
		3	**-30489.47**	**61147.56**	**71814.21**
2.3	80% 2 States - 20% 3 States	2	-50564.49	101161.76	110608.53
		3	**-30345.92**	**60818.21**	**71440.15**
2.4	90% 2 States - 10% 3 States	2	-40810.01	81650.51	91015.99
		3	**-30172.50**	**60477.85**	**71551.59**
2.5	95% 2 States - 5% 3 States	2	-36169.86	72372.65	81673.22
		3	**-30133.04**	**60365.64**	**63062.98**
2.6	98% 2 States - 2% 3 States	**2**	**-32077.28**	64180.14	73436.39
		3	-32087.83	**64177.01**	**73406.71**
2.7	100% 2 States - 0% 3 States	**2**	**-29951.47**	59925.04	**69171.60**
		3	-29953.43	**59921.84**	69243.06

Table 4.10: Experiments 2: Model Comparison

In Appendix F we report the parameter estimates for each experiment.

The results show that as the proportion of customers of K_2 grows, the probability of going from state 2 to state 3 (computed using the population mean μ_θ) decreases. Figure 4.3 shows how the 95% CI of q_{23} decreases when the proportion of customers in K_2 increases.

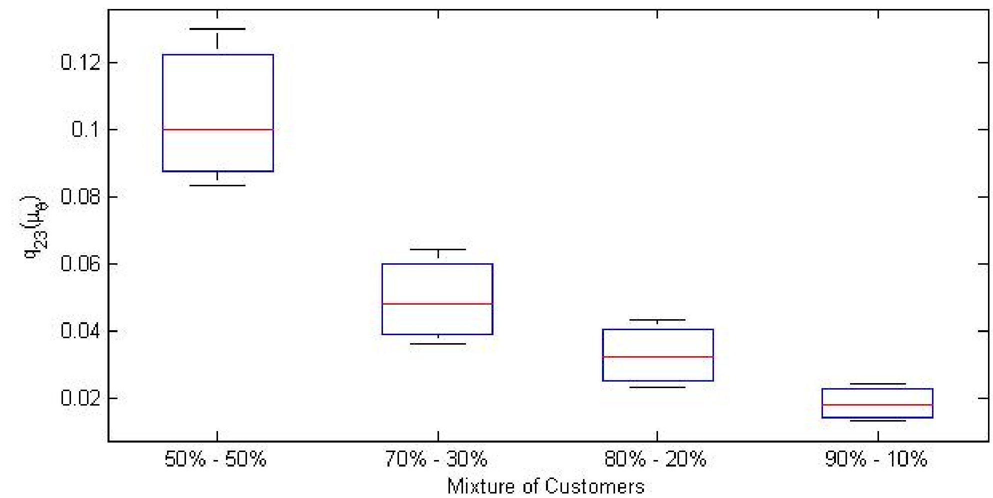


Figure 4.3: 95% confidence intervals of $q_{23}(\cdot)$ using population mean μ_θ by mixture of customers

This effect has two main causes. First, the weight of the parameters of segment 1 increases on the estimation of the population parameters[6]. Second, there is a decreasing trend of q_{23} for customers in K_2 as shown on Figure 4.4. This plot shows the same decreasing trend than the probabilities computed using the population mean μ_θ. On the other hand, this relationship cannot be observed clearly for customers K_3 (Figure 4.5), as there are true parameters q_{23} for segment 2. However, the variance of this estimation increases due to the decrease in the number of observations (customers) of this segment.

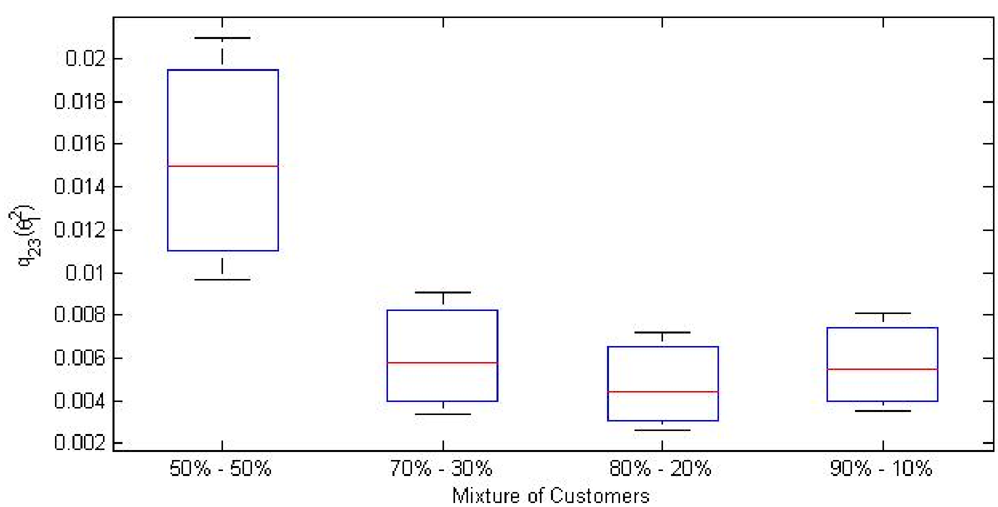


Figure 4.4: 95% confidence intervals of $q_{23}(\cdot)$ for customers in K_2 by mixture of customers

[6]In the Gibbs step that updates μ_θ on the M-H algorithm described in Section C the average of individual parameters ($\bar{\theta}$), captures the values of the individual parameters at the aggregate level, therefore if the proportion of customers with 2 states increases, the behavior of this group becomes more relevant on the computation of $\bar{\theta}$ and μ_θ.

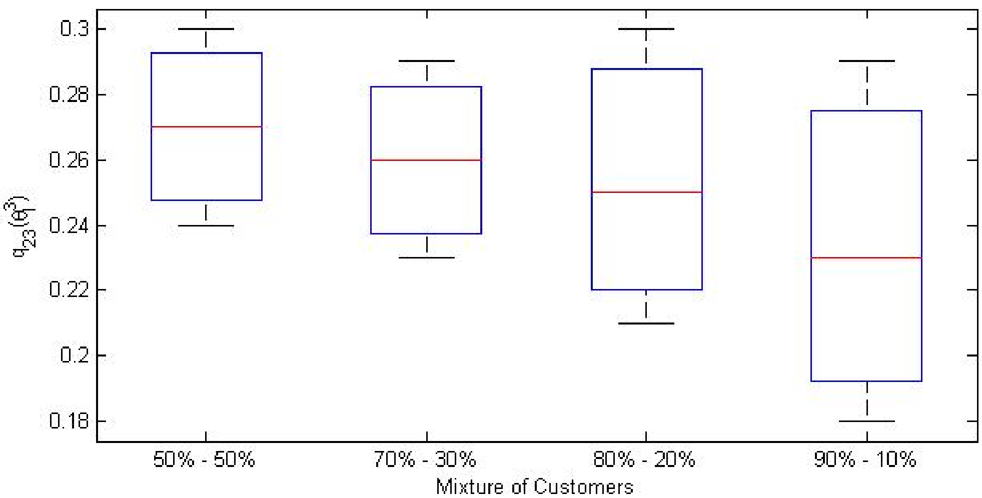


Figure 4.5: 95% confidence intervals of $q_{23}(\cdot)$ for customers in K_3 by mixture of customers

Using Table 4.10 we can also conclude that the HMM with 3 states is chosen in all experiments with more than 5 customers in segment 2. The main reason is that the 3-state HMM is more flexible than a 2-state HMM. Therefore, the behavior of a customer with 2 states can be captured by a 3-state model but for a customer with 3 states, the 2 state model can not capture correctly the behavior. We explain this effect using the results of P and Q on Appendix F.

In this context the HMM with 3 states performs considerably better at the individual level for customers with 3 states than a 2 state HMM, and it does not perform worse for customers with 2 states. Then, the greater number of parameters on 3 state HMM is the only penalty that can make a 3-state model worse than a 2 state model. For this reason, the 3-state model is the best model on almost every mixture tested in this simulated exercise, and all states are identified as different between each other on those estimations[7]. The only exception is the 98%-2% proportion, where the lower state is estimated by two states and the state on the middle is estimates as the higher state (see Appendix F.2.5).

In summary, increasing the proportion of customers with less states can lead to worse aggregate results. In this case, we showed that the wrong inferences described from Experiment 1.1 can increase if the proportion of customers with fewer states increases.

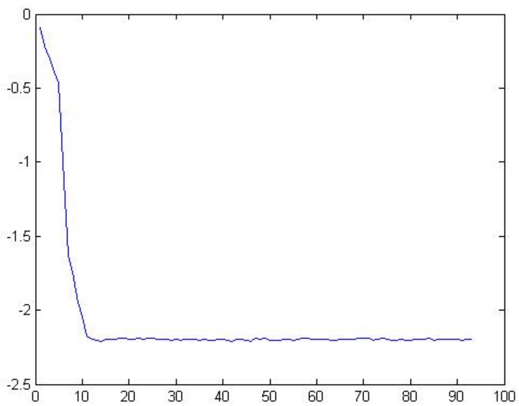
Next, we investigate the effect of including heterogeneity in the potential bias.

[7]When no heterogeneity is accounted for α^0.

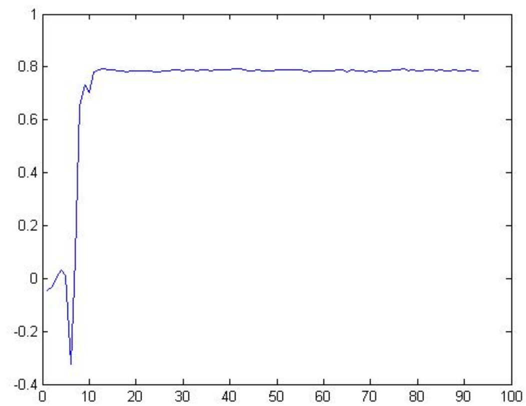
4.2.3 Heterogeneity in the Conditional Probabilities: Experiment 3.1

For models of Experiments 1.1 and 2.x the conditional behavior probabilities are fixed for all customers (not accounting for heterogeneity on α^0). However, there are examples in the literature that model otherwise. Netzer et al. [2008] introduces a HMM where the intercepts of the parameters that control the conditional behavior are estimated at the individual level.

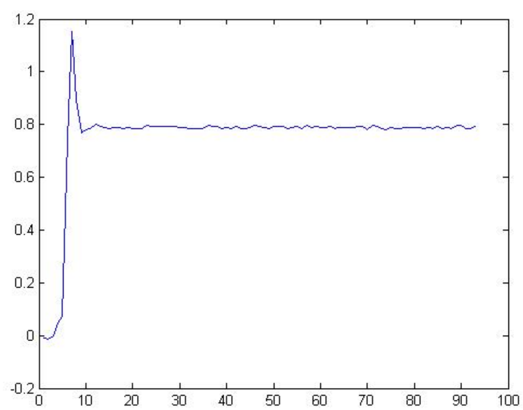
As expected, homogeneous parameters Φ converge faster in comparison with the individual level parameters θ_k. We show in Figure 4.6 how parameter α_s^0 in Experiment 1.1 converges on the first 1900 iterations[8] still in the burn-in period of the estimation algorithm (first 40000 iterations).



(a) Convergence of α_1^0



(b) Convergence of α_2^0



(c) Convergence of α_3^0

Figure 4.6: Convergence of α^0

Fast convergence of α^0 allows to identify each state properly, therefore the transition matrix can be correctly recovered. If conditional probabilities are not correctly identified, then it is impossible for the model to capture the transition between the states properly.

[8] The algorithm saves the draws every 20 iterations to avoid high correlations, therefore plots in Figure 4.6 have approx. 90 iterations shown in the horizontal axis.

In this experiment we study the effects of accounting for heterogeneity in the conditional probabilities (α^0) on the estimation results when there is heterogeneity in the number of states.

The choice of the number of states that fits the data best was based on the log-marginal likelihood (LML) criterion, the deviance information criterion (DIC) and the Markov switching criterion (MSC). Given these criteria, we choose the 3 states model as the right model (see Table 4.11).

Model	**LML**	**DIC**	**MSC**
2 states	-48745.88	97999.16	152197.16
3 states	**-33804.60**	**68316.14**	**85675.77**

Table 4.11: Experiment 3.1: Model Comparison

As explained in Section 4.2.1 we show results for Q and P given the different dimension of simulated parameters. However, we report the parameter estimates of the 3-state HMM in Table E.9, on Appendix E.3. We report in Table 4.12 posterior mean and 95% CI for transition matrix Q, computed separately using customers in K_2, customers in K_3 and population mean μ_θ.

In addition, we report in Table 4.13 posterior mean and 95% CI for conditional probabilities p_{ks}, computed separately using customers in K_2, customers in K_3 and population mean μ_θ.

		Posterior Mean			**Simulated**		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.71	0.02	0.27	0.70	0.30	NA
		[0.69 0.73]	[0.01 0.03]	[0.26 0.29]			
$\mathbf{K_2}$	**2**	0.27	0.24	0.49	0.30	0.70	NA
		[0.20 0.33]	[0.18 0.32]	[0.42 0.54]			
	3	0.30	0.01	0.69	NA	NA	NA
		[0.28 0.32]	[0.01 0.02]	[0.67 0.71]			
	1	0.70	0.18	0.12	0.70	0.20	0.10
		[0.68 0.72]	[0.17 0.20]	[0.10 0.13]			
$\mathbf{K_3}$	**2**	0.20	0.51	0.28	0.20	0.50	0.30
		[0.18 0.23]	[0.49 0.54]	[0.26 0.31]			
	3	0.12	0.18	0.70	0.10	0.20	0.70
		[0.10 0.13]	[0.17 0.20]	[0.68 0.72]			
	1	0.75	0.07	0.19			
		[0.72 0.76]	[0.05 0.09]	[0.17 0.20]			
K	**2**	0.24	0.37	0.38		NA	
		[0.21 0.27]	[0.32 0.43]	[0.34 0.42]			
	3	0.20	0.05	0.75			
		[0.18 0.22]	[0.04 0.07]	[0.73 0.77]			

Table 4.12: Experiment 3.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	0.10	0.16	0.51	0.10	0.50	NA
	[0.10 0.10]	[0.15 0.18]	[0.51 0.51]			
$\mathbf{K_3}$	0.10	0.49	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.49 0.49]	[0.90 0.90]			
K	0.10	0.26	0.71		NA	
	[0.10 0.10]	[0.22 0.29]	[0.68 0.74]			

Table 4.13: Experiment 3.1 conditional probabilities posterior mean and 95% CI

As described in Section 4.2.1, the conditional probabilities are correctly identified on Experiment 1.1. The model identifies three states, with probabilities 0.1, 0.5 and 0.9. Customers with two states are modeled with the same states but the probability of making transitions to the higher state is 1%. For further references to this, we define this result of identifying correctly the three state as EF1.

In this experiment, results are different. As the intercept α^0 is not fixed across customers in this experiment all parameters of the model are heterogeneous. Besides, the relationship of parameters across customers is weaker than in Experiment 1.1 since in this experiment there is no fixed parameter across customers. Only heterogeneous parameters such as population mean μ_θ and covariance matrix Σ_θ account for a relationship between parameters of different customers. Consequently, the individual parameters of a specific customer are estimated to capture the behavior of each hidden state depending on the data of that customer. For customers with three states, the model identifies three different latent behavior and account for them on the estimation. Customers with two states have data generated by only two latent states. However, the model forces to capture the behavior using three states for all customers. This means that for customers in K_2, which were not simulated with a state of probability 0.9, the estimation procedure identifies a state that is equal to one of the two lower states. For further references of this effect, we define this result as EF2.

Table 4.13 indicates the presence of EF2 on Experiment 3.1. For customers in K_2, states 1 and 2 have both conditional probabilities p_{ks} that represent the lower simulated state with probability $p_{ks} = 0.1$. State 3 represents the second state simulated with probability $p_{ks} = 0.5$. Given this estimation results for parameters α^0, transition probabilities are estimated accordingly. Table 4.12 shows that the probabilities of moving from state 1 and 3 to state 2 are 2% and 1%, respectively. Customers in state 1 and 3 have probability of approximately 70% of remaining in the same state, as it has been simulated for both states for customers in K_2. These results show that transitions of customers in K_2 occur mostly between states 1 and 3, i.e., between two states with p_{ks} 0.1 and 0.5 as simulated.

Although for customers in K_3, Table 4.12 and Table 4.13 show good recovery of both transition matrices and conditional probabilities, duplicated states for customers in K_2 generates bias on aggregate results. At the aggregate level (considering the results using μ_θ) parameters Q and p_{ks} are affected by the mixture of states that do not capture the same behavior in terms of conditional probabilities. Table 4.13 shows that the lower state is the only one recovered in terms of p_{ks}. The other estimated states represent behaviors where p_{ks} is 0.25 and 0.7, both values of p_{ks} do not represent any of the two segments of customers. In addition, Table 4.12 shows also a transition matrix that do not represent either the dynamics of customers in K_2 or K_3. Furthermore, μ_θ is estimated as the mean of the distribution of θ_k over the population of customers, therefore the transition matrix at the aggregate level can not be interpreted as the average of transition matrix but as the transition matrix of the average of the individual parameters, given that the average is estimated at the parameter level (θ_k) and logistic transformation for computing Q is nonlinear. The most intuitive examples of it are the transition probabilities involving state 2, e.g. from state 1 to state 2. The probability of moving from state 1 to state 2 is reported as 7% at the aggregate level, however at the individual level it is 2% for customers in K_2 and 18% for customers in K_3, which do not average 7%. Moreover, those probabilities do not represent the same dynamics in terms of changes in parameter p_{ks} of the Binomial distribution. Similar effects could be observed on transition probabilities q_{13}, q_{22}, q_{23}, q_{31} and q_{32}.

Confidence intervals of q_{11} are another example of bias on the aggregate level. Considering results over K_2 and K_3 separately, q_{11} is 0.71 and 0.70 respectively, but we observe that q_{11} computed using μ_θ is between 0.72 and 0.76 with 95% of probability. This effect is caused by the differences of the probabilities q_{12} and q_{13} between customers in K_2 and K_3. q_{12} is 0.02 for K_2 but 0.18 for K_3, and q_{13} is 0.27 for K_2 but 0.12 for K_3. In addition, transition matrix at aggregate level is computed using μ_θ, therefore logistic transformation is performed over the mean of parameters and does not represent the average of transition matrices but the transition matrix of the average of parameters. Accordingly, q_{11} increases while the other probabilities of the same row take values between the values of customers in K_2 and K_3.

To illustrate this, we introduce a simple example, where we change the parametrization of Q. We define $e_1 = 0$, e_2 and e_3 such as $q_{1s} = \frac{exp(e_s)}{\sum_l exp(e_l)}$. Considering this definition, we can compute e_s from q_{1s} as follows:

$$e_s = log\left(\frac{q_{1s}}{q_{11}}\right)$$

Using the values of q_{1s} for customers in K_2 and K_3, we obtain $e_1^2 = 0$, $e_2^2 = -3.57$, $e_3^2 = -0.97$ and $e_1^3 = 0$, $e_2^3 = -1.36$, $e_3^3 = -1.76$.

Then, we average e_s of K_2 and K_3 to obtain $e_1^\mu = 0$, $e_2^\mu = -2.46$, $e_3^\mu = -1.37$. If we compute q_{1s} from e_s^μ, we obtain transition probabilities 75%, 6% and 19% of moving to the first, second and third state. It is interesting to observe how even both customer segments have probabilities q_{11} similar between them ($\sim 70\%$), aggregate results could show biased values. This effect is caused by averaging on the parameters and compute a logistic transformation on those averaged values.

Taken together, we conclude that, for this experiment, using aggregate results could lead to misinterpretation of the customer behavior at the individual level. It is important to highlight that on Experiment 1.1 the parameters are only biased on transition matrix given that states were fixed in terms of conditional probabilities for all customers, whereas on Experiment 3.1 conditional probabilities are also biased at aggregate level.

4.2.4 Covariates: Experiments 4.x

As there are many applications that incorporate marketing actions in the HMM (Netzer et al. [2008] and Montoya et al. [2010]), we include covariates in the model. In this section we present the results of several experiments that incorporate covariates both in the transition matrix and the conditional probabilities in the model described in Section 3.1. Throughout the experiments of this section, we modify the presence of heterogeneity in the components of the model. Experiments 4.5.1 and 4.5.2 differ on the magnitude of the effects simulated (see Appendix A.6).

		Heterogeneity		
Experiment	**Covariate**[a] X_{kt}	**Intercept of** P α_s^0	**Effects on** Q $\boldsymbol{\rho}_{ks}$	**Effects on** P $\boldsymbol{\alpha}_{ks}$
4.1	Continuous	×	×	×
4.2	Discrete	×	×	×
4.3	Continuous	×	×	√
4.4	Continuous	×	√	×
4.5.1	Continuous	√	×	×
4.5.2	Continuous	√	×	×
4.6	Continuous	√	√	√

Table 4.14: Simulation experiments 4.x

[a] If a experiment include covariates, this column indicates whether these covariates are continuous or discrete.

For all experiments in this section, LML, DIC and MSC imply that the 3-states HMM fits better than 2-states HMM.

In Experiment 4.1, both covariates effects ($\boldsymbol{\rho}$, $\boldsymbol{\alpha}$) and conditional probabilities parameters (α^0) are recovered correctly (see Appendix E.4.1). Heterogeneous parameters at the individual level (τ) are correctly recovered as well, but not at the aggregate level. Considering the results of Q and p_{ks} in Appendix F.4.1, we could observe EF1[9] on experiment 4.1. This means that even if there is no customer with 2 states reaching the third state, conditional probability p_{ks} for each state is the same for all customers. Consequently, the effects of the covariates can be estimated correctly because states have the same conditional distribution for all customers and covariate effects are modeled as homogeneous.

As estimation of Experiment 1.1, customers with 2 states have a low probability (~ 0) of transitioning from state 1 or 2 to state 3, therefore those customers move between states 1 and 2 following a Markov process as simulated (with transition matrix $\sim \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$). Dynamics of customers with 3 states are correctly recovered. However, as well as in Experiment 1.1, results of Experiment 4.1 show that the transition matrix computed at the aggregate level (using μ_θ) could lead to misinterpretation of customer dynamics. The probability of transitioning from state 2 to state 3 is biased at the aggregate level as it was in Experiment 1.1. In this case, aggregate results show that there is a 4% probability of transitioning to state 3 from state 2, but 50% of the

[9] Described in Section 4.2.1. States are fixed for all customers due to the homogeneity on α^0.

customers have a 28% probability of making such transition, and the other 50% have probability 0% of transitioning to state 3 from state 2.

Experiments 4.2, 4.3 and 4.4 present similar results. Covariate effect parameters are recovered in most cases, and when some parameters are not recovered, those simulated parameters are slightly off the posterior confidence intervals. Individual level parameters are also correctly recovered, as well as the intercept of conditional probabilities α^0. These experiments also present the EF1, i.e., states are identified as the same for all customers.

Considering the results of Section 4.2.3, it is interesting to see the results of Experiment 4.5. The main objective of this experiment, is to observe if states can be correctly identified for all customers if covariate effects are homogeneous. For this case two instance were studied. The first one (4.5.1) has similar values for simulated covariate effects across states, whereas in the second instance (4.5.2) we simulated using significantly different covariate effects values across states (for details about simulated parameters see Appendix A.6).

		Posterior Mean[a]			**Simulated**		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70 [0.68 0.71]	0.04 [0.04 0.07]	0.26 [0.23 0.28]	0.70	0.30	NA
$\mathbf{K_2}$	**2**	0.27 [0.24 0.36]	0.12 [0.10 0.16]	0.61 [0.50 0.65]	0.30	0.70	NA
	3	0.25 [0.23 0.27]	0.06 [0.04 0.10]	0.69 [0.65 0.71]	NA	NA	NA
	1	0.70 [0.69 0.72]	0.19 [0.17 0.20]	0.11 [0.10 0.12]	0.70	0.20	0.10
$\mathbf{K_3}$	**2**	0.22 [0.21 0.25]	0.48 [0.45 0.51]	0.30 [0.26 0.32]	0.20	0.50	0.30
	3	0.13 [0.12 0.14]	0.19 [0.17 0.21]	0.69 [0.66 0.70]	0.10	0.20	0.70
	1	0.70 [0.69 0.71]	0.10 [0.09 0.13]	0.20 [0.17 0.22]			
K	**2**	0.24 [0.22 0.30]	0.26 [0.21 0.30]	0.50 [0.43 0.54]		NA	
	3	0.18 [0.16 0.20]	0.10 [0.09 0.14]	0.72 [0.68 0.73]			

Table 4.15: Experiment 4.5.1 transition matrix posterior mean and 95% CI

[a] To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	0.10	0.46	0.54	0.10	0.50	NA
	[0.10 0.10]	[0.45 0.47]	[0.54 0.56]			
$\mathbf{K_3}$	0.11	0.51	0.89	0.10	0.50	0.90
	[0.10 0.11]	[0.51 0.52]	[0.88 0.90]			
K	0.10	0.49	0.70		NA	
	[0.10 0.11]	[0.47 0.50]	[0.66 0.73]			

Table 4.16: Experiment 4.5.1 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.69	0.31	0.00	0.70	0.30	NA
		[0.68 0.71]	[0.29 0.32]	[0.00 0.00]			
$\mathbf{K_2}$	**2**	0.29	0.70	0.00	0.30	0.70	NA
		[0.28 0.31]	[0.69 0.72]	[0.00 0.00]			
	3	0.24	0.20	0.55	NA	NA	NA
		[0.21 0.36]	[0.18 0.26]	[0.41 0.60]			
	1	0.69	0.20	0.10	0.70	0.20	0.10
		[0.68 0.71]	[0.19 0.22]	[0.09 0.12]			
$\mathbf{K_3}$	**2**	0.22	0.50	0.28	0.20	0.50	0.30
		[0.20 0.24]	[0.46 0.53]	[0.25 0.32]			
	3	0.11	0.19	0.69	0.10	0.20	0.70
		[0.10 0.13]	[0.18 0.22]	[0.66 0.71]			
	1	0.69	0.28	0.02			
		[0.68 0.70]	[0.27 0.30]	[0.02 0.04]			
K	**2**	0.26	0.71	0.04		NA	
		[0.24 0.27]	[0.69 0.73]	[0.02 0.06]			
	3	0.17	0.20	0.63			
		[0.15 0.22]	[0.18 0.25]	[0.54 0.65]			

Table 4.17: Experiment 4.5.2 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	0.10	0.51	0.90	0.10	0.50	NA
	[0.10 0.10]	[0.51 0.52]	[0.88 0.92]			
$\mathbf{K_3}$	0.10	0.49	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.49 0.50]	[0.90 0.90]			
K	0.10	0.50	0.90		NA	
	[0.10 0.10]	[0.49 0.51]	[0.89 0.91]			

Table 4.18: Experiment 4.5.2 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

We show in Tables 4.15 and 4.16 posterior mean and CI of Q and p_{ks} for Experiment 4.5.1. In addition, we show in Tables 4.17 4.18 the corresponding results for Experiment 4.5.2.

We observe that if covariate effects are significantly different across states, then those parameters fixed the states for all customers, i.e., the states have conditional probabilities 0.1, 0.5 and 0.9 for all customers even if α^0 is modeled as heterogeneous. This means that the estimated model follows EF1. However, when covariate effects are similar, those parameters do not provide differentiation from states, therefore EF2[10] is present in the estimation model and duplicated states are identified for customers in K_2 (for details see Appendix F.4.5 and F.4.6). It is important to recall that, even though these experiments do not have a different simulation structure and they only differ on the values of simulated parameters, experiments 4.5.1 and 4.5.2 present very different results in terms of how the states are identified across customers.

For experiment 4.5.2, only the transition matrix at the aggregate level is biased (and only the probabilities related to transitions to the higher state), whereas for experiment 4.5.1 conditional probabilities are also biased at the aggregate level, and the transition matrix is completely biased due to different identification of states for customers in K_2 and K_3.

Following the previous analysis, in Experiment 4.6 we incorporate heterogeneity on all parameters. Tables F.29 and F.30 in Appendix F show that state with conditional probability $p_{ks} = 0.5$ is identified for two states for customers in K_2 (EF2). This effect causes not only misinterpretation on the aggregate results of Q and P but also underestimation of the effects on the transition matrix of covariates for the higher state ($\boldsymbol{\rho}$), as shown in Table E.16 (Appendix E.4.6). Parameter $\boldsymbol{\rho}_3$ is simulated as 0.8 and posterior mean is 0.37 (95% CI: [0.29 0.45]). This biased result is caused due to different interpretations of state 3 for customers in K_2 and K_3. For customers in K_2, the third state correspond to the middle simulated state, which has a simulated effect of $\boldsymbol{\rho}_2 = 0$. On the other hand, for customers in K_3 the same state correspond to the higher simulated state which was simulated with $\boldsymbol{\rho}_3 = 0.8$.

Figure 4.7 shows the histogram of individual level posterior means for $\boldsymbol{\rho}_3$, with customers in K_2 marked in red, and customers in K_3 marked in blue. The black line represent the posterior mean of μ_θ and the yellow lines, its 95% CI. Using this plot, we can show how aggregate level covariate effect on the higher state is underestimated. We highlight that distribution of that parameter for customers in K_2 is approximately centered in 0, whereas the same distribution for customers in K_3 is approximately centered in 0.8. However, aggregate results show that the effect is estimated between 0.29 and 0.45.

Although, we showed similar results in Experiment 1.1 for parameters τ_{21} and τ_{22}, the causes that generate both bimodal distributions are different. In Experiment 1.1, states are identified as the same for all customers and the bias on aggregate results is caused for different values of individual level estimated parameters between two customer segments (EF1). On the other hand, in Experiment 4.6 the bias is mainly caused by the fact that states capture different behavior for customers in K_2 and K_3 (EF2), due to duplicated states for customers in K_2, which increases the differences on individual level parameters between customers of different segments.

In summary, we showed that when there is homogeneity on α^0, EF1 is present: states are identified as the same for all customers, individual level parameters are correctly recovered, covari-

[10] Described in Section 4.2.3. States are not fixed for all customers, and duplicated states are estimated for customers in K_2.

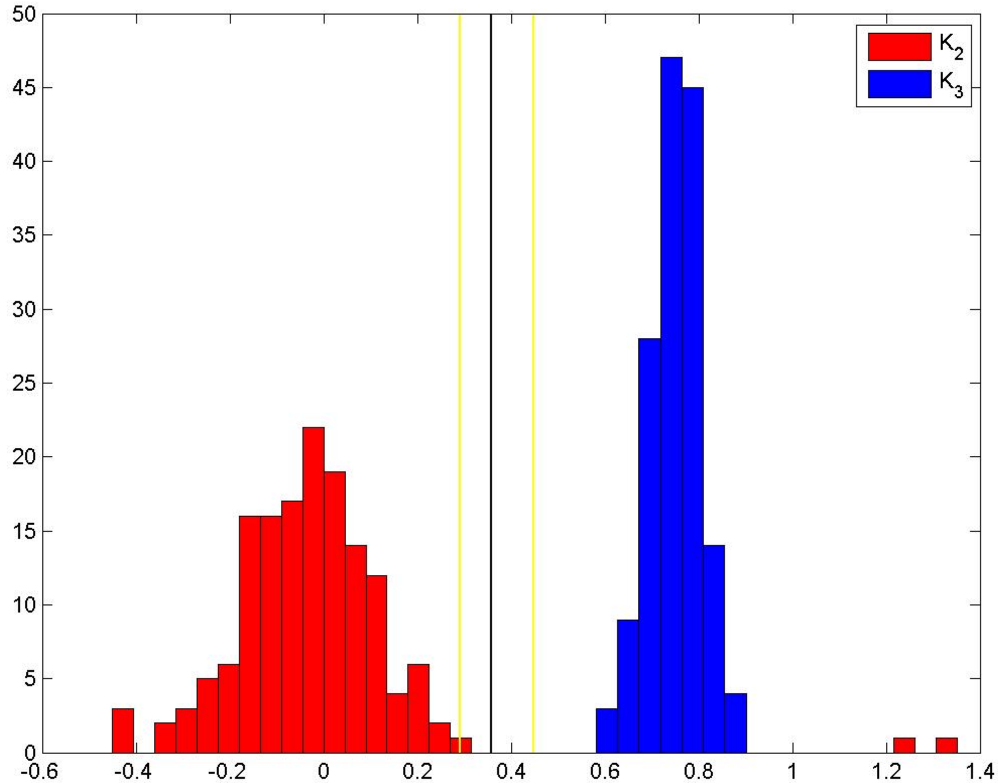


Figure 4.7: Experiment 4.6: Heterogeneity in $\boldsymbol{\rho}_3$

ate effects are mostly recovered (except for some cases in which the simulated values are slightly off the CI), but aggregate result are biased. In addition, when there is heterogeneity on α^0, if covariate effects are homogeneous and different between states, those parameters allow the correct identification of states across customers and EF1 is present as well. However, when α^0 is heterogeneous but covariate effects are similar across states EF2 is observed: states are identified differently across customers, customers with fewer simulated states identify two different states and a third that is a duplicated state of the others, and aggregate results are biased not only for transition matrix but for conditional probabilities as well. In addition, when there is heterogeneity on all parameters, we observe EF2, and covariate effects are also biased at aggregate level.

Chapter 5

Proposed Model

We identified in the previous chapter, a bias on the recovery of the parameters at the aggregate level when the heterogeneity in the number of the states of a HMM is not incorporated in the estimated model. When there is no heterogeneity on the intercept of the conditional probabilities and the mixture of customers with different states is balanced (50%-50%), allowing for enough heterogeneity on the parameters can help to recover the parameters at the individual level, but not at the aggregate level. This is because the model can not distinguish between the two latent class with different number of states. Therefore, a model which can capture the dynamics and behavior at the aggregate level considering the segments of customers with different number of states is needed. The following models attempt to improve the results on the estimation of the parameters when there is heterogeneity on the states.

5.1 Mixture of Gaussians as a prior

When using a Normal distribution as prior distribution of the parameters, the report of results at the aggregate level usually includes only the results of the mean μ_θ and the covariance matrix Σ_θ which leads to wrong conclusions of customer behavior. Moreover, computing the transition matrix using μ_θ can lead to mix the probability of transitioning from the middle state to the higher state between the customers who make those transitions ($q_{23} = 0.3$) and the customers who do not ($q_{23} = 0$) (see Figure 4.2).

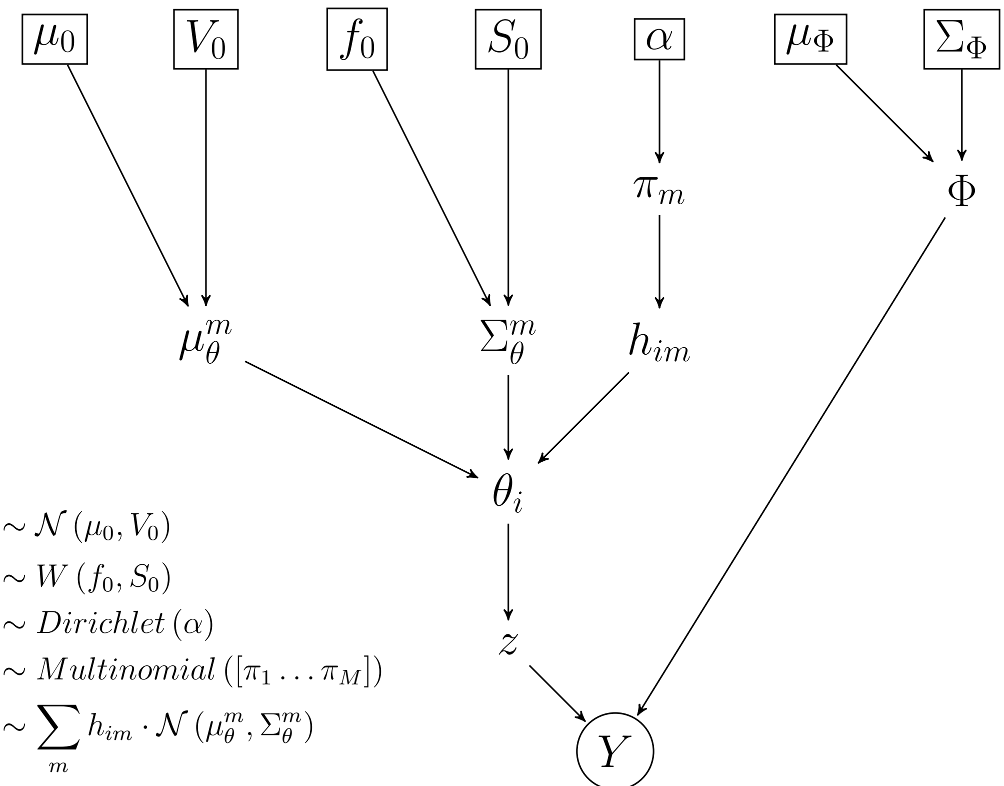
In this context, we suggest an approach that uses a mixture of Gaussians as the prior of the individual level parameters (MOGP). The model is defined using the same notation as in Section 3.1.

It is important to highlight that this model does not capture directly different number of states for different customers, but allows detecting if there is a segment of customers with low probability of reaching a certain state. Then it does not mix the results of different segments at the aggregate level but shows them as different components of the mixture of Gaussians instead.

5.1.1 Model specification

Let consider a fixed number of states S. Let θ_i be the individual-level parameters and Φ the population-level parameters of the HMM described in Section 3.1. Let $M \in \mathbb{N}_+$ be the number of components of the mixture of Gaussians. Let μ_θ^m and Σ_θ^m be mean and covariance matrix of the m'th component of the heterogeneity distribution on θ_i. We define a latent variable h_{im}, where $h_{im} = 1$ if parameter θ_i is drawn from component m of the mixture, and $h_{im} = 0$ otherwise. Let π_m be the probability that a vector is drawn from component m, where $\sum_{m=1}^{M} \pi_m = 1$ and $\pi_m \geq 0$.

We modify the model only in terms of the prior distribution of θ_i. Using description in Appendix B, prior and full distribution can be written as follows [Koop, 2003]:



$$
\begin{aligned}
\mu_\theta^m &\sim \mathcal{N}(\mu_0, V_0) \\
(\Sigma_\theta^m)^{-1} &\sim W(f_0, S_0) \\
[\pi_1 \dots \pi_M] &\sim Dirichlet(\alpha) \\
[h_{i1} \dots h_{iM}] &\sim Multinomial([\pi_1 \dots \pi_M]) \\
\theta_i &\sim \sum_m h_{im} \cdot \mathcal{N}(\mu_\theta^m, \Sigma_\theta^m) \\
\Phi &\sim \mathcal{N}(\mu_\Phi, \Sigma_\Phi)
\end{aligned}
$$

□ : Prior hyperparameters
○ : Observable Data

Figure 5.1: Mixture of Gaussians Hierarchical Heterogeneous Model

5.1.2 Results

Basic model: Experiment 1.1

We tested this model using the simulated data from Experiment 1.1.

For MOGP, the number of the states that fits the data best is 3 (see Appendix G.3). The number of components of the Mixture of Gaussians used by that model is 2 (see Appendix G.2).

We compare this model with the best model using a fix number of states, which is the 3-State model (3HMM). As we show in Table 5.1, according to all criteria MOGP is better model than 3HMM, given that likelihood function is the same for both models but a bimodal prior allows for more heterogeneity and generates a better estimation at the individual level.

Model	**LMD**	**DIC**	**MSC**
3HMM	-30750.53	61699.86	72471.99
MOGP	**-30317.22**	**60671.90**	**70311.26**

Table 5.1: Model Comparison

We report in Table 5.2 the posterior mean and CI of Q for customers in K_2 and K_3 and for population mean of each component of the prior distribution μ_θ^1 and μ_θ^1. In addition, we report in Table 5.3 the posterior mean and CI of p_{ks}.

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.29	0.01	0.70	0.30	NA
		[0.68 0.71]	[0.28 0.31]	[0.01 0.01]			
$\mathbf{K_2}$	**2**	0.31	0.68	0.01	0.30	0.70	NA
		[0.29 0.32]	[0.67 0.70]	[0.01 0.01]			
	3	0.31	0.18	0.50	NA	NA	NA
		[0.27 0.36]	[0.13 0.25]	[0.44 0.56]			
	1	0.65	0.22	0.13	0.70	0.20	0.10
		[0.61 0.69]	[0.19 0.25]	[0.11 0.15]			
$\mathbf{K_3}$	**2**	0.19	0.51	0.30	0.20	0.50	0.30
		[0.16 0.21]	[0.48 0.55]	[0.27 0.34]			
	3	0.12	0.21	0.68	0.10	0.20	0.70
		[0.10 0.14]	[0.18 0.24]	[0.64 0.71]			
	1	0.70	0.29	0.01			
		[0.68 0.71]	[0.28 0.31]	[0.01 0.01]			
Component 1	**2**	0.31	0.68	0.01		NA	
(51%)		[0.29 0.32]	[0.66 0.70]	[0.01 0.01]			
	3	0.32	0.18	0.50			
		[0.27 0.36]	[0.13 0.25]	[0.44 0.56]			
	1	0.65	0.22	0.14			
		[0.60 0.68]	[0.19 0.25]	[0.12 0.16]			
Component 2	**2**	0.18	0.50	0.31		NA	
(49%)		[0.15 0.21]	[0.46 0.55]	[0.28 0.35]			
	3	0.12	0.21	0.68			
		[0.10 0.14]	[0.18 0.24]	[0.64 0.71]			

Table 5.2: MOGP: Experiment 1.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table 5.3: MOGP: Experiment 1.1 conditional probabilities posterior mean and 95% CI

The main advantage of this model is that enables to represent correctly the behavior of customers at the aggregate level. Although this model does not allow for heterogeneity on the number of states, it can be seen that a subset of customers does not reach the higher state. The first component of the mixture of Gaussians represents the behavior of the customers in K_2 and the second component represents the behavior of the customers in K_3. Using this model, the transitions matrices between states can be interpreted correctly at the aggregate level.

Mixture of customers: Experiment 2.3

In the simulation shown in the previous section, the mixture of customers with 2 and 3 states is balanced. In Section 4.2.2 we show that when the proportion of customers with fewer states increases, the bias at the aggregate level increases as well. We showed that parameters are correctly recovered at the individual level but not at the aggregate level. Specifically, the probability of going from state 2 to the higher state decreases as the proportion of customers with two states increases.

We test the performance of this model using data from Experiment 2.3 as an example of different mixture of customers (80% customers in K_2 and 20% customers in K_3). We report in Table 5.4 the model selection criteria LML, DIC, and MSC.

Model	**LML**	**DIC**	**MSC**
3HMM	-30345.92	60818.21	71440.15
MOGP	-30307.93	60668.66	70317.84

Table 5.4: Model Comparison

As it can be observed in Table 5.5, the main advantage of this model is that the two different aggregate components allow to identify two segments, each of them representing a different conditional behavior.

		Posterior Mean			**Simulated**		
		1	**2**	**3**	**1**	**2**	**3**
K_2	**1**	0.70 [0.68 0.71]	0.29 [0.28 0.31]	0.01 [0.01 0.01]	0.70	0.30	NA
	2	0.31 [0.29 0.32]	0.68 [0.67 0.70]	0.01 [0.01 0.01]	0.30	0.70	NA
	3	0.41 [0.37 0.46]	0.17 [0.13 0.22]	0.42 [0.34 0.49]	NA	NA	NA
K_3	**1**	0.65 [0.61 0.68]	0.22 [0.19 0.25]	0.13 [0.11 0.15]	0.70	0.20	0.10
	2	0.18 [0.15 0.22]	0.51 [0.47 0.55]	0.30 [0.27 0.34]	0.20	0.50	0.30
	3	0.12 [0.10 0.14]	0.21 [0.18 0.24]	0.68 [0.64 0.71]	0.10	0.20	0.70
Component 1 (20%)	**1**	0.65 [0.60 0.68]	0.22 [0.19 0.25]	0.14 [0.12 0.16]		NA	
	2	0.18 [0.15 0.21]	0.51 [0.46 0.55]	0.31 [0.28 0.36]			
	3	0.11 [0.09 0.14]	0.21 [0.18 0.24]	0.68 [0.64 0.71]			
Component 2 (80%)	**1**	0.70 [0.68 0.71]	0.29 [0.28 0.31]	0.01 [0.01 0.01]		NA	
	2	0.31 [0.29 0.32]	0.68 [0.67 0.70]	0.01 [0.01 0.01]			
	3	0.41 [0.37 0.46]	0.17 [0.13 0.22]	0.42 [0.34 0.49]			

Table 5.5: MOGP Experiment 2.3 transition matrix posterior mean and 95% CI

MOGP identifies two segments of customers, in an approximate proportion of 80%-20%, where customers in the 80% segment make transitions mostly among the first 2 states and customers of the other segment can move along all states. In terms of conditional distribution probabilities p_{ks}, states are correctly identified (see Table 5.6). Moreover, states are fixed for all customers given that conditional probabilities are homogeneous.

	Posterior Mean			**Simulated**		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table 5.6: MOGP: Experiment 2.3 conditional probabilities posterior mean and 95% CI

Finally, and using results from this model on the data of experiments 2.1 and 2.4, the probability of moving from state 2 to 3 does not have a significant decrease while the proportion of customers with two states increases (see Figures 5.2 and 5.3).

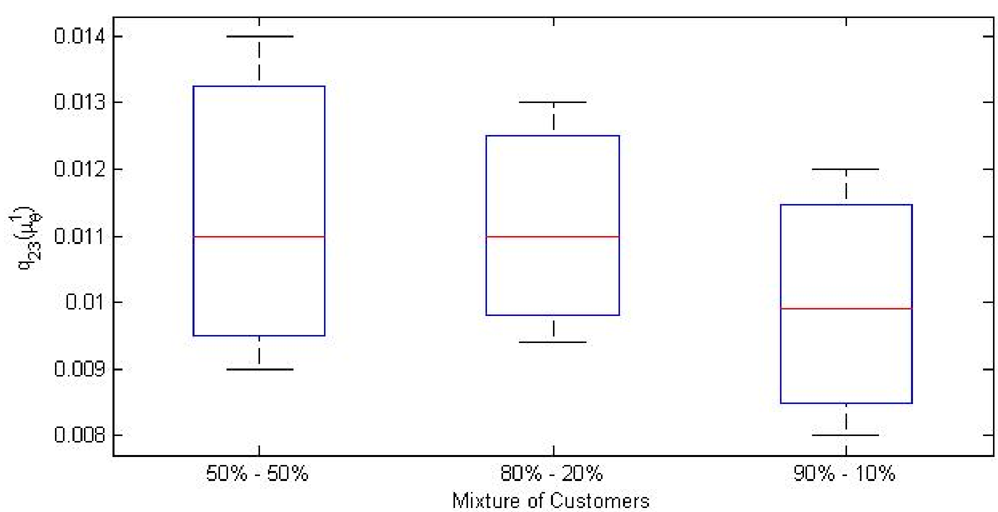


Figure 5.2: Quantiles of $q_{23}(\mu_\theta^1)$ by mixtures of customers, for Component 1 of MOGP)

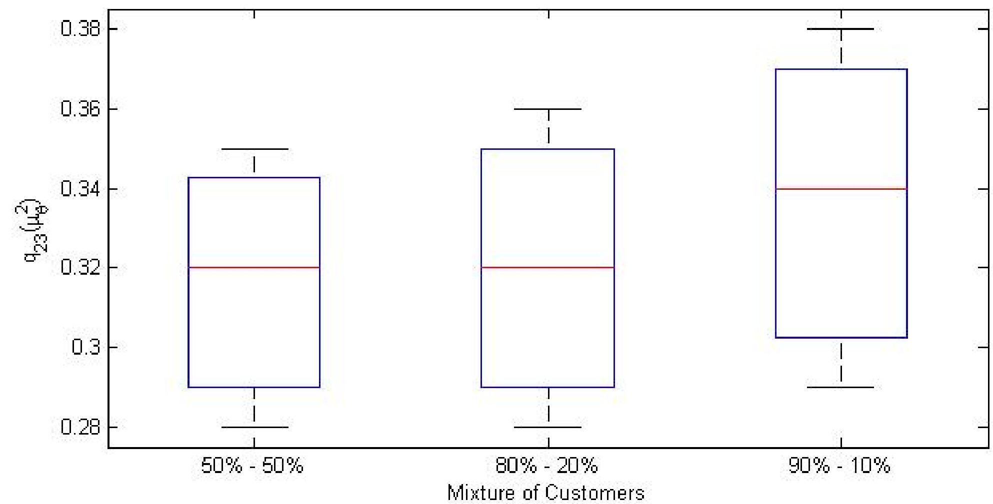


Figure 5.3: Quantiles of $q_{23}(\mu_\theta^2)$ by mixtures of customers, for Component 2 of MOGP)

Heterogeneity in the conditional probabilities: Experiment 3.1

In Section 4.2.3, we showed that introducing heterogeneity on the conditional probabilities generates duplicated states and these states do not identify the same conditional behavior for all customers, which increases the bias at aggregate level (EF2). We applied the MOGP model to simulated data from Experiment 3.1. LML, DIC and MSC indicates MOGP as the best model. We report in Tables

Model	**LMD**	**DIC**	**MSC**
3HMM	-33804.60	68316.14	85675.77
MOGP	-30636.06	61851.26	76145.69

Table 5.7: Model Comparison

5.8 and 5.9 posterior mean and CI for Q and p_{ks}.

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.06	0.23	0.70	0.30	NA
		[0.69 0.72]	[0.05 0.08]	[0.21 0.26]			
$\mathbf{K_2}$	**2**	0.32	0.22	0.46	0.30	0.70	NA
		[0.25 0.37]	[0.18 0.29]	[0.42 0.49]			
	3	0.32	0.06	0.62	NA	NA	NA
		[0.30 0.35]	[0.03 0.11]	[0.56 0.66]			
	1	0.69	0.19	0.12	0.70	0.20	0.10
		[0.67 0.71]	[0.18 0.21]	[0.11 0.13]			
$\mathbf{K_3}$	**2**	0.21	0.52	0.27	0.20	0.50	0.30
		[0.19 0.23]	[0.50 0.55]	[0.25 0.29]			
	3	0.12	0.19	0.69	0.10	0.20	0.70
		[0.10 0.13]	[0.18 0.21]	[0.67 0.71]			
	1	0.69	0.20	0.12			
		[0.67 0.71]	[0.18 0.22]	[0.11 0.13]			
Component 1	**2**	0.20	0.53	0.27		NA	
(50%)		[0.18 0.23]	[0.50 0.55]	[0.25 0.29]			
	3	0.12	0.20	0.69			
		[0.10 0.13]	[0.18 0.21]	[0.67 0.71]			
	1	0.70	0.06	0.23			
		[0.68 0.72]	[0.05 0.08]	[0.21 0.26]			
Component 2	**2**	0.32	0.22	0.46		NA	
(50%)		[0.25 0.37]	[0.18 0.29]	[0.42 0.50]			
	3	0.32	0.06	0.62			
		[0.30 0.35]	[0.03 0.11]	[0.56 0.66]			

Table 5.8: MOGP Experiment 3.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
Component 1	0.10	0.49	0.90	0.10	0.50	0.90
(50%)	[0.10 0.10]	[0.48 0.51]	[0.89 0.91]			
Component 2	0.10	0.48	0.52	0.10	0.50	
(50%)	[0.10 0.10]	[0.46 0.50]	[0.51 0.54]			

Table 5.9: MOGP: Experiment 3.1 conditional probabilities posterior mean and 95% CI

The results show that the MOGP model does not solve the problem detected for 3HMM when assuming the same number of states for all customers. EF2 is present in this model as well, and the posterior mean of conditional probabilities show that component 1 identifies two states (2nd and 3rd) as the middle state simulated (0.5). As it is the case for 3HMM, this effect is caused in MOGP model because we assume that all customers have the same number of states. Although this model allows to capture aggregate information of different segments by identifying them on different components of the distribution, it does not account for heterogeneity in the number of states. However, this model does not average aggregate results from different segments (which is the case for 3HMM).

In summary, the MOGP allows to identify the segments on the population, however it does not avoid the duplicity on the states for customers with fewer states than the estimated model.

Covariates: Experiments 4.x

In Section 4.2.4 we showed that EF2 is present when all parameters are heterogeneous or when the intercept of conditional probabilities α^0 is heterogeneous and the covariate effects are homogeneous and similar across states. We estimate the MOGP model for Experiment 4.6 and found consistent results. First, as in the case of previous experiments, MOGP is a better model according to DIC and MSC measures (see Table 5.10).

Model	LML	DIC	MSC
3HMM	-30483.06	62163.14	83787.11
MOGP	-30574.38	62099.28	82964.79

Table 5.10: Model Comparison

We show in Table 5.11 and Table 5.12 that MOGP estimates duplicated states for customers in K_2 which is captured for the second component of the mixture of Gaussians distribution. Component 1 captures correctly the parameters of customers in K_3.

		Posterior Mean[a] 1	Posterior Mean[a] 2	Posterior Mean[a] 3	Simulated 1	Simulated 2	Simulated 3
$\mathbf{K_2}$	**1**	0.69 [0.68 0.71]	0.07 [0.06 0.08]	0.24 [0.22 0.25]	0.70	0.30	NA
	2	0.41 [0.38 0.44]	0.19 [0.17 0.22]	0.40 [0.36 0.43]	0.30	0.70	NA
	3	0.29 [0.28 0.31]	0.10 [0.09 0.12]	0.61 [0.58 0.63]	NA	NA	NA
$\mathbf{K_3}$	**1**	0.69 [0.68 0.71]	0.21 [0.19 0.22]	0.10 [0.09 0.11]	0.70	0.20	0.10
	2	0.21 [0.20 0.23]	0.53 [0.51 0.55]	0.26 [0.24 0.28]	0.20	0.50	0.30
	3	0.11 [0.11 0.12]	0.21 [0.19 0.23]	0.68 [0.66 0.70]	0.10	0.20	0.70
Component 1 (51%)	**1**	0.69 [0.67 0.71]	0.21 [0.19 0.22]	0.10 [0.09 0.11]		NA	
	2	0.21 [0.20 0.23]	0.53 [0.51 0.56]	0.26 [0.23 0.28]			
	3	0.11 [0.11 0.12]	0.21 [0.19 0.23]	0.68 [0.65 0.70]			
Component 2 (49%)	**1**	0.69 [0.68 0.71]	0.07 [0.06 0.08]	0.24 [0.22 0.25]		NA	
	2	0.41 [0.38 0.44]	0.19 [0.16 0.22]	0.40 [0.36 0.44]			
	3	0.30 [0.28 0.31]	0.10 [0.09 0.12]	0.60 [0.58 0.63]			

Table 5.11: MOGP Experiment 4.6 transition matrix posterior mean and 95% CI

[a] To report these values, we use only the intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
Component 1	0.10	0.50	0.90	0.10	0.50	0.90
(51%)	[0.10 0.11]	[0.49 0.52]	[0.90 0.91]			
Component 2	0.10	0.48	0.52	0.10	0.50	
(49%)	[0.10 0.11]	[0.47 0.50]	[0.50 0.53]			

Table 5.12: MOGP: Experiment 4.6 Latent Class 3 conditional probabilities posterior mean and 95% CI

[a] To report these values, we use only the intercept, i.e., we assume that covariates $X_{kt} = 0$.

To illustrate the effect of duplicated states on the estimation of covariate effects we report the parameter estimates by component in Table 5.13 and Table 5.14.

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean Component 1**	**95% CI**
Transition Matrix Parameters					
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.81	[0.73 0.89]
High transition threshold state 1	τ_{k12}	.	0.30	0.32	[0.24 0.40]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-1.31	[-1.40 -1.22]
High transition threshold state 2	τ_{k22}	.	0.80	0.87	[0.81 0.92]
Low transition threshold state 3	τ_{k31}	.	-2.20	-2.06	[-2.14 -1.98]
High transition threshold state 3	τ_{k32}	.	0.30	0.27	[0.21 0.34]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.50	0.47	[0.39 0.56]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.00	0.00	-0.05	[-0.14 0.02]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.80	0.79	[0.71 0.88]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.17	[-2.21 -2.13]
Intercept of state 2	α_2^0	0.79	0.79	0.78	[0.76 0.80]
Intercept of state 3	α_3^0	.	0.79	0.80	[0.78 0.83]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.50	0.48	[0.45 0.50]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.10	0.10	0.10	[0.08 0.12]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.10	0.09	[0.06 0.12]

Table 5.13: MOGP Experiment 4.6: Component 1 parameter estimates

Parameter		Simulated K_2	Simulated K_3	Posterior Mean Component 2	95% CI
Transition Matrix Parameters					
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.81	[0.74 0.89]
High transition threshold state 1	τ_{k12}	.	0.30	-1.05	[-1.17 -0.93]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-0.37	[-0.49 -0.24]
High transition threshold state 2	τ_{k22}	.	0.80	-0.26	[-0.41 -0.12]
Low transition threshold state 3	τ_{k31}	.	-2.20	-0.87	[-0.95 -0.78]
High transition threshold state 3	τ_{k32}	.	0.30	-0.80	[-0.94 -0.68]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.50	0.43	[0.35 0.50]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.00	0.00	0.13	[0.00 0.24]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.80	-0.10	[-0.18 -0.02]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.16	[-2.20 -2.12]
Intercept of state 2	α_2^0	0.79	0.79	0.73	[0.71 0.76]
Intercept of state 3	α_3^0	.	0.79	-1.90	[-1.99 -1.81]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.50	0.50	[0.48 0.53]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.10	0.10	0.11	[0.07 0.15]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.10	0.09	[0.07 0.11]

Table 5.14: MOGP Experiment 4.6: Component 2 parameter estimates

Component 1 identifies correctly the behavior of customers with 3 states, including all covariate effects, given that we estimate the model with the correct number of states for those customers. On the other hand, component 2 represents the behavior of customers with two states. However, state 2 and 3 captures the behavior of the true middle state of $p_{ks} = 0.5$. Therefore, covariate effects of states 2 and 3 attempt to capture independently the same effect. Specifically, the true value for $\boldsymbol{\rho}_2$ is 0, but the estimated values of $\boldsymbol{\rho}_2$ and $\boldsymbol{\rho}_3$ are 0.13 and -0.10. In this scenario, heterogeneity in $\boldsymbol{\rho}_2$ is captured by two different states.

In summary, using a model for detecting the aggregate behavior of different segments, as MOGP, allows for a correct interpretation of the results at the aggregate level by avoiding to average parameter estimates from different segments. This model recovers correctly the aggregate behavior when the states are fixed and identify the same conditional behavior for all customers. However, it does not eliminate the effect of duplicated states (EF2) when conditional probabilities parameters are heterogeneous. Given this effect, covariate effect ρ_2 is biased. We introduce heterogeneity specifically on the number of states in the following section to capture aggregate behavior correctly.

5.2 Latent class of Hidden Markov Models

The model we described in Section 5.1 does not account for heterogeneity in the number of states of the HMM. In this section, we proposed a model that attempts to estimate the number of states of the HMM for each customer.

Discrete heterogeneity is usually captured by using a Latent class model, in which there are group of customers (classes) with similar characteristics within the group but different across groups. We define a latent class model that considers a HMM of different number of states on each class.

One important fact is that in Latent class models usually the model is the same for all classes, but parameters are different and capture heterogeneity across customers within the same model.

The previous model we described in Section 3.1 includes continuous heterogeneity within a HMM, using a hierarchical Bayesian model. In this model instead, the latent class component captures structural heterogeneity, specifically, classes where each class contains a HMM with a different number of states, i.e., dynamics across different classes not only differ on the values of their parameters but are also structurally different.

5.2.1 Model specification

Accordingly, we formalized the latent class of hidden Markov models (LCHMM) as follows.

Let consider latent class $m = 1 \dots M$ a HMM with S_m number of states. For simplicity, we assume $S_m = m$.

Following the notation used in Section 3.1, let θ_i^m be the individual-level parameters for HMM m and customer i, and Φ^m be the population parameters for HMM m. In addition, we define $\Theta^m = \{\theta_i^m\}_{i=1}^K$ as the set of individual parameters θ_i^m for HMM m.

We define $\pi_m \geq 0$ as the membership probability of a random customer belongs to model m ($\sum\limits_{m=1}^{M} \pi_m = 1$). We parametrize π_m using a multivariate logistic transformation over $\alpha \in \mathbb{R}^{M-1}$ as follows:

$$
\begin{aligned}
\pi_m &= \frac{exp\left(\alpha_m\right)}{\sum\limits_{\tilde{m}=1}^{M-1} exp\left(\alpha_{\tilde{m}}\right)} \qquad\qquad m = 1, \dots, M-1 \\
\pi_M &= 1 - \sum_{m=1}^{M-1} \pi_m
\end{aligned} \tag{5.1}
$$

Let $\Theta = \{\Theta^m\}_{m=1}^M$ be the set of individual level parameters and $\Phi = \left\{\{\Phi^m\}_{m=1}^M, \alpha\right\}$ the population parameters.

Let $z_{i,t}^m$ be the state of customer i on period t on the HMM m. Note that each customer has parameters for M HMMs, therefore each customer has a path of hidden states through periods for each HMM m.

The likelihood function at the individual level $L_i(\theta_i, \Phi|Y_i)$ is computed by conditioning on the latent class model $m = 1, \ldots, M$.

$$L_i(\theta_i, \Phi|Y_i) = \sum_{m=1}^{M} \pi_m \, L_i(\theta_i^m, \Phi_m|Y_i, S = S_m) \tag{5.2}$$

where $L_i(\cdot, \cdot|Y_i, n = n_m)$ is the individual likelihood function of the HMM with S_m states for customer i described in Equation 3.13. Therefore, the likelihood of the model is computed using the product of the likelihood of each customer as follows:

$$\begin{aligned} L(\Theta, \Phi|Y) &= \prod_{i \in K} L_i(\Theta_i, \Phi|Y_i) \\ &= \prod_{i \in K} \left(\sum_{m=1}^{M} \pi_m \, L_i(\theta_i^m, \Phi_m|Y_i, n = n_m) \right) \end{aligned} \tag{5.3}$$

Finally, we estimate this model using a Bayesian hierarchical approach to account for unobserved heterogeneity at the individual level. The MCMC algorithm we implemented is described in Appendix H.

5.2.2 Results

Basic model: Experiment 1.1

We tested this model using the simulated data from Experiment 1.1. For LCHMM, three Latent class were used ($m = 1, 2, 3$): single state HMM, 2-state HMM and 3-state HMM. This model is compared with the best model using a fix number of states, which is the 3-State model (3HMM). In this case, LML and DIC were used to choose the best model, given that LCHMM is not a HMM itself but a latent class model instead, therefore MSC does not apply as a valid information criterion.

Model	**LML**	**DIC**
3HMM	**-30750.53**	**61699.86**
LCHMM	-30882.14	61761.18

Table 5.15: Model Comparison

Although LCHMM is not a better model than 3HMM (according to LML and DIC), we will show next that LCHMM provides better results at aggregate level, in terms of posterior means of transition matrices and conditional probabilities.

We report the posterior mean and the CI for probability of class membership π_m in Table 5.16. In addition, we report in Tables 5.17 and 5.18 the posterior means and the CI for Q and p_{ks}, for the HMM of each class.

Latent Class	Membership Probability π_m
1-state HMM	0.01 [0.00 0.02]
2-state HMM	0.50 [0.45 0.56]
3-state HMM	0.49 [0.43 0.54]

Table 5.16: π_m Results for LCHMM model *Results format: (2.5%, 50%, 97.5%)*

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
Latent Class 2 (50%)	**1**	0.71 [0.70 0.72]	0.29 [0.28 0.30]		0.70	0.30	NA
	2	0.32 [0.31 0.32]	0.68 [0.68 0.69]		0.30	0.70	NA
	3				NA	NA	NA
Latent Class 3 (49%)	**1**	0.69 [0.68 0.70]	0.21 [0.20 0.22]	0.10 [0.10 0.11]	0.70	0.20	0.10
	2	0.20 [0.19 0.20]	0.51 [0.51 0.52]	0.29 [0.28 0.30]	0.20	0.50	0.30
	3	0.10 [0.10 0.11]	0.21 [0.21 0.22]	0.68 [0.67 0.69]	0.10	0.20	0.70

Table 5.17: LCHMM Experiment 1.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
Latent Class 1 (1%)	0.50 [0.12 0.87]				NA	
Latent Class 2 (50%)	0.10 [0.10 0.10]	0.50 [0.50 0.50]		0.10	0.50	
Latent Class 3 (49%)	0.10 [0.10 0.10]	0.50 [0.50 0.50]	0.90 [0.90 0.90]	0.10	0.50	0.90

Table 5.18: LCHMM: Experiment 1.1 conditional probabilities posterior mean and 95% CI

As expected, customers have either 2 or 3 states. The probability that a customer has no dynamics on their hidden state, i.e., that he has only 1 state, is 0.01. Moreover, a random customer has a 0.5 probability of having 2 hidden states and 0.49 probability of having 3 hidden states.

Even though the states are not fixed across classes (we do not impose in the model a relationship between states of different HMMs), the second class represents correctly the 2 lower simulated states for customers in K_2 and the third class identifies all three simulated states for customers in K_3 (see p_{ks} posterior mean and CI in Table 5.18). In addition, latent class 2 recovers correctly the transition matrix of customers in K_2 and latent class 3 recovers the transition matrix of customers in K_3 as well.

One of the advantages of using this model is to be able to identify the heterogeneity in the number of states. Using the results of this model we compute the individual posterior membership probability to test if the number of hidden states simulated for each customer are correctly identified. We compute π_m using the posterior parameter estimates at the individual level. The individual posterior membership probability can be computed using the same procedures as in any latent class model.

Let $x_k \in \{1 \dots M\}$ be the latent class of customer k. The membership probability that a random customer belongs to class m, π_m is described in Equation 5.1. We define: $\bar{\theta}_k^m$ as the posterior mean of θ_k^m, $\bar{\Phi}^m$ as the posterior mean of Φ^m, and $\bar{\alpha}$ as the posterior mean of α. The posterior individual probability (which we denote as π_{km}) can be computed proportional to the likelihood of the latent class m, times the posterior aggregate membership probability $\pi_m(\bar{\alpha})$ as follows:

$$\begin{aligned} \pi_{km} &= \mathbb{P}(x_k = m | Y_k) \\ &= \frac{\pi_m(\bar{\alpha})\, L_k(\bar{\theta}_k^m, \bar{\Phi}^m | Y_k, n = n_m)}{\sum_{m'=1}^{M} \pi_{m'}(\bar{\alpha})\, L_k(\bar{\theta}_k^{m'}, \bar{\Phi}^{m'} | Y_k, n = n_{m'})} \\ &\propto \pi_m(\bar{\alpha})\, L_k(\bar{\theta}_k^m, \bar{\Phi}^m | Y_k, n = n_m) \end{aligned} \tag{5.4}$$

Using the membership probability π_{km}, we assign customers to the latent class with the highest probability, i.e., if $h_i^* \in \{1 \dots M\}$ is the individual posterior latent class, then $h_k^* = \arg\max_m \pi_{km}$.

We report in Table 5.19 the number of customers of sets K_2 and K_3 assigned to each latent class using π_{km}. In this table, it can be seen that most of the customers are correctly assigned to the true number of states. The single customer that was incorrectly assigned did not reach to the higher state in any simulation period despite having a positive probability of making that move; therefore, as expected, it was identified as a customer with two hidden states.

Number of states

	Predicted		
True	**1**	**2**	**3**
1	0	0	0
2	0	150	0
3	0	1	149

Table 5.19: Customers assigned to each latent class using posterior membership probabilities

Mixture of customers: Experiment 2.3

As shown for model MOGP in Section 5.1.2, we estimate the LCHMM with data from Experiment 2.3 (80% customers in K_2 and 20% customers in K_3) to test if this model identifies the two segments at the aggregate level. Additionally, we test if the probability of going from state 2 to the higher state decreases as the proportion of customers with two states increases when using LCHMM.

We report in Table 5.20 the model selection criteria LML and DIC. The results are not conclusive about which model is better.

Model	LML	DIC
3HMM	**-30345.92**	60818.21
LCHMM	-30406.51	**60811.52**

Table 5.20: Model Comparison

We report in Tables 5.21 and 5.22 posterior means and CI for Q and p_{ks}, for the HMM of each class.

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
Latent Class 2 (80%)	**1**	0.70 [0.70 0.71]	0.30 [0.29 0.30]		0.70	0.30	NA
	2	0.31 [0.31 0.32]	0.69 [0.68 0.69]		0.30	0.70	NA
	3				NA	NA	NA
Latent Class 3 (19%)	**1**	0.68 [0.67 0.69]	0.22 [0.21 0.23]	0.10 [0.10 0.11]	0.70	0.20	0.10
	2	0.20 [0.20 0.21]	0.50 [0.49 0.51]	0.30 [0.29 0.31]	0.20	0.50	0.30
	3	0.10 [0.10 0.11]	0.21 [0.20 0.22]	0.69 [0.68 0.70]	0.10	0.20	0.70

Table 5.21: LCHMM Experiment 2.3 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
Latent Class 1 (1%)	0.49 [0.12 0.87]				NA	
Latent Class 2 (80%)	0.10 [0.10 0.10]	0.50 [0.50 0.50]		0.10	0.50	
Latent Class 3 (19%)	0.10 [0.10 0.10]	0.50 [0.50 0.51]	0.90 [0.90 0.90]	0.10	0.50	0.90

Table 5.22: LCHMM: Experiment 2.3 conditional probabilities posterior mean and 95% CI

Table 5.21 shows that the model identifies two segments of customers, in a 80%-20% proportion where customers in the 80% segment have two states whereas customers in the 20% segment have three states. Although according to LML and DIC this model is not clearly better than 3HMM and MOGP, LCHMM explicitly identifies a segment with 2 states, by capturing heterogeneity on the number of states. Additionally, Table 5.25 shows the correct recovery of p_{ks} for each state on all relevant latent classes (2 and 3).

In addition, the probability of going from state 2 to 3 does not have a significant decrease while the proportion of customers with two states grows as shown in Figure 5.4.

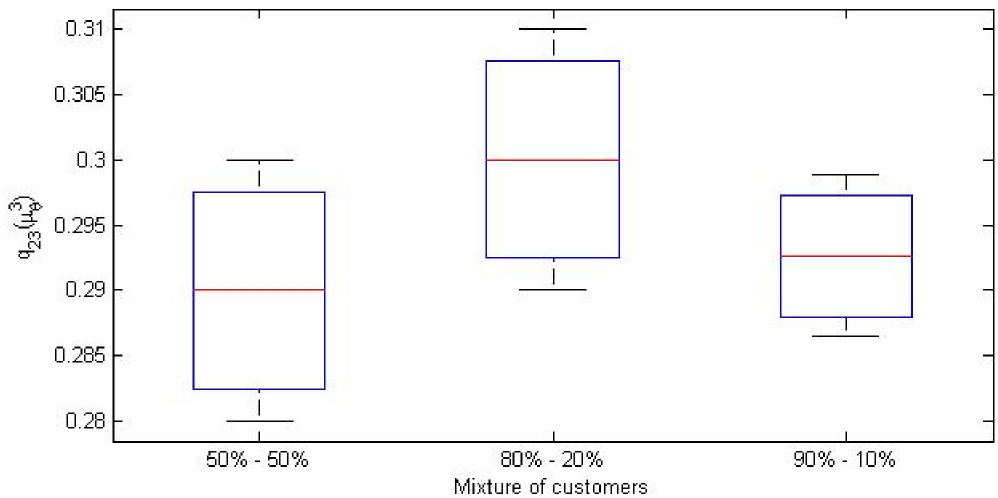


Figure 5.4: Quantiles of $q_{23}(\mu_\theta^3)$ by mixtures of customers, for 3 states class of LCHMM

Heterogeneity in the conditional probabilities: Experiment 3.1

Aggregate parameter estimates of models 3HMM and MOGP on this experiment are considerably biased as shown in Sections 4.2.3 and 5.1.2, as a result of duplicated states and averaging over customers from different segments. In this context, we test LCHMM using simulated data from experiment 3.1, where conditional probabilities are heterogeneous across customers.

We show in Table 5.23 that LCHMM is the best model.

Model	**LML**	**DIC**
3HMM	-33804.60	68316.14
LCHMM	**-30973.70**	**62337.79**

Table 5.23: Model Comparison

We report posterior mean and CI of Q and p_{ks} in Tables 5.24 and 5.25 .

Results show that LCHMM does not estimate duplicated states. Although conditional probabilities are heterogeneous, latent classes 2 and 3 recover correctly those probabilities for each segment. Conditional behavior of customers in K_2 is captured by the latent class with 2 hidden states, and conditional behavior of customers in K_3 is captured by the latent class with 3 hidden states. In addition, results for transition matrix show that dynamics are correctly recovered as well. In contrast with aggregate results of 3HMM and MOGP, results of transition matrix and conditional probabilities using population parameter μ_θ capture correctly the behavior of all segments of customers. Therefore, using a model with latent class of different number of states allows for delivering accurate interpretations of the group of customers.

		Posterior Mean			Simulated		
		1	2	3	1	2	3
	1	0.71	0.29		0.70	0.30	NA
		[0.71 0.71]	[0.29 0.29]				
Latent Class 2	2	0.31	0.69		0.30	0.70	NA
(50%)		[0.30 0.31]	[0.69 0.70]				
	3				NA	NA	NA
	1	0.70	0.20	0.10	0.70	0.20	0.10
		[0.70 0.70]	[0.19 0.20]	[0.10 0.11]			
Latent Class 3	2	0.20	0.50	0.30	0.20	0.50	0.30
(49%)		[0.20 0.21]	[0.49 0.50]	[0.30 0.31]			
	3	0.10	0.20	0.70	0.10	0.20	0.70
		[0.10 0.10]	[0.20 0.21]	[0.70 0.71]			

Table 5.24: LCHMM Experiment 3.1 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	2	3	1	2	3
Latent Class 1	0.50				NA	
(1%)	[0.50 0.51]					
Latent Class 2	0.10	0.52		0.10	0.50	
(50%)	[0.10 0.10]	[0.51 0.53]				
Latent Class 3	0.10	0.49	0.90	0.10	0.50	0.90
(49%)	[0.10 0.10]	[0.48 0.50]	[0.90 0.91]			

Table 5.25: LCHMM: Experiment 3.1 conditional probabilities posterior mean and 95% CI

In summary, introducing heterogeneity in the structural model can avoid duplicity on states, it helps identifying correctly customers with two and three states, and it allows obtaining accurate results at aggregate level.

Covariates: Experiments 4.x

In Section 4.2.4 we showed that EF2 is present when all parameters are heterogeneous or when the intercept of the conditional probabilities α^0 is heterogeneous and the covariate effects are homogeneous and similar across states. We applied LCHMM to Experiment 4.6 and found consistent results with the previous section. In Table 5.26 we report the model comparison criteria. As in previous experiments, LCHMM is worse than 3HMM in terms of information criteria. However, as we show later, this model provides more accurate aggregate results.

Model	LML	DIC
3HMM	**-30483.06**	**62163.14**
LCHMM	-31235.94	62834.56

Table 5.26: Model Comparison

In Section 4.2.4 we show that states do not represent the same for all customers when the effects are heterogeneous or homogeneous but not different enough among states. However, if the effects are different, then the states represent the same for all avoiding the duplication of states for segment 1. We report in Tables 5.27 and 5.28 posterior mean and CI for Q and p_{ks} of the model. The results show that LCHMM captures correctly segments with 2 and 3 states.

		Posterior Mean[a]			**Simulated**		
		1	**2**	**3**	**1**	**2**	**3**
Latent Class 2 (50%)	**1**	0.70 [0.70 0.70]	0.30 [0.30 0.30]		0.70	0.30	NA
	2	0.30 [0.30 0.30]	0.70 [0.70 0.70]		0.30	0.70	NA
	3				NA	NA	NA
Latent Class 3 (49%)	**1**	0.70 [0.69 0.70]	0.20 [0.20 0.20]	0.10 [0.10 0.10]	0.70	0.20	0.10
	2	0.20 [0.20 0.21]	0.51 [0.51 0.51]	0.29 [0.28 0.29]	0.20	0.50	0.30
	3	0.10 [0.10 0.10]	0.20 [0.20 0.21]	0.70 [0.69 0.70]	0.10	0.20	0.70

Table 5.27: LCHMM Experiment 4.6 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			**Simulated**		
	1	**2**	**3**	**1**	**2**	**3**
Latent Class 1 (1%)	0.10 [0.10 0.10]				NA	
Latent Class 2 (50%)	0.10 [0.10 0.11]	0.51 [0.50 0.52]		0.10	0.50	
Latent Class 3 (49%)	0.10 [0.10 0.11]	0.51 [0.50 0.52]	0.91 [0.90 0.91]	0.10	0.50	0.90

Table 5.28: LCHMM: Experiment 4.6 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

In addition, most covariate effects are correctly recovered as reported in Tables 5.29 and 5.30.

Parameter		**Simulated $\mathbf{K_2}$**	**Posterior Mean Latent Class 2**	**95% CI**
Transition Matrix Parameters				
Low transition threshold state 1	τ_{k11}	0.85	0.85	[0.85 0.86]
Low transition threshold state 2	τ_{k21}	-0.85	-0.85	[-0.86 -0.84]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.48	[0.47 0.48]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.00	0.01	[0.01 0.02]
Conditional Probabilities Parameters				
Intercept of state 1	α_1^0	-2.20	-2.16	[-2.18 -2.14]
Intercept of state 2	α_2^0	0.79	0.79	[0.78 0.81]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.50	[0.49 0.51]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.10	0.10	[0.09 0.10]

Table 5.29: Experiment 4.6: Parameter estimates

Parameter		**Simulated $\mathbf{K_3}$**	**Posterior Mean Latent Class 3**	**95% CI**
Transition Matrix Parameters				
Low transition threshold state 1	τ_{k11}	0.85	0.82	[0.82 0.83]
High transition threshold state 1	τ_{k12}	0.30	0.31	[0.30 0.32]
Low transition threshold state 2	τ_{k21}	-1.39	-1.36	[-1.37 -1.35]
High transition threshold state 2	τ_{k22}	0.80	0.82	[0.82 0.83]
Low transition threshold state 3	τ_{k31}	-2.20	-2.20	[-2.20 -2.19]
High transition threshold state 3	τ_{k32}	0.30	0.31	[0.31 0.32]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.49	[0.49 0.50]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.00	-0.01	[-0.01 -0.00]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	0.80	0.81	[0.81 0.82]
Conditional Probabilities Parameters				
Intercept of state 1	α_1^0	-2.20	-2.16	[-2.18 -2.14]
Intercept of state 2	α_2^0	0.79	0.78	[0.77 0.79]
Intercept of state 3	α_3^0	0.79	0.80	[0.79 0.82]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.47	[0.47 0.48]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.10	0.10	[0.10 0.10]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	0.10	0.09	[0.08 0.09]

Table 5.30: Experiment 4.6: Parameter estimates

In summary, introducing latent classes with different HMM on each class increases importantly the number of parameters, therefore information criteria such as DIC can indicate that this model is worse than a model with the same number of states. However, this model captures heterogeneity on the number of states, avoiding the effect of duplicated states and providing more accurate aggregate results that lead to better conclusions from the population behavior.

Chapter 6

Empirical illustration

In this chapter we describe an application of both proposed models in empirical data. For this purpose we test the models using data from physicians' prescriptions behavior of a new drug [Montoya et al., 2010]. Data comprised new prescriptions of a specific new drug at the physician-level and monthly detailing and sampling activities from the pharmaceutical company over a 2 years period. Moreover, data contain the total number of prescriptions given by each physician in the same category of the new drug, this helps to incorporate changes in demand. We estimate the model described in Section 3.1, and both proposed models, modified to incorporate the data specific structure. We compare all three models to identify differences among the parameter estimates.

6.1 Data description

We use the data set of the empirical application analyzed by Montoya et al. [2010]. The dataset contains new prescriptions and marketing efforts of a drug from 300 physicians for the 24 month period after its introduction to the market. The dataset contains for each month and each physician, the number of prescriptions given, and the number of details and samples received from the pharmaceutical company. Detailing corresponds to the number of contacts with a physician in order to give information about the drug. Sampling corresponds to the number of free samples of the new drug given by the pharmaceutical company to an specific physician. Additionally, the dataset contains the total number of prescriptions on the category given by each physician at each month. Each physician on the sample has received at least one detail and one sample. For further details and descriptive statistics of this data set, see Montoya et al. [2010].

6.2 Model specification

We use the specifications described in Chapter 3.1, and Sections 5.1.1 and 5.2.1 with the following definitions:

1. **Initial Probabilities:** Customers starts on lower state.

$$\Pi_k = [1\,0\ldots 0]$$

2. **Covariates:** Standardized Detailing and Sampling

$$X_{kt} = [f(\text{Detailing}_{kt})\; f(\text{Sampling}_{kt})]$$

where $f(x) = \frac{\ln(x+1)-\mu}{\sigma}$, with $\mu = \text{mean}(\ln(x+1))$ and $\sigma = \text{std}(\ln(x+1))$.

3. **Observable data:** Number of prescriptions

$$Y_{kt}: \text{ Number of new prescriptions of new drug written by physician } k \text{ on month } t$$

4. **Number of trials of Binomial distribution:** Total category prescriptions

$$N_{kt}: \text{ Total number of prescriptions written by physician } k \text{ on month } t \text{ in the category}$$

6.3 Model estimation

We use a MCMC procedure to estimate each model using proper definitions of prior distributions, following the estimation procedure for each model on the simulation exercise.

6.4 Results

Using results from [Montoya et al., 2010], we tested the models with constant number of states ($S = 2$ and $S = 3$). As we show in Table 6.1 that the model with 3 states is the best model.[1]

Model	LML	DIC	MSC	Validation log-likelihood
2 States	-9071.50	18488.08	26460.80	-2439.83
3 States	**-8852.25**	**17265.21**	**22304.10**	**-2305.60**

Table 6.1: Empirical data: Homogeneous number of states models Comparison

Using this result, we report in Table 6.2 LML, DIC, MSC and validation log-likelihood for 3HMM, MOGP and LCHH. Given those criteria, it is not clear which is the best model. The fact that LCHMM and MOGP have more parameters than 3HMM, may induce over-fitting on proposed models, specially for LCHMM, given that more covariates imply more parameters per latent class and LCHMM is estimated with three classes (HMMs with 1, 2 and 3 states).

We report in Tables 6.3 and 6.4 the differences between those models on posterior means and CI of Q and p_{ks} of each model.

[1]These models were estimated using the first 20 periods of each physician, and periods 21-24 were used to compute the validation log-likelihood.

Model	LML	DIC	MSC	Validation log-likelihood
3HMM	-8852.25	17265.21	22304.10	**-2305.60**
MOGP	-8963.54	**17201.93**	**21237.77**	-2322.41
LCHMM	**-8432.88**	17477.22	–	-2753.77

Table 6.2: Empirical data: Model Comparison

Model	Component / Class		Posterior Mean[a] 1	2	3
3HMM	–	**1**	0.62 [0.58 0.65]	0.35 [0.32 0.37]	0.03 [0.02 0.07]
		2	0.19 [0.16 0.44]	0.72 [0.47 0.76]	0.09 [0.06 0.13]
		3	0.30 [0.24 0.40]	0.38 [0.22 0.44]	0.32 [0.25 0.43]
MOGP	**Component 1** (66.67%)	**1**	0.66 [0.56 0.69]	0.29 [0.22 0.32]	0.05 [0.03 0.18]
		2	0.29 [0.23 0.53]	0.62 [0.28 0.71]	0.09 [0.05 0.32]
		3	0.29 [0.24 0.35]	0.38 [0.11 0.40]	0.33 [0.30 0.63]
	Component 2 (33.33%)	**1**	0.48 [0.45 0.64]	0.33 [0.22 0.37]	0.19 [0.14 0.24]
		2	0.30 [0.25 0.41]	0.26 [0.18 0.30]	0.44 [0.30 0.50]
		3	0.10 [0.09 0.30]	0.18 [0.12 0.28]	0.72 [0.42 0.77]
LCHMM	**Latent Class 2** (70.39%)	**1**	0.59 [0.52 0.67]	0.41 [0.33 0.48]	
		2	0.17 [0.13 0.23]	0.83 [0.77 0.87]	
		3			
	Latent Class 3 (29.31%)	**1**	0.39 [0.28 0.51]	0.55 [0.42 0.66]	0.06 [0.01 0.18]
		2	0.23 [0.15 0.32]	0.33 [0.18 0.52]	0.44 [0.24 0.62]
		3	0.09 [0.06 0.13]	0.09 [0.05 0.18]	0.82 [0.71 0.88]

Table 6.3: Empirical data: Transition matrix posterior mean and 95% CI

[a]To report these values, we use the average covariates $X_{kt} = 0$.

Model	Component / Class	Posterior Mean[a]		
		1	**2**	**3**
3HMM		0.01 [0.01 0.01]	0.07 [0.07 0.08]	0.21 [0.19 0.23]
		1	**2**	**3**
MOGP	-	0.01 [0.01 0.02]	0.07 [0.06 0.07]	0.17 [0.14 0.18]
		1	**2**	**3**
LCHMM	**Latent Class 2** (70.39%)	0.00 [0.00 0.01]	0.07 [0.07 0.08]	
	Latent Class 3 (29.31%)	0.01 [0.01 0.02]	0.03 [0.02 0.04]	0.19 [0.19 0.22]

Table 6.4: Empirical data: Conditional probabilities posterior mean and 95% CI

[a]To report these values, we use the average covariates $X_{kt} = 0$.

Following the labels used by Montoya et al. [2010], we label the lower state "Inactive", the second state "Infrequent" and the higher state, "Frequent". Using the results of the conditional probabilities, we can conclude that states do not exactly represent the same behavior across models in terms of the conditional probabilities. The first two states are comparable between 3HMM, MOGP models. Nevertheless, the frequent state is slightly different between those models. In LCHMM, a random customer has less than 1% probability of having only one state, which indicates that customers' behavior is dynamic. 70% of customers have 2 states, and those states capture the inactive and infrequent states estimated by 3HMM and MOGP. For latent class 3, the higher state captures the behavior of frequent state of model 3HMM (slightly different) and infrequent state of this class has a lower conditional probability than class 2 and the other two models.

At the transition matrix obtained by model 3HMM, we can suggest that the "frequent" state is a highly unstable state, only with a 34% probability of staying there on the next period. However, LCHMM indicates that for 29.3% of customers, that state has a 82% probability of staying in the "frequent" state the next period; and 70.3% are moving only between the first two states. Similar results show the MOGP model. Moreover, the probability of moving from "infrequent" to "frequent" state is underestimated at the aggregate level for that 29% of customers. 3HMM reports a 9% probability of making that move, while MOGP and LCHMM reports a 44% probability of reaching to the higher state using that movement for an approximate 30% of customers. This indicates that there is a misinterpretation of the results when there are observed at the aggregate level. Although LCHMM and MOGP tend to overfit to training data, those models give a better interpretation of the transition matrix results at the aggregate level, allowing to make more accurate conclusions from them. In addition, these results suggest that over-fitting of LCHMM is driven by customers which have not reached the frequent state on the estimation periods, but they move to the higher state on the validation period. However, if a customer never moves to the higher state over the estimation periods, most models that account for heterogeneity in the number of states will probably identify such customer as a customer with 2 states.

We also report posterior means of covariate effects for each model in Table 6.5.

Parameter		3HMM	MOGP		LCHMM		
			Comp. 1	Comp. 2	1 state	2 states	3 states
			(66.67%)	(33.33%)	(0.30%)	(70.39%)	(29.31%)
Transition Matrix							
Detailing State 1	ρ_1^d	0.23	0.08	-0.01	.	0.37	0.48
Detailing State 2	ρ_2^d	-0.05	0.04	0.70	.	0.02	0.00
Detailing State 3	ρ_3^d	0.02	-0.12	-0.11	.	.	-0.13
Sampling State 1	ρ_1^s	0.12	0.19	0.40	.	0.19	0.32
Sampling State 2	ρ_2^s	0.13	-0.09	-0.33	.	0.09	0.16
Sampling State 3	ρ_3^s	0.21	0.36	0.20	.	.	0.51
Conditional Probabilities							
Detailing State 1	$\boldsymbol{\alpha}_1^d$	0.18	-0.02	0.07	0.00	-0.05	0.56
Detailing State 2	$\boldsymbol{\alpha}_2^d$	0.02	-0.00	0.21	.	0.02	-0.03
Detailing State 3	$\boldsymbol{\alpha}_3^d$	-0.05	-0.06	-0.02	.	.	-0.02
Sampling State 1	$\boldsymbol{\alpha}_1^s$	0.24	0.18	0.70	0.01	0.21	-0.14
Sampling State 2	$\boldsymbol{\alpha}_2^s$	0.05	0.06	-0.38	.	-0.07	0.17
Sampling State 3	$\boldsymbol{\alpha}_3^s$	0.01	-0.06	0.14	.	.	0.11

Table 6.5: Empirical data: Covariates Estimates Results by model

Regarding the covariate effects on the transition matrix, the 3HMM underestimates the detailing effect on the inactive state, in comparison with LCHMM. In addition, the 3HMM underestimates the sampling effect on the frequent state, in comparison to component 1 of MOGP and class 3 of LCHMM. On the other hand, aggregate covariates effects on conditional probabilities are also biased. 3HMM indicates that detailing effect on inactive state is 0.18. However, using LCHMM, aggregate results show that 70% of customers (2 states) do not present a significant effect, whereas for the rest 30% with 3 states, the effect of detailing is higher (0.56).

In summary, results using empirical data show that assuming the same number of states induces to biased aggregate results. Incorporating more heterogeneity on the model allows for capturing the behavior of different segments, in particular, a segment with fewer states. Both proposed models identify a segment of customers ($\sim$ 30%) with 44% probability of moving to the higher state, whereas the basic model underestimates the same probability. In addition, the basic model underestimates the probability of staying in the frequent state. Moreover, the basic model also present biased aggregate results for covariate effects.

Chapter 7

Conclusions and Further Research

7.1 Conclusions

Estimating a HMM is usually a difficult task, specially due to the number of parameters and the correlation between transition and conditional distribution parameters. First, to estimate a HMM it is necessary to determine the number of states of the HMM and that task is usually made using criteria such as Log Marginal Likelihood (LML), Deviation Criterion (DIC) and Markov Switching Criterion. On the other hand, in marketing applications, observable behavior of each customer is usually modeled using a HMM, and heterogeneity is incorporated at the parameter level, using a Bayesian approach. Selecting the number of states for HMM in marketing involves assuming that all customers have the same number of states. If there is presence of heterogeneity in the number of states, the procedure to estimate a HMM may not deliver the correct results.

Monte Carlo simulations with heterogeneity in the number of states (customers with two or three states) show that heterogeneity at the parameter level helps estimating correctly at the individual level a HMM with different number of states across customers, when assuming the same number of states for all. However, aggregate results should be used carefully. The number of states which fits best the data is the highest individual level number of states in data. This means that the model that fits best the data is, for each customer, either the correct one for that customer or a HMM with more states. In the experiments we showed, the best model is the 3 state HMM, when data include customers with 2 and 3 states[1]. This model recovers correctly parameters of customers who have the same number of states than the HMM chosen. For customer with fewer states than the HMM selected, two effects can appear:

1. States have the same interpretation of states of customers with 3 states, but customers with 2 states do not reach to the states that they do not actually have (EF1).
2. To complete the number of states of the HMM selected for those customers, some states are duplicated, meaning that there are states that have the same interpretation and transitions between each other, and moves between those states have no changing effect on the distribution that is conditional on the hidden states (EF2).

[1]Except for experiment 2.7 with 100% customers with 2 states described in Section 4.2.2

When there is homogeneity on the conditional probabilities, results show the first effect (EF1). In this case, parameters associated with the conditional probabilities are correctly estimated, as well as the individual level parameters associated with the transition matrix. However, aggregate results of heterogeneous parameters are estimated as the mean of a mixture of individual results of customers who reach to every states and customers who do not. Given that, aggregate results induce interpretations about the population behavior that do not correspond to any individual customer in the mix.

One of these results at the aggregate level is the probability of making transitions to the higher state. At the individual level this probability is estimated correctly for customers with 3 states, and is recovered as zero for customers with 2 states. However, the aggregate parameter of this probability is estimated between the correct value of 3 state customers and zero, which does not represent transitions of any subsets of customers mentioned before. This finding can lead to conclude that, on average, it is difficult to move customers to the highest state, when in fact for some customers this probability is higher and for the rest this value is zero. Additionally, this aggregate probability decreases when the proportion of customers with two states increases, due to the higher weight of those customers in the total population.

When we incorporate heterogeneity on the conditional probabilities, there are not common parameters across customers in the model[2] and individual parameters are free to adapt to data of each customer. In this case, results show the second effect (EF2). Estimating a state duplicated from another, generates a transition matrix for those customers on which transitions between duplicated states must be interpreted as transitions in which the customer remains in the same state. This effect makes more difficult to understand the behavior of those customers at the individual level. Consequently, aggregate results are affected. The transition matrix obtained from aggregate parameters differ completely from the individual level transition matrices, given that states have different interpretations for different customers. Conditional probabilities at the aggregate level, as well, represent a mixture of conditional probabilities of the two sets of customers resulting in probabilities that no represent any customer in the set. Incorporating covariates could help identifying the states, as homogeneous effects fix the states across customers. If the effects are homogeneous and different then the result is (EF1). When those effects are homogeneous and not different enough, the states do not represent the same for all customers and the results show the second effect (EF2). If the effects are heterogeneous, even with effects different enough across states, the states can not be identified as the same across customers and the results also show effect (EF2).

The main issue, generated by not accounting for heterogeneity on the number of states, relies on how to interpret the results at aggregate level. Observing multimodal distribuion at the heterogeneity distribution is the first signal to detect bias on aggregate results. Usually Normal distributions are used as priors. Fitting a mixture of Gaussians could help detecting a multimodal mixture on the posterior distribution of the parameters obtained by MCMC. Additionally, it is also helpful to incorporate more heterogeneity on the model. Using a mixture of Gaussians as a prior instead of Normal distributions helps solving the first effect (EF1) by reporting clearly at aggregate level the presence of two or more groups with different behaviors, and be able to detect subsets of customers that never reach to certain states. Moreover, another procedure to introduce more heterogeneity to solve the problems described, is to incorporate structural heterogeneity. The latent class of hidden Markov model is a method to introduce discrete structural heterogeneity, in this case, with a HMM of different size by each class. However, LCHMM increases the number of parameters, which implies obtaining worse results in terms of information criteria. On the other

[2]Besides the heterogeneity distribution.

hand, LCHMM identifies the heterogeneity on the number of states, which prevent biased aggregate results and eliminate both effects (EF1) and (EF2).

Applying those models to empirical data helped to understand better those effects. Individual level parameters were correctly estimated using a HMM with fixed number of states. At the aggregate level, results show that "frequent" state is more stable than reported. 70% of customers do not reach the higher state, and the rest have a 44% of moving from "infrequent" to "frequent" state instead of 9% as reported by the classical HMM.

In summary, we conclude that assuming the same number of states usually does not generate bias at the individual level, but conclusions from the aggregate level have to be made with precaution. In addition, introducing heterogeneity to all parameters could let free the states for each customer making more difficult to make conclusions from the results. Specifically, the model could identify transitions between different states (changes of behavior) when there is no change in observable behavior (when a customer makes moves between duplicated states). Improvements to support the model involves incorporating more heterogeneity in parameter estimation and methods to detect at the aggregate level the heterogeneity across customers.

7.2 Further Research

There are several research areas that could extend this work.

First, all HMM studied in this work have a first order discrete time Markov chain for modeling the latent variable. More research can be done by studying the heterogeneity on the number of states for hidden states modeled by continuous time Markov chains and higher ordered Markov chains.

Second, an embedded algorithm could be used to estimate both parameters and number of states simultaneously, such as reversible jump Markov chain Monte Carlo, and discover the effects of heterogeneity on the number of states when the model is estimated using this method. Additionally, reversible jump Markov chain Monte Carlo could be extended to capture the heterogeneity on the number of states, by estimating the number of states at the individual level, taking precautions on the label of states across customers (and parameters accordingly μ_θ, Σ_θ and ϕ) when birth, death, join and split moves are made.

Finally, we have focused the analyses on the estimation of the model parameters. It could be interesting to investigate whether the heterogeneity of the number of states affects the procedure to optimize the marketing mix allocation, and to measure the potential impact on the optimal solution.

Bibliography

Ansari, A., Montoya, R., and Netzer, O. (2012). Dynamic learning in behavioral games: A hidden markov mixture of experts approach. *Quantitative Marketing and Economics*, pages 1–29.

Ascarza, E. and Hardie, B. G. (2013). A joint model of usage and churn in contractual settings. *Marketing Science*, 32(4):570–590.

Atchade, Y. (2006). An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in applied Probability*, 8(2):235–254.

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

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Green, P., Noad, R., and Smart, N. (2005). Further hidden markov model cryptanalysis. In *Cryptographic Hardware and Embedded Systems–CHES 2005*, pages 61–74. Springer.

Greene, W. (1997). *Econometric analysis*. Prentice-Hall international editions. Prentice Hall.

Gunter, S. and Bunke, H. (2003). Optimizing the number of states, training iterations and gaussians in an hmm-based handwritten word recognizer. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pages 472–476. IEEE.

Koop, G. (2003). *Bayesian econometrics*. J. Wiley.

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.

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.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete valued time series*, volume 70. CRC Press.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*, chapter Assessing the Number of Components in Mixture Models. Wiley. com.

Montgomery, A., Li, S., Srinivasan, K., and Liechty, J. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4):579–595.

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.

Netzer, O., Lattin, J., and Srinivasan, V. (2008). A hidden markov model of customer relationship dynamics. *Marketing Science*, 27(2):185–204.

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

Robert, C., Ryden, T., and Titterington, D. (2002). Bayesian inference in hidden markov models through the reversible jump markov chain monte carlo method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):57–75.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing (Wiley Series in Probability and Statistics)*. Wiley, 1 edition.

Scott, S. L. (2002). Bayesian methods for hidden markov models. *Journal of the American Statistical Association*, 97(457).

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.

Appendix A

Simulated mean and standard deviation of HMM parameters

This chapter contains simulated mean and standard deviation of heterogeneous parameters and simulated values for homogeneous parameters for each experiment.

A.1 Simulation Experiments 1.1 and 2.x

Table A.1 shows simulated mean and standard deviation of parameters for experiments 1.1 and 2.x. Recall that experiments 2.x differ from experiment 1 only on the proportion of customers with 2 and 3 states.

Parameter		Mean	Std. dev.
2 State Customers			
Transition Matrix Parameters			
Intercept of probability 1 $\rightarrow$ 1	τ_{k11}	0.85	0.13
Intercept of probability 2 $\rightarrow$ 2	τ_{k22}	0.85	0.13
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
3 State Customers			
Transition Matrix Parameters			
Intercept of probability 1 $\rightarrow$ 1	τ_{k11}	1.95	0.20
Intercept of probability 1 $\rightarrow$ 2	τ_{k12}	0.69	0.12
Intercept of probability 2 $\rightarrow$ 1	τ_{k21}	-0.41	0.09
Intercept of probability 2 $\rightarrow$ 2	τ_{k22}	0.51	0.10
Intercept of probability 3 $\rightarrow$ 2	τ_{k32}	0.69	0.12
Intercept of probability 3 $\rightarrow$ 3	τ_{k33}	1.95	0.20
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
Intercept of state 3	α_3^0	0.79	.

Table A.1: Mean and std. deviation of simulation experiments 1.1 and 2.x

A.2 Simulation Experiment 3.1

Table A.2 shows simulated mean and standard deviation of heterogeneous parameters for experiment 3.1.

Parameter		**Mean**	**Std. dev.**
2 State Customers			
Transition Matrix Parameters			
Intercept of probability 1 → 1	τ_{k11}	0.85	0.13
Intercept of probability 2 → 2	τ_{k22}	0.85	0.13
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
3 State Customers			
Transition Matrix Parameters			
Intercept of probability 1 → 1	τ_{k11}	1.95	0.20
Intercept of probability 1 → 2	τ_{k12}	0.69	0.12
Intercept of probability 2 → 1	τ_{k21}	-0.41	0.09
Intercept of probability 2 → 2	τ_{k22}	0.51	0.10
Intercept of probability 3 → 2	τ_{k32}	0.69	0.12
Intercept of probability 3 → 3	τ_{k33}	1.95	0.20
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
Intercept of state 3	α_3^0	0.79	0.13

Table A.2: Mean and std. deviation of simulation experiment 3.1

A.3 Simulation Experiments 4.1 and 4.2

Table A.3 shows simulated mean and standard deviation of heterogeneous parameters for experiments 4.1 and 4.2.

Parameter		Mean	Std. dev.
2 State Customers			
Transition Matrix Parameters			
Transition threshold parameter - state 1	τ_{k11}	0.85	0.13
Transition threshold parameter - state 2	τ_{k22}	-0.85	0.13
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	.
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	.
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	.
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	.
3 State Customers			
Transition Matrix Parameters			
Low transition threshold parameter - state 1	τ_{k11}	0.85	0.13
High transition threshold parameter - state 1	τ_{k12}	0.30	0.08
Low transition threshold parameter - state 2	τ_{k21}	-1.39	0.17
High transition threshold parameter - state 2	τ_{k22}	0.80	0.13
Low transition threshold parameter - state 3	τ_{k32}	-2.20	0.21
High transition threshold parameter - state 3	τ_{k33}	0.30	0.08
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	.
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	.
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	0.41	.
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
Intercept of state 3	α_3^0	0.79	.
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	.
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	.
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	0.63	.

Table A.3: Mean and std. deviation of simulation experiments 4.1 and 4.2

A.4 Simulation Experiment 4.3

Table A.4 shows simulated mean and standard deviation of heterogeneous parameters for experiment 4.3.

Parameter		**Mean**	**Std. dev.**
2 State Customers			
Transition Matrix Parameters			
Transition threshold parameter - state 1	τ_{k11}	0.85	0.13
Transition threshold parameter - state 2	τ_{k22}	-0.85	0.13
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	.
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	.
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.07
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.10
3 State Customers			
Transition Matrix Parameters			
Low transition threshold parameter - state 1	τ_{k11}	0.85	0.13
High transition threshold parameter - state 1	τ_{k12}	0.30	0.08
Low transition threshold parameter - state 2	τ_{k21}	-1.39	0.17
High transition threshold parameter - state 2	τ_{k22}	0.80	0.13
Low transition threshold parameter - state 3	τ_{k32}	-2.20	0.21
High transition threshold parameter - state 3	τ_{k33}	0.30	0.08
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	.
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	.
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	0.41	.
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	.
Intercept of state 2	α_2^0	0.79	.
Intercept of state 3	α_3^0	0.79	.
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.07
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.10
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	0.63	0.11

Table A.4: Mean and std. deviation of simulation experiment 4.3

A.5 Simulation Experiment 4.4

Table A.5 shows simulated mean and standard deviation of heterogeneous parameters for experiment 4.4.

Parameter		**Mean**	**Std. dev.**
2 State Customers			
Transition Matrix Parameters			
Transition threshold parameter - state 1	τ_{k11}	0.85	0.13
Transition threshold parameter - state 2	τ_{k22}	-0.85	0.13
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.10
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.10
3 State Customers			
Transition Matrix Parameters			
Low transition threshold parameter - state 1	τ_{k11}	0.85	0.13
High transition threshold parameter - state 1	τ_{k12}	0.30	0.08
Low transition threshold parameter - state 2	τ_{k21}	-1.39	0.17
High transition threshold parameter - state 2	τ_{k22}	0.80	0.13
Low transition threshold parameter - state 3	τ_{k32}	-2.20	0.21
High transition threshold parameter - state 3	τ_{k33}	0.30	0.08
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.10
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.10
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	0.41	0.09

Table A.5: Mean and std. deviation of simulation experiment 4.4

A.6 Simulation Experiment 4.5

Table A.6 shows simulated mean and standard deviation of heterogeneous parameters for experiment 4.5.

Parameter		**Mean**	**Std. dev.**
2 State Customers			
Transition Matrix Parameters			
Transition threshold parameter - state 1	τ_{k11}	0.85	0.13
Transition threshold parameter - state 2	τ_{k22}	-0.85	0.13
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
3 State Customers			
Transition Matrix Parameters			
Low transition threshold parameter - state 1	τ_{k11}	0.85	0.13
High transition threshold parameter - state 1	τ_{k12}	0.30	0.08
Low transition threshold parameter - state 2	τ_{k21}	-1.39	0.17
High transition threshold parameter - state 2	τ_{k22}	0.80	0.13
Low transition threshold parameter - state 3	τ_{k31}	-2.20	0.21
High transition threshold parameter - state 3	τ_{k32}	0.30	0.08
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
Intercept of state 3	α_3^0	0.79	0.13

Table A.6: Mean and std. deviation of simulation experiment 4.5

A.7 Simulation Experiment 4.6

Table A.7 shows simulated mean and standard deviation of heterogeneous parameters for experiment 4.6.

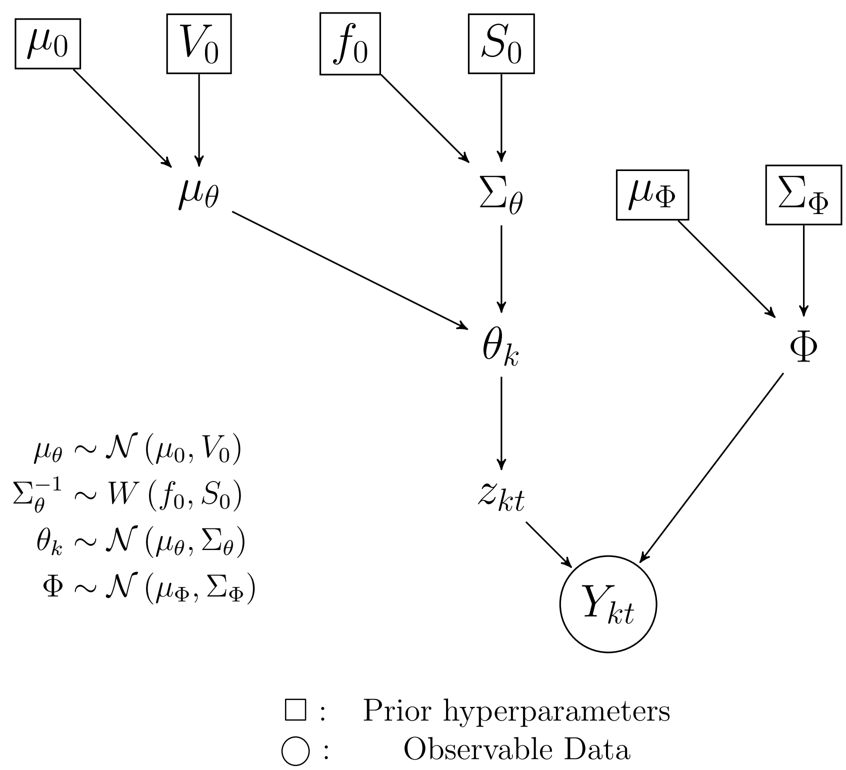
Parameter		**Mean**	**Std. dev.**
2 State Customers			
Transition Matrix Parameters			
Transition threshold parameter - state 1	τ_{k11}	0.85	0.13
Transition threshold parameter - state 2	τ_{k22}	-0.85	0.13
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.10
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.10
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.07
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.10
3 State Customers			
Transition Matrix Parameters			
Low transition threshold parameter - state 1	τ_{k11}	0.85	0.13
High transition threshold parameter - state 1	τ_{k12}	0.30	0.08
Low transition threshold parameter - state 2	τ_{k21}	-1.39	0.17
High transition threshold parameter - state 2	τ_{k22}	0.80	0.13
Low transition threshold parameter - state 3	τ_{k32}	-2.20	0.21
High transition threshold parameter - state 3	τ_{k33}	0.30	0.08
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.10
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.10
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	0.41	0.09
Conditional Probabilities Parameters			
Intercept of state 1	α_1^0	-2.20	0.21
Intercept of state 2	α_2^0	0.79	0.13
Intercept of state 3	α_3^0	0.79	0.13
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.07
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.10
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	0.63	0.11

Table A.7: Mean and std. deviation of simulation experiment 4.6

Appendix B

Prior and Full Conditionals Distributions

Figure B.1 shows each component of the model, with prior distribution and hyperparameters for population parameters (μ_θ,Σ_θ) and fixed parameters (Φ).



$$
\begin{aligned}
\mu_\theta &\sim \mathcal{N}\left(\mu_0, V_0\right)\\
\Sigma_\theta^{-1} &\sim W\left(f_0, S_0\right)\\
\theta_k &\sim \mathcal{N}\left(\mu_\theta, \Sigma_\theta\right)\\
\Phi &\sim \mathcal{N}\left(\mu_\Phi, \Sigma_\Phi\right)
\end{aligned}
$$

Figure B.1: Hierarchical Heterogeneous Estimation Model

Given this structure, full conditionals distribution of parameters θ_k, Φ, μ_θ, and Σ_θ are described as follows:

$$\mathbb{P}\left(\theta_k|\mu_\theta, \Sigma_\theta, \Phi, \{Y_{kt}\}_{t\in T}\right) \propto \exp\left(-\frac{1}{2}(\theta_k - \mu_\theta)^T \Sigma_\theta^{-1}(\theta_k - \mu_\theta)\right) L_k\left(\theta_k, \Phi | \{Y_{kt}\}_{t\in T}\right) \tag{B.1}$$

$$\mathbb{P}\left(\Phi|\mu_\theta, \Sigma_\theta, \{\theta_k\}_{k\in K}, Y\right) \propto \exp\left(-\frac{1}{2}(\Phi - \mu_\Phi)^T \Sigma_\Phi^{-1}(\Phi - \mu_\Phi)\right) L\left(\{\theta_k\}_{k\in K}, \Phi|Y\right) \tag{B.2}$$

$$\mu_\theta \sim \mathcal{N}(\mu_n, V_n) \tag{B.3}$$

$$\Sigma_\theta^{-1} \sim W(df_1, S_1) \tag{B.4}$$

where:

$$V_n = \left(V_0^{-1} + K\Sigma_\theta^{-1}\right)^{-1} \tag{B.5}$$

$$\mu_n = V_n\left(\mu_0 V_0^{-1} + K\bar{\theta}\Sigma_\theta^{-1}\right) \tag{B.6}$$

$$S_1 = \left(S_0^{-1} + \sum_{k=1}^{K}(\theta_k - \mu_\theta)(\theta_k - \mu_\theta)^T\right)^{-1} \tag{B.7}$$

$$\bar{\theta} = \frac{1}{K}\sum_{k=1}^{K}\theta_k \tag{B.8}$$

MCMC procedure generates draws from these distribution using a Gaussian random-walk Metropolis Hastings algorithm with an adaptive step, described in Appendix C, with an acceptance ratio of approximately 20% (see Appendix D).

We set uninformative prior hyperparameters as follows:

$$\begin{aligned}
\mu_0 &= \begin{bmatrix}0.69 & 0.33 & -0.69 & 0.33 & -0.69 & 0.33 & 0 & \dots & 0\end{bmatrix} \\
V_0 &= \frac{1}{5}\mathbb{I}_{n_\theta} \\
df_0 &= K + 5 \\
S_0 &= (df_0 - K - 1)\mathbb{I}_{n_\theta} \\
\mu_\Phi &= \begin{bmatrix}0 & -4 & -4 & 0 & \dots & 0\end{bmatrix} \\
\Sigma_\Phi &= \mathbb{I}_{n_\Phi}
\end{aligned}$$

Appendix C

Markov chain Monte Carlo algorithm

The Bayesian approach used to estimate the parameters is a hierarchical Bayesian Markov chain Monte Carlo (MCMC), specifically the adaptive Metropolis-Hastings procedure introduced by Atchadé Atchade [2006] with both Gibbs moves and Metropolis-Hastings acceptance criterion, and used by Netzer Netzer et al. [2008] and Montoya Montoya et al. [2010].

Each iteration of the algorithm consists in updating the parameters value to obtain draws from the posterior distribution, in four major steps:

1. Update Φ (Metropolis-Hastings acceptance criterion)
2. Update Σ_θ (Gibbs move from full conditionals)
3. Update μ_θ (Gibbs move from full conditionals)
4. Update θ_k for each k separately (Metropolis-Hastings acceptance criterion)

A more detailed explanation of each of the steps of the algorithm follows next.

Consider the $(i+1)$'th iteration, and the parameters obtained in the i'th iteration $(\left\{\theta_k^i\right\}, \Phi^i, \mu_\theta^i, \Sigma_\theta^i)$.

For the following equations, recall that n_θ and n_Φ represent the length of vectors θ_k and Φ respectively. Therefore, θ_k and μ_k are $n_\theta \times 1$ vectors, Φ is a $n_\Phi \times 1$ vector and Σ_θ is a $n_\theta \times n_\theta$ matrix.

1. <u>Update Φ (Metropolis-Hastings acceptance criterion)</u>

   Let $x^i = \left\{\left\{(\theta_k)^i\right\}_{k\in K}, \Phi^i\right\}$ the point where the MCMC is, before updating Φ.
   The proposed candidate is computed using a random walk as following:

$$\Phi^c = \Phi^i + \sigma_{i\phi} \cdot z$$

with $z \sim \mathcal{N}(0, \Lambda_{i\phi})$

The choices of $\sigma_{i\phi}$ and $\Lambda_{i\phi}$ are made by the algorithm using the adaptive part described in Appendix D.

Then the acceptance probability is $\alpha_{i\Phi} = \min\left\{1, \alpha_\Phi\left(\Phi^i, \Phi^c\right)\right\}$ where:

$$\alpha_\Phi\left(\Phi^i, \Phi^c\right) = \frac{L\left(\left\{\theta_k^i\right\}_{k\in K}, \Phi^c|Y\right) e^{-\frac{1}{2}(\Phi^c-\mu_\Phi)^T\Sigma_\Phi^{-1}(\Phi^c-\mu_\Phi)}}{L\left(\left\{\theta_k^i\right\}_{k\in K}, \Phi^i|Y\right) e^{-\frac{1}{2}(\Phi^i-\mu_\Phi)^T\Sigma_\Phi^{-1}(\Phi^i-\mu_\Phi)}}$$

Let $u \sim \mathcal{U}(0,1)$. If $u < \alpha_{i\Phi}$ then the candidate is accepted, i.e., $\Phi^{i+1} = \Phi^c$, otherwise is rejected, i.e., $\Phi^{i+1} = \Phi^i$. Also $x_{new}^i = \left\{\left\{\theta_k^i\right\}_{k\in K}, \Phi^{i+1}\right\}$.

x_{new}^i is not the $(i+1)$'th draw of the MCMC given that the movements on θ_k still have to be considered.

2. Update Σ_θ (Gibbs move from full conditionals)

   Let:

$$f_1 = f_0 + K$$

$$S_1 = \left(\sum_{k\in K}\left(\theta_k^i - \mu_\theta^i\right)\left(\theta_k^i - \mu_\theta^i\right)^T + {S_0}^{-1}\right)^{-1}$$

   Then the algorithm just draw from a Wishart distribution with parameters (f_1, S_1):

$$\left(\Sigma_\theta^{i+1}\right)^{-1} \sim W(f_1, S_1)$$

   (i.e., $\Sigma_\theta^{i+1} \sim IW\left(f_1, S_1^{-1}\right)$).

3. Update μ_θ (Gibbs move from full conditionals)

   Let:

$$\bar{\theta} = \frac{1}{K}\sum_{k\in K}\theta_k^i$$

$$V_1 = \left[V_0^{-1} + K\cdot\left(\Sigma_\theta^{i+1}\right)^{-1}\right]^{-1}$$

$$\mu_1 = V_1\left[V_0^{-1}\cdot\mu_0 + K\cdot\left(\Sigma_\theta^{i+1}\right)^{-1}\cdot\bar{\theta}\right]$$

   Then the algorithm just draw from a Multivariate Normal distribution with parameters (μ_1, V_1):

$$\mu_\theta^{i+1} \sim \mathcal{N}(\mu_1, V_1)$$

4. Update θ_k for each k separately (Metropolis-Hastings acceptance criterion)

   For each $k \in K$ the algorithm does the following:

   Propose a candidate using a random walk aproach:

$$\theta_k^c = \theta_k^i + \sigma_{i\theta_k}\cdot z$$

   with $z \sim \mathcal{N}(0, \Lambda_{i\theta_k})$

   The choices of $\sigma_{i\theta_k}$ and $\Lambda_{i\theta_k}$ are made by the algorithm using the adaptive part described in Atchadé 2006.

Then the acceptance probability is $\alpha_{i\theta_k} = \min\left\{1, \alpha_{\theta_k}\left(\theta_k^i, \theta_k^c\right)\right\}$ where:

$$\alpha_{\theta_k}\left(\theta_k^i, \theta_k^c\right\} = \frac{L_k\left(\theta_k^c, \Phi^{i+1} | \left\{Y_{kt}\right\}_{t\in T}\right) e^{-\frac{1}{2}\left(\theta_k^c - \mu_\theta^{i+1}\right)^T \left(\Sigma_\theta^{i+1}\right)^{-1}\left(\theta_k^c - \mu_\theta^{i+1}\right)}}{L_k\left(\theta_k^i, \Phi^{i+1} | \left\{Y_{kt}\right\}_{t\in T}\right) e^{-\frac{1}{2}\left(\theta_k^i - \mu_\theta^{i+1}\right)^T \left(\Sigma_\theta^{i+1}\right)^{-1}\left(\theta_k^i - \mu_\theta^{i+1}\right)}}$$

Let $u \sim \mathcal{U}(0,1)$. If $u < \alpha_{i\theta_k}$ then the candidate is accepted, i.e., $\theta_k^{i+1} = \theta_k^c$, otherwise is rejected, i.e., $\theta_k^{i+1} = \theta_k^i$.

Finally, when the movements for each k are made, accepted or rejected, the algorithm generated the $(i+1)$'th draw of the posterior distribution: $x^{i+1} = \left\{\left\{\theta_k^{i+1}\right\}_{k\in K}, \Phi^{i+1}\right\}$, μ_θ^{i+1} and Σ_θ^{i+1}.

Appendix D

Adaptative Process

This chapter describes the adaptive process that selects parameters $(\sigma_{i\phi}, \Lambda_{i\phi}, \sigma_{i\theta_k}, \Lambda_{i\theta_k})$ for random walk steps to improve convergence introduced by Atchadé (2006).

The adaptive process is computed on each iteration after all four updating parameter steps.

Let $A_1 = 10^7$, $\varepsilon_1 = 10^{-7}$ and $\varepsilon_2 = 10^{-6}$. Let $\bar{\tau} = 0.2$ the target acceptance rate of the candidates of the MCMC. Consider iteration i'th, then let $\gamma_i = 10/i$.

The adaptive process modifies on each iteration i, $(u_i, \sigma_i, \Gamma_i)$ the parameters needed to compute (σ_i, Λ_i) for the random walk step. Λ_i is computed using $\Lambda_i = \Gamma_i + \varepsilon_2$ where d is the numbers of rows (columns) of Λ_i and Γ_i.

Given that the algorithm has random walk steps on Φ and on θ_k for all k, it is needed a process for each type of random walk step, i.e., $(u_{i\phi}, \sigma_{i\phi}, \Gamma_{i\phi})$ and $(u_{i\theta_k}, \sigma_{i\theta_k}, \Gamma_{i\theta_k})$ for all k to compute $(\sigma_{i\Phi}, \Lambda_{i\Phi})$ and $(\sigma_{i\theta_k}, \Lambda_{i\theta_k})$ for all k.

The parameters $(u_i, \sigma_i, \Gamma_i)$ will be restricted to move only inside the space $B(0, A_1) \times [\varepsilon_1, A_1] \times \Theta_\Gamma$, where $B(0, A_1)$ is the ball of center 0 and radius A_1 in $\mathbb{R}^d$ (where d is the dimension of Φ or θ_k for each case), and Θ_Γ is the space of all semipositive-definite (spf) matrices Γ of $d \times d$ where $|\Gamma| \leq A_1$ [1]

Let p_1, p_2 and p_3 the three projection functions that keep $(u_i, \sigma_i, \Gamma_i)$ in $B(0, A_1) \times [0, \infty) \times \Theta_\Gamma$:

$$
\begin{array}{ccc}
p_1 : \mathbb{R} \longrightarrow [0, \infty) & p_2 : \mathcal{M}_{d \times d},\ \text{spf} \longrightarrow \Theta_\Gamma & p_3 : \mathbb{R}^d \longrightarrow B(0, A_1) \\
p_1(\sigma) = |\sigma| & p_2(\Sigma) = \begin{cases} \Sigma & \text{if } |\Sigma| \leq A_1 \\ \frac{A_1}{|\Sigma|}\Sigma & \text{if } |\Sigma| > A_1 \end{cases} & p_3(u) = \begin{cases} u & \text{if } |u| \leq \varepsilon_1 \\ \frac{A_1}{|u|}u & \text{if } |u| > A_1 \end{cases}
\end{array}
$$

The original Atchadé algorithm uses other p_1 function but the absolute value shows also good results.

[1] $|\Gamma|$ is the norm of a matrix Γ defined by $|\Gamma| = \left\{ \sum_{ij} \left| {\Gamma_{ij}}^2 \right| \right\}^{1/2}$

Given that all four steps on $(i+1)$'th iteration have been made, it means that the $(i+1)$'th draw from the posterior distribution is known ($x^{i+1} = \left\{ \{\theta_k^{i+1}\}_{k \in K}, \Phi^{i+1} \right\}$), and the values of $(u_{i\phi}, \sigma_{i\phi}, \Gamma_{i\phi})$ and $\{(u_{i\theta_k}, \sigma_{i\theta_k}, \Gamma_{i\theta_k})\}_{k \in K}$ of the previous iteration i.

The three steps of the adaptive procedure follow next:

1. Update $(u_{i\phi}, \sigma_{i\phi}, \Gamma_{i\phi})$
$$u_{i+1,\phi} = p_3 \left(u_{i\phi} + \gamma_i \left(\Phi^{i+1} - u_{i\phi} \right) \right)$$
$$\Gamma_{i+1,\phi} = p_2 \left(\Gamma_{i\phi} + \gamma_i \left((\Phi^{i+1} - u_{i\phi})(\Phi^{i+1} - u_{i\phi})^T - \Gamma_{i\phi} \right) \right)$$
$$\sigma_{i+1,\phi} = \left| \sigma_{i\phi} + \gamma_i \left(\alpha_{i\Phi} - \sigma_{i\phi} \right) \right|$$
where $\alpha_{i\Phi} = \min \left\{ 1, \alpha_\Phi \left(\Phi^i, \Phi^c \right) \right\}$ is the probability of accepting candidate Φ^c on the Metropolis-Hastings step for Φ.

2. For each $k \in K$: Update $(u_{i\theta_k}, \sigma_{i\theta_k}, \Gamma_{i\theta_k})$
$$u_{i+1,\theta_k} = p_3 \left(u_{i\theta_k} + \gamma_i \left(\theta_k^{i+1} - u_{i\theta_k} \right) \right)$$
$$\Gamma_{i+1,\theta_k} = p_2 \left(\Gamma_{i\theta_k} + \gamma_i \left((\theta_k^{i+1} - u_{i\theta_k})(\theta_k^{i+1} - u_{i\theta_k})^T - \Gamma_{i\theta_k} \right) \right)$$
$$\sigma_{i+1,\theta_k} = \left| \sigma_{i\theta_k} + \gamma_i \left(\alpha_{i\theta_k} - \sigma_{i\theta_k} \right) \right|$$
where $\alpha_{i\theta_k} = \min \left\{ 1, \alpha_{\theta_k} \left(\theta_k^i, \theta_k^c \right) \right\}$ is the probability of accepting candidate θ_k^c on the Metropolis-Hastings step for θ_k.

3. Set $\Lambda_{i+1,\phi} = \Gamma_{i+1,\phi} + \varepsilon_2 \mathbb{I}_n$ and $\Lambda_{i+1,\theta_k} = \Gamma_{i+1,\theta_k} + \varepsilon_2 \mathbb{I}_{n(n-1)}$ for each k.

In practice, this procedure is only computed from iteration $i > 100$ and for $i \leq 100$ $\sigma_{i\theta_k} = 0.05$, $\sigma_{i\phi} = 0.01$, $\Gamma_{i\theta_k} = \mathbb{I}_{n(n-1)}$, $\Gamma_{i\phi} = \mathbb{I}_n$ (recall that n is the number of states, meaning that $n(n-1)$ is the dimension of θ_k and n the dimension of Φ).

Appendix E

Parameter estimates Monte Carlo simulations

E.1 Experiment 1: Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**	
Transition Matrix Parameters						
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	2.91	[2.71	3.12]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	1.82	[1.59	2.04]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	0.85	[0.57	1.12]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	1.85	[1.60	2.09]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	0.38	[0.19	0.53]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	1.67	[1.49	1.83]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21	-2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78	0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78	0.80]

Table E.1: Experiment 1.1: Parameter estimates

E.2 Experiment 2.x: Parameter estimates

E.2.1 Experiment 2.1: Parameter estimates

Parameter		Simulated K_2	Simulated K_3	Posterior Mean	95% CI
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	2.91	[2.71 3.12]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	1.82	[1.59 2.04]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	0.85	[0.57 1.12]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	1.85	[1.60 2.09]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	0.38	[0.19 0.53]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	1.67	[1.49 1.83]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78 0.80]

Table E.2: Experiment 2.1: Parameter estimates

E.2.2 Experiment 2.2: Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	3.28	[3.06 3.51]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	2.23	[1.99 2.47]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	1.72	[1.40 2.03]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	2.67	[2.36 2.96]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	0.07	[-0.13 0.28]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	1.12	[0.90 1.35]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78 0.80]

Table E.3: Experiment 2.2: Parameter estimates

E.2.3 Experiment 2.3: Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	3.51	[3.27 3.75]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	2.49	[2.23 2.74]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	2.17	[1.81 2.50]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	3.08	[2.75 3.39]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	0.00	[-0.24 0.20]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	0.82	[0.59 1.02]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.77 0.80]

Table E.4: Experiment 2.3: Parameter estimates

E.2.4 Experiment 2.4: Parameter estimates

Parameter		Simulated $\mathbf{K_2}$	Simulated $\mathbf{K_3}$	Posterior Mean	95% CI	
Transition Matrix Parameters						
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	3.93	[3.72	4.16]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	2.96	[2.72	3.19]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	2.80	[2.47	3.12]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	3.67	[3.36	3.97]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	-0.42	[-0.68	-0.15]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	0.46	[0.19	0.73]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21	-2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78	0.79]
Intercept of state 3	α_3^0	.	0.79	0.78	[0.77	0.80]

Table E.5: Experiment 2.4: Parameter estimates

E.2.5 Experiment 2.5: Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	4.13	[3.92 4.31]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	3.21	[2.97 3.41]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	3.05	[2.77 3.35]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	3.88	[3.61 4.15]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	-0.38	[-0.69 -0.03]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	0.09	[-0.20 0.37]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79	0.78	[0.75 0.80]

Table E.6: Experiment 2.5: Parameter estimates

E.2.6 Experiment 2.6: Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95	0.15	[-0.18 0.52]
Intercept of probability 1 → 2	τ_{k12}	.	0.69	-0.06	[-0.55 0.24]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	-0.25	[-0.49 0.10]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	0.56	[0.38 0.71]
Intercept of probability 3 → 2	τ_{k32}	.	0.69	-1.35	[-1.63 -1.05]
Intercept of probability 3 → 3	τ_{k33}	.	1.95	1.06	[0.96 1.16]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.21	[-2.23 -2.20]
Intercept of state 2	α_2^0	0.79	0.79	-4.31	[-5.98 -3.00]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78 0.80]

Table E.7: Experiment 2.6: Parameter estimates

E.2.7 Experiment 2.7: Parameter estimates

Parameter		**Simulated**		**Posterior Mean**	**95% CI**
		K_2	**K_3**		
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	1.95		[ ]
Intercept of probability 1 → 2	τ_{k12}	.	0.69		[ ]
Intercept of probability 2 → 1	τ_{k21}	.	-0.41	0.84	[0.77 0.90]
Intercept of probability 2 → 2	τ_{k22}	0.85	0.51	0.79	[0.71 0.86]
Intercept of probability 3 → 2	τ_{k32}	.	0.69		[ ]
Intercept of probability 3 → 3	τ_{k33}	.	1.95		[ ]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79		[ ]

Table E.8: Experiment 2.7: Parameter estimates

E.3 Experiment 3.1 Parameter estimates

Parameter		Simulated $\mathbf{K_2}$	Simulated $\mathbf{K_3}$	Posterior Mean	95% CI	
Transition Matrix Parameters						
Intercept of probability $1 \to 1$	τ_{k11}	0.85	1.95	1.39	[1.28	1.48]
Intercept of probability $1 \to 2$	τ_{k12}	.	0.69	-1.05	[-1.35	-0.74]
Intercept of probability $2 \to 1$	τ_{k21}	.	-0.41	-0.46	[-0.61	-0.33]
Intercept of probability $2 \to 2$	τ_{k22}	0.85	0.51	-0.03	[-0.26	0.21]
Intercept of probability $3 \to 2$	τ_{k32}	.	0.69	-1.31	[-1.67	-0.97]
Intercept of probability $3 \to 3$	τ_{k33}	.	1.95	1.33	[1.22	1.43]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.24	-2.16]
Intercept of state 2	α_2^0	0.79	0.79	0.12	[-0.05	0.25]
Intercept of state 3	α_3^0	.	0.79	0.66	[0.62	0.71]

Table E.9: Experiment 3.1: Parameter estimates

E.4 Experiment 4.x Parameter estimates

E.4.1 Experiment 4.1 Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**	
Transition Matrix Parameters						
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.83	[0.77	0.90]
High transition threshold state 1	τ_{k12}	.	0.30	1.06	[0.92	1.21]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-1.05	[-1.13	-0.96]
High transition threshold state 2	τ_{k22}	.	0.80	1.42	[1.30	1.54]
Low transition threshold state 3	τ_{k31}	.	-2.20	-1.99	[-2.11	-1.85]
High transition threshold state 3	τ_{k32}	.	0.30	0.28	[0.11	0.41]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.51	0.52	[0.45	0.60]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.48	0.46	[0.38	0.53]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.41	0.40	[0.29	0.52]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21	-2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78	0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78	0.79]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.26	0.26	[0.25	0.27]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.46	0.46	[0.46	0.47]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.63	0.62	[0.60	0.63]

Table E.10: Experiment 4.1: Parameter estimates

E.4.2 Experiment 4.2 Parameter estimates

Parameter		Simulated $\mathbf{K_2}$	$\mathbf{K_3}$	Posterior Mean	95% CI	
Transition Matrix Parameters						
Intercept of probability 1 → 1	τ_{k11}	0.85	0.85	0.71	[0.63	0.80]
Intercept of probability 1 → 2	τ_{k12}	.	0.30	1.17	[1.02	1.32]
Intercept of probability 2 → 1	τ_{k21}	-0.85	-1.39	-1.11	[-1.21	-1.01]
Intercept of probability 2 → 2	τ_{k22}	.	0.80	1.45	[1.32	1.57]
Intercept of probability 3 → 2	τ_{k32}	.	-2.20	-2.18	[-2.33	-2.03]
Intercept of probability 3 → 3	τ_{k33}	.	0.30	0.36	[0.24	0.48]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.51	0.42	[0.29	0.54]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.48	0.54	[0.39	0.67]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.41	0.46	[0.26	0.65]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.22	-2.18]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78	0.80]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78	0.80]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.26	0.26	[0.24	0.28]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.46	0.46	[0.45	0.47]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.63	0.63	[0.60	0.66]

Table E.11: Experiment 4.2: Parameter estimates

E.4.3 Experiment 4.3 Parameter estimates

Parameter		Simulated $\mathbf{K_2}$	Simulated $\mathbf{K_3}$	Posterior Mean	95% CI
Transition Matrix Parameters					
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.94	[0.87 1.00]
High transition threshold state 1	τ_{k12}	.	0.30	1.03	[0.90 1.16]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-1.02	[-1.09 -0.94]
High transition threshold state 2	τ_{k22}	.	0.80	1.41	[1.28 1.53]
Low transition threshold state 3	τ_{k31}	.	-2.20	-1.71	[-1.94 -1.47]
High transition threshold state 3	τ_{k32}	.	0.30	0.21	[0.09 0.32]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.51	0.53	[0.46 0.60]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.48	0.52	[0.44 0.59]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.41	0.38	[0.27 0.49]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.80]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78 0.79]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.26	0.27	[0.25 0.28]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.46	0.46	[0.45 0.48]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.63	0.57	[0.52 0.62]

Table E.12: Experiment 4.3: Parameter estimates

E.4.4 Experiment 4.4 Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Intercept of probability 1 → 1	τ_{k11}	0.85	0.85	0.93	[0.86 0.99]
Intercept of probability 1 → 2	τ_{k12}	.	0.30	1.05	[0.91 1.19]
Intercept of probability 2 → 1	τ_{k21}	-0.85	-1.39	-1.03	[-1.11 -0.95]
Intercept of probability 2 → 2	τ_{k22}	.	0.80	1.40	[1.28 1.52]
Intercept of probability 3 → 2	τ_{k32}	.	-2.20	-1.90	[-2.02 -1.78]
Intercept of probability 3 → 3	τ_{k33}	.	0.30	0.29	[0.20 0.39]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.51	0.51	0.52	[0.45 0.59]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.48	0.48	0.53	[0.46 0.60]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.41	0.39	[0.27 0.51]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.20	[-2.21 -2.19]
Intercept of state 2	α_2^0	0.79	0.79	0.79	[0.78 0.79]
Intercept of state 3	α_3^0	.	0.79	0.79	[0.78 0.80]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.26	0.26	0.27	[0.26 0.28]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.46	0.46	0.47	[0.46 0.48]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.63	0.62	[0.60 0.64]

Table E.13: Experiment 4.4: Parameter estimates

E.4.5 Experiment 4.5 Parameter estimates

Experiment 4.5.1 Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.85	[0.79 0.92]
High transition threshold state 1	τ_{k12}	.	0.30	-0.63	[-0.83 -0.44]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-1.14	[-1.26 -1.01]
High transition threshold state 2	τ_{k22}	.	0.80	0.12	[-0.06 0.30]
Low transition threshold state 3	τ_{k31}	.	-2.20	-1.53	[-1.64 -1.41]
High transition threshold state 3	τ_{k32}	.	0.30	-0.52	[-0.77 -0.33]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.53	0.53	0.55	[0.48 0.62]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.52	0.52	0.55	[0.44 0.66]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.61	0.60	[0.52 0.69]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.17	[-2.21 -2.13]
Intercept of state 2	α_2^0	0.79	0.79	0.75	[0.73 0.77]
Intercept of state 3	α_3^0	.	0.79	-0.12	[-0.32 0.06]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.53	0.53	0.52	[0.51 0.53]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.63	0.63	0.64	[0.62 0.65]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.71	0.66	[0.65 0.67]

Table E.14: Experiment 4.5.1: Parameter estimates

Experiment 4.5.2 Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**	
Transition Matrix Parameters						
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.82	[0.76	0.87]
High transition threshold state 1	τ_{k12}	.	0.30	1.07	[0.92	1.21]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-1.07	[-1.16	-0.99]
High transition threshold state 2	τ_{k22}	.	0.80	1.47	[1.35	1.60]
Low transition threshold state 3	τ_{k31}	.	-2.20	-1.59	[-1.73	-1.44]
High transition threshold state 3	τ_{k32}	.	0.30	0.07	[-0.02	0.15]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.50	0.46	[0.39	0.53]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.80	0.80	0.93	[0.85	1.02]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.10	0.12	[0.01	0.24]
Conditional Probabilities Parameters						
Intercept of state 1	α_1^0	-2.20	-2.20	-2.21	[-2.23	-2.18]
Intercept of state 2	α_2^0	0.79	0.79	0.80	[0.78	0.81]
Intercept of state 3	α_3^0	.	0.79	0.78	[0.74	0.83]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.50	0.50	[0.49	0.51]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.80	0.80	0.80	[0.79	0.81]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.10	0.09	[0.08	0.11]

Table E.15: Experiment 4.5.2: Parameter estimates

E.4.6 Experiment 4.6 Parameter estimates

Parameter		**Simulated** $\mathbf{K_2}$	$\mathbf{K_3}$	**Posterior Mean**	**95% CI**
Transition Matrix Parameters					
Low transition threshold state 1	τ_{k11}	0.85	0.85	0.82	[0.77 0.87]
High transition threshold state 1	τ_{k12}	.	0.30	-0.39	[-0.51 -0.27]
Low transition threshold state 2	τ_{k21}	-0.85	-1.39	-0.70	[-0.81 -0.59]
High transition threshold state 2	τ_{k22}	.	0.80	0.30	[0.18 0.40]
Low transition threshold state 3	τ_{k31}	.	-2.20	-1.50	[-1.59 -1.41]
High transition threshold state 3	τ_{k32}	.	0.30	-0.22	[-0.31 -0.12]
Covariate effect on state 1	$\boldsymbol{\rho}_{k1}$	0.50	0.50	0.46	[0.41 0.51]
Covariate effect on state 2	$\boldsymbol{\rho}_{k2}$	0.00	0.00	-0.04	[-0.10 0.02]
Covariate effect on state 3	$\boldsymbol{\rho}_{k3}$	.	0.80	0.37	[0.29 0.45]
Conditional Probabilities Parameters					
Intercept of state 1	α_1^0	-2.20	-2.20	-2.17	[-2.19 -2.14]
Intercept of state 2	α_2^0	0.79	0.79	0.75	[0.73 0.77]
Intercept of state 3	α_3^0	.	0.79	-0.49	[-0.68 -0.32]
Covariate effect on state 1	$\boldsymbol{\alpha}_1$	0.50	0.50	0.49	[0.47 0.51]
Covariate effect on state 2	$\boldsymbol{\alpha}_2$	0.10	0.10	0.11	[0.08 0.13]
Covariate effect on state 3	$\boldsymbol{\alpha}_3$	.	0.10	0.09	[0.08 0.11]

Table E.16: Experiment 4.6: Parameter estimates

Appendix F

Transition matrix and conditional probabilities posterior mean and 95% CI

F.1 Experiment 1.1 and 2.1

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.29	0.01	0.70	0.30	NA
		[0.68 0.72]	[0.27 0.31]	[0.01 0.01]			
$\mathbf{K_2}$	**2**	0.31	0.67	0.01	0.30	0.70	NA
		[0.29 0.33]	[0.65 0.69]	[0.01 0.02]			
	3	0.16	0.16	0.68	NA	NA	NA
		[0.13 0.21]	[0.12 0.20]	[0.61 0.74]			
	1	0.70	0.21	0.08	0.70	0.20	0.10
		[0.68 0.72]	[0.19 0.23]	[0.07 0.10]			
$\mathbf{K_3}$	**2**	0.18	0.55	0.27	0.20	0.50	0.30
		[0.16 0.20]	[0.52 0.58]	[0.24 0.30]			
	3	0.11	0.21	0.68	0.10	0.20	0.70
		[0.10 0.13]	[0.19 0.22]	[0.66 0.70]			
	1	0.72	0.24	0.04			
		[0.70 0.73]	[0.23 0.26]	[0.03 0.05]			
K	**2**	0.24	0.65	0.10		NA	
		[0.22 0.26]	[0.63 0.68]	[0.08 0.13]			
	3	0.13	0.19	0.68			
		[0.12 0.15]	[0.16 0.21]	[0.65 0.71]			

Table F.1: Experiment 1.1 transition matrix posterior mean and 95% CI

	Posterior Mean			**Simulated**		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table F.2: Experiment 1.1 conditional probabilities posterior mean and 95% CI

F.2 Experiments 2.x

F.2.1 Experiment 2.2

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.29	0.01	0.70	0.30	NA
		[0.68 0.71]	[0.28 0.31]	[0.00 0.01]			
$\mathbf{K_2}$	**2**	0.31	0.69	0.01	0.30	0.70	NA
		[0.29 0.32]	[0.67 0.70]	[0.00 0.01]			
	3	0.33	0.18	0.49	NA	NA	NA
		[0.26 0.45]	[0.13 0.22]	[0.37 0.57]			
	1	0.70	0.21	0.09	0.70	0.20	0.10
		[0.67 0.73]	[0.19 0.23]	[0.08 0.11]			
$\mathbf{K_3}$	**2**	0.19	0.55	0.26	0.20	0.50	0.30
		[0.16 0.21]	[0.52 0.59]	[0.23 0.29]			
	3	0.11	0.22	0.66	0.10	0.20	0.70
		[0.10 0.13]	[0.20 0.25]	[0.64 0.69]			
	1	0.72	0.25	0.03			
		[0.70 0.74]	[0.23 0.27]	[0.02 0.03]			
K	**2**	0.27	0.69	0.05		NA	
		[0.24 0.29]	[0.66 0.71]	[0.04 0.06]			
	3	0.19	0.21	0.60			
		[0.17 0.24]	[0.18 0.24]	[0.53 0.63]			

Table F.3: Experiment 2.2 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table F.4: Experiment 2.2 conditional probabilities posterior mean and 95% CI

F.2.2 Experiment 2.3

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	**1**	0.70 [0.69 0.71]	0.29 [0.28 0.31]	0.01 [0.00 0.01]	0.70	0.30	NA
	2	0.31 [0.29 0.32]	0.69 [0.67 0.70]	0.00 [0.00 0.01]	0.30	0.70	NA
	3	0.40 [0.33 0.47]	0.22 [0.16 0.26]	0.38 [0.32 0.47]	NA	NA	NA
$\mathbf{K_3}$	**1**	0.70 [0.66 0.73]	0.21 [0.18 0.24]	0.09 [0.08 0.11]	0.70	0.20	0.10
	2	0.20 [0.17 0.23]	0.55 [0.51 0.59]	0.25 [0.21 0.30]	0.20	0.50	0.30
	3	0.11 [0.09 0.13]	0.22 [0.19 0.25]	0.67 [0.64 0.70]	0.10	0.20	0.70
K	**1**	0.72 [0.70 0.74]	0.26 [0.24 0.28]	0.02 [0.02 0.03]		NA	
	2	0.28 [0.26 0.30]	0.69 [0.67 0.71]	0.03 [0.02 0.04]			
	3	0.23 [0.20 0.27]	0.23 [0.19 0.26]	0.53 [0.49 0.58]			

Table F.5: Experiment 2.3 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10 [0.10 0.10]	0.50 [0.50 0.50]	0.90 [0.90 0.90]	0.10	0.50	0.90

Table F.6: Experiment 2.3 conditional probabilities posterior mean and 95% CI

F.2.3 Experiment 2.4

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	**1**	0.70 [0.68 0.71]	0.29 [0.28 0.31]	0.01 [0.01 0.01]	0.70	0.30	NA
	2	0.31 [0.29 0.32]	0.69 [0.67 0.70]	0.01 [0.00 0.01]	0.30	0.70	NA
	3	0.46 [0.39 0.56]	0.17 [0.14 0.21]	0.37 [0.28 0.45]	NA	NA	NA
$\mathbf{K_3}$	**1**	0.73 [0.68 0.77]	0.21 [0.18 0.26]	0.06 [0.04 0.08]	0.70	0.20	0.10
	2	0.20 [0.16 0.24]	0.57 [0.51 0.62]	0.23 [0.18 0.29]	0.20	0.50	0.30
	3	0.09 [0.07 0.12]	0.23 [0.19 0.27]	0.68 [0.62 0.72]	0.10	0.20	0.70
K	**1**	0.72 [0.70 0.73]	0.27 [0.25 0.29]	0.01 [0.01 0.02]			
	2	0.29 [0.27 0.31]	0.69 [0.67 0.71]	0.02 [0.01 0.02]		NA	
	3	0.31 [0.27 0.38]	0.20 [0.17 0.24]	0.49 [0.41 0.54]			

Table F.7: Experiment 2.4 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10 [0.10 0.10]	0.50 [0.50 0.50]	0.90 [0.90 0.90]	0.10	0.50	0.90

Table F.8: Experiment 2.4 conditional probabilities posterior mean and 95% CI

F.2.4 Experiment 2.5

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	**1**	0.70 [0.68 0.71]	0.30 [0.28 0.31]	0.01 [0.01 0.01]	0.70	0.30	NA
	2	0.31 [0.29 0.32]	0.69 [0.67 0.70]	0.01 [0.01 0.01]	0.30	0.70	NA
	3	0.46 [0.36 0.51]	0.23 [0.18 0.32]	0.32 [0.24 0.41]	NA	NA	NA
$\mathbf{K_3}$	**1**	0.74 [0.68 0.79]	0.21 [0.17 0.27]	0.05 [0.03 0.08]	0.70	0.20	0.10
	2	0.22 [0.18 0.29]	0.55 [0.47 0.61]	0.23 [0.16 0.31]	0.20	0.50	0.30
	3	0.08 [0.05 0.12]	0.23 [0.18 0.29]	0.69 [0.62 0.75]	0.10	0.20	0.70
K	**1**	0.71 [0.69 0.72]	0.28 [0.26 0.30]	0.01 [0.01 0.01]			
	2	0.30 [0.28 0.32]	0.69 [0.66 0.70]	0.01 [0.01 0.02]		NA	
	3	0.36 [0.29 0.41]	0.25 [0.20 0.33]	0.39 [0.32 0.47]			

Table F.9: Experiment 2.5 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10 [0.10 0.10]	0.50 [0.50 0.50]	0.90 [0.89 0.90]	0.10	0.50	0.90

Table F.10: Experiment 2.5 conditional probabilities posterior mean and 95% CI

F.2.5 Experiment 2.6

		Posterior Mean			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.38	0.30	0.32	0.70	0.30	NA
		[0.29 0.56]	[0.14 0.39]	[0.29 0.34]			
$\mathbf{K_2}$	**2**	0.22	0.50	0.28	0.30	0.70	NA
		[0.18 0.30]	[0.40 0.55]	[0.26 0.32]			
	3	0.24	0.06	0.70	NA	NA	NA
		[0.22 0.27]	[0.04 0.08]	[0.68 0.71]			
	1	0.38	0.30	0.32	0.70	0.20	0.10
		[0.28 0.57]	[0.14 0.40]	[0.28 0.35]			
$\mathbf{K_3}$	**2**	0.22	0.49	0.29	0.20	0.50	0.30
		[0.16 0.31]	[0.39 0.57]	[0.25 0.33]			
	3	0.22	0.06	0.72	0.10	0.20	0.70
		[0.20 0.26]	[0.04 0.07]	[0.69 0.75]			
	1	0.38	0.30	0.32			
		[0.29 0.56]	[0.14 0.39]	[0.29 0.34]			
K	**2**	0.22	0.50	0.28		NA	
		[0.18 0.30]	[0.40 0.55]	[0.26 0.32]			
	3	0.24	0.06	0.70			
		[0.22 0.27]	[0.04 0.08]	[0.68 0.71]			

Table F.11: Experiment 2.6 transition matrix posterior mean and 95% CI

	Posterior Mean			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50		0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]				

Table F.12: Experiment 2.6 conditional probabilities posterior mean and 95% CI

F.2.6 Experiment 2.7

		Posterior Mean		Simulated	
		1	**2**	**1**	**2**
	1	0.70	0.30	0.70	0.30
		[0.68 0.71]	[0.29 0.32]		
$\mathbf{K_2}$	**2**	0.31	0.69	0.30	0.70
		[0.30 0.33]	[0.67 0.70]		
	1	0.70	0.30	0.70	0.30
		[0.68 0.71]	[0.29 0.32]		
K	**2**	0.31	0.69	0.30	0.70
		[0.30 0.33]	[0.67 0.70]		

Table F.13: Experiment 2.7 transition matrix posterior mean and 95% CI

	Posterior Mean		Simulated		
	1	**2**	**1**	**2**	**3**
K	0.10	0.50	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]			

Table F.14: Experiment 2.7 conditional probabilities posterior mean and 95% CI

F.3 Experiment 3.1

		Posterior Mean 1	Posterior Mean 2	Posterior Mean 3	Simulated 1	Simulated 2	Simulated 3
K_2	**1**	0.71 [0.69 0.73]	0.02 [0.01 0.03]	0.27 [0.26 0.29]	0.70	0.30	NA
	2	0.27 [0.20 0.33]	0.24 [0.18 0.32]	0.49 [0.42 0.54]	0.30	0.70	NA
	3	0.30 [0.28 0.32]	0.01 [0.01 0.02]	0.69 [0.67 0.71]	NA	NA	NA
K_3	**1**	0.70 [0.68 0.72]	0.18 [0.17 0.20]	0.12 [0.10 0.13]	0.70	0.20	0.10
	2	0.20 [0.18 0.23]	0.51 [0.49 0.54]	0.28 [0.26 0.31]	0.20	0.50	0.30
	3	0.12 [0.10 0.13]	0.18 [0.17 0.20]	0.70 [0.68 0.72]	0.10	0.20	0.70
K	**1**	0.75 [0.72 0.76]	0.07 [0.05 0.09]	0.19 [0.17 0.20]			
	2	0.24 [0.21 0.27]	0.37 [0.32 0.43]	0.38 [0.34 0.42]		NA	
	3	0.20 [0.18 0.22]	0.05 [0.04 0.07]	0.75 [0.73 0.77]			

Table F.15: Experiment 3.1 transition matrix posterior mean and 95% CI

	Posterior Mean 1	Posterior Mean 2	Posterior Mean 3	Simulated 1	Simulated 2	Simulated 3
K_2	0.10 [0.10 0.10]	0.16 [0.15 0.18]	0.51 [0.51 0.51]	0.10	0.50	NA
K_3	0.10 [0.10 0.10]	0.49 [0.49 0.49]	0.90 [0.90 0.90]	0.10	0.50	0.90
K	0.10 [0.10 0.10]	0.26 [0.22 0.29]	0.71 [0.68 0.74]		NA	

Table F.16: Experiment 3.1 conditional probabilities posterior mean and 95% CI

F.4 Experiments 4.x

F.4.1 Experiment 4.1

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
K_2	**1**	0.69 [0.67 0.70]	0.31 [0.29 0.33]	0.00 [0.00 0.00]	0.70	0.30	NA
	2	0.30 [0.28 0.31]	0.70 [0.69 0.72]	0.00 [0.00 0.00]	0.30	0.70	NA
	3	0.13 [0.11 0.16]	0.24 [0.17 0.33]	0.63 [0.52 0.70]	NA	NA	NA
K_3	**1**	0.71 [0.69 0.73]	0.21 [0.19 0.22]	0.09 [0.08 0.10]	0.70	0.20	0.10
	2	0.23 [0.21 0.25]	0.49 [0.46 0.53]	0.28 [0.24 0.31]	0.20	0.50	0.30
	3	0.11 [0.10 0.12]	0.20 [0.18 0.22]	0.69 [0.67 0.71]	0.10	0.20	0.70
K	**1**	0.70 [0.68 0.71]	0.28 [0.26 0.30]	0.02 [0.01 0.03]		NA	
	2	0.26 [0.24 0.28]	0.70 [0.67 0.72]	0.04 [0.03 0.07]			
	3	0.12 [0.11 0.14]	0.22 [0.18 0.26]	0.66 [0.61 0.70]			

Table F.17: Experiment 4.1 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10 [0.10 0.10]	0.50 [0.50 0.50]	0.90 [0.90 0.90]	0.10	0.50	0.90

Table F.18: Experiment 4.1 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.2 Experiment 4.2

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.67	0.33	0.00	0.70	0.30	NA
		[0.64 0.69]	[0.31 0.36]	[0.00 0.00]			
$\mathbf{K_2}$	**2**	0.30	0.70	0.00	0.30	0.70	NA
		[0.28 0.32]	[0.68 0.72]	[0.00 0.00]			
	3	0.10	0.25	0.65	NA	NA	NA
		[0.08 0.13]	[0.19 0.37]	[0.52 0.73]			
	1	0.68	0.23	0.10	0.70	0.20	0.10
		[0.65 0.70]	[0.21 0.25]	[0.08 0.11]			
$\mathbf{K_3}$	**2**	0.21	0.45	0.34	0.20	0.50	0.30
		[0.19 0.23]	[0.42 0.49]	[0.30 0.38]			
	3	0.10	0.19	0.71	0.10	0.20	0.70
		[0.09 0.11]	[0.18 0.21]	[0.68 0.73]			
	1	0.67	0.31	0.02			
		[0.65 0.69]	[0.29 0.33]	[0.01 0.03]			
K	**2**	0.25	0.71	0.04		NA	
		[0.23 0.27]	[0.68 0.73]	[0.02 0.07]			
	3	0.10	0.22	0.68			
		[0.09 0.12]	[0.19 0.27]	[0.62 0.72]			

Table F.19: Experiment 4.2 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table F.20: Experiment 4.2 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.3 Experiment 4.3

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.30	0.00	0.70	0.30	NA
		[0.68 0.72]	[0.28 0.31]	[0.00 0.00]			
$\mathbf{K_2}$	**2**	0.31	0.69	0.00	0.30	0.70	NA
		[0.29 0.32]	[0.68 0.71]	[0.00 0.00]			
	3	0.23	0.25	0.52	NA	NA	NA
		[0.17 0.36]	[0.20 0.29]	[0.38 0.58]			
	1	0.74	0.18	0.08	0.70	0.20	0.10
		[0.72 0.75]	[0.17 0.20]	[0.07 0.10]			
$\mathbf{K_3}$	**2**	0.23	0.45	0.32	0.20	0.50	0.30
		[0.21 0.25]	[0.42 0.48]	[0.29 0.35]			
	3	0.10	0.20	0.70	0.10	0.20	0.70
		[0.09 0.11]	[0.19 0.22]	[0.67 0.72]			
	1	0.72	0.26	0.02			
		[0.71 0.73]	[0.24 0.27]	[0.02 0.03]			
K	**2**	0.27	0.69	0.04		NA	
		[0.25 0.28]	[0.67 0.71]	[0.03 0.07]			
	3	0.15	0.23	0.62			
		[0.13 0.20]	[0.21 0.27]	[0.54 0.65]			

Table F.21: Experiment 4.3 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table F.22: Experiment 4.3 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.4 Experiment 4.4

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.30	0.00	0.70	0.30	NA
		[0.68 0.72]	[0.28 0.32]	[0.00 0.00]			
$\mathbf{K_2}$	**2**	0.30	0.69	0.00	0.30	0.70	NA
		[0.29 0.32]	[0.68 0.71]	[0.00 0.00]			
	3	0.17	0.26	0.58	NA	NA	NA
		[0.14 0.19]	[0.20 0.32]	[0.50 0.63]			
	1	0.73	0.18	0.09	0.70	0.20	0.10
		[0.71 0.75]	[0.17 0.20]	[0.07 0.10]			
$\mathbf{K_3}$	**2**	0.23	0.46	0.31	0.20	0.50	0.30
		[0.21 0.25]	[0.43 0.49]	[0.28 0.34]			
	3	0.10	0.21	0.69	0.10	0.20	0.70
		[0.09 0.11]	[0.19 0.23]	[0.67 0.71]			
	1	0.72	0.26	0.02			
		[0.70 0.73]	[0.24 0.28]	[0.01 0.03]			
K	**2**	0.26	0.69	0.05		NA	
		[0.25 0.28]	[0.67 0.71]	[0.03 0.07]			
	3	0.13	0.23	0.64			
		[0.12 0.14]	[0.20 0.27]	[0.60 0.67]			

Table F.23: Experiment 4.4 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
K	0.10	0.50	0.90	0.10	0.50	0.90
	[0.10 0.10]	[0.50 0.50]	[0.90 0.90]			

Table F.24: Experiment 4.4 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.5 Experiment 4.5.1

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
	1	0.70	0.04	0.26	0.70	0.30	NA
		[0.68 0.71]	[0.04 0.07]	[0.23 0.28]			
$\mathbf{K_2}$	**2**	0.27	0.12	0.61	0.30	0.70	NA
		[0.24 0.36]	[0.10 0.16]	[0.50 0.65]			
	3	0.25	0.06	0.69	NA	NA	NA
		[0.23 0.27]	[0.04 0.10]	[0.65 0.71]			
	1	0.70	0.19	0.11	0.70	0.20	0.10
		[0.69 0.72]	[0.17 0.20]	[0.10 0.12]			
$\mathbf{K_3}$	**2**	0.22	0.48	0.30	0.20	0.50	0.30
		[0.21 0.25]	[0.45 0.51]	[0.26 0.32]			
	3	0.13	0.19	0.69	0.10	0.20	0.70
		[0.12 0.14]	[0.17 0.21]	[0.66 0.70]			
	1	0.70	0.10	0.20			
		[0.69 0.71]	[0.09 0.13]	[0.17 0.22]			
K	**2**	0.24	0.26	0.50		NA	
		[0.22 0.30]	[0.21 0.30]	[0.43 0.54]			
	3	0.18	0.10	0.72			
		[0.16 0.20]	[0.09 0.14]	[0.68 0.73]			

Table F.25: Experiment 4.5.1 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	0.10	0.46	0.54	0.10	0.50	NA
	[0.10 0.10]	[0.45 0.47]	[0.54 0.56]			
$\mathbf{K_3}$	0.11	0.51	0.89	0.10	0.50	0.90
	[0.10 0.11]	[0.51 0.52]	[0.88 0.90]			
K	0.10	0.49	0.70		NA	
	[0.10 0.11]	[0.47 0.50]	[0.66 0.73]			

Table F.26: Experiment 4.5.1 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.6 Experiment 4.5.2

		Posterior Mean[a]			Simulated		
		1	2	3	1	2	3
$\mathbf{K_2}$	1	0.69 [0.68 0.71]	0.31 [0.29 0.32]	0.00 [0.00 0.00]	0.70	0.30	NA
	2	0.29 [0.28 0.31]	0.70 [0.69 0.72]	0.00 [0.00 0.00]	0.30	0.70	NA
	3	0.24 [0.21 0.36]	0.20 [0.18 0.26]	0.55 [0.41 0.60]	NA	NA	NA
$\mathbf{K_3}$	1	0.69 [0.68 0.71]	0.20 [0.19 0.22]	0.10 [0.09 0.12]	0.70	0.20	0.10
	2	0.22 [0.20 0.24]	0.50 [0.46 0.53]	0.28 [0.25 0.32]	0.20	0.50	0.30
	3	0.11 [0.10 0.13]	0.19 [0.18 0.22]	0.69 [0.66 0.71]	0.10	0.20	0.70
K	1	0.69 [0.68 0.70]	0.28 [0.27 0.30]	0.02 [0.02 0.04]			
	2	0.26 [0.24 0.27]	0.71 [0.69 0.73]	0.04 [0.02 0.06]		NA	
	3	0.17 [0.15 0.22]	0.20 [0.18 0.25]	0.63 [0.54 0.65]			

Table F.27: Experiment 4.5.2 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	2	3	1	2	3
$\mathbf{K_2}$	0.10 [0.10 0.10]	0.51 [0.51 0.52]	0.90 [0.88 0.92]	0.10	0.50	NA
$\mathbf{K_3}$	0.10 [0.10 0.10]	0.49 [0.49 0.50]	0.90 [0.90 0.90]	0.10	0.50	0.90
K	0.10 [0.10 0.10]	0.50 [0.49 0.51]	0.90 [0.89 0.91]		NA	

Table F.28: Experiment 4.5.2 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

F.4.7 Experiment 4.6

		Posterior Mean[a]			Simulated		
		1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	**1**	0.69 [0.68 0.70]	0.07 [0.06 0.08]	0.24 [0.23 0.25]	0.70	0.30	NA
	2	0.47 [0.44 0.50]	0.19 [0.16 0.22]	0.34 [0.30 0.36]	0.30	0.70	NA
	3	0.26 [0.25 0.28]	0.11 [0.10 0.12]	0.62 [0.61 0.64]	NA	NA	NA
$\mathbf{K_3}$	**1**	0.70 [0.68 0.71]	0.20 [0.19 0.22]	0.10 [0.09 0.11]	0.70	0.20	0.10
	2	0.22 [0.20 0.25]	0.52 [0.50 0.55]	0.26 [0.23 0.28]	0.20	0.50	0.30
	3	0.12 [0.11 0.15]	0.21 [0.19 0.22]	0.67 [0.65 0.69]	0.10	0.20	0.70
K	**1**	0.69 [0.68 0.70]	0.12 [0.11 0.14]	0.18 [0.17 0.20]		NA	
	2	0.33 [0.31 0.36]	0.33 [0.30 0.36]	0.34 [0.30 0.37]			
	3	0.18 [0.17 0.20]	0.15 [0.14 0.16]	0.67 [0.65 0.68]			

Table F.29: Experiment 4.6 transition matrix posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

	Posterior Mean[a]			Simulated		
	1	**2**	**3**	**1**	**2**	**3**
$\mathbf{K_2}$	0.10 [0.10 0.10]	0.47 [0.47 0.48]	0.51 [0.51 0.52]	0.10	0.50	NA
$\mathbf{K_3}$	0.10 [0.10 0.10]	0.50 [0.50 0.51]	0.90 [0.90 0.91]	0.10	0.50	0.90
K	0.10 [0.10 0.11]	0.49 [0.48 0.50]	0.64 [0.61 0.67]		NA	

Table F.30: Experiment 4.6 conditional probabilities posterior mean and 95% CI

[a]To report these values, we use only intercept, i.e., we assume that covariates $X_{kt} = 0$.

Appendix G

Mixture of Gaussians as a Prior Model

G.1 Full Conditional Distributions

Given the prior structure described in Section 5.1.1 and following Koop [2003], we describe full conditionals distribution of parameters θ_k, Φ, μ_θ^m, Σ_θ^m and π_m.

$$\mathbb{P}\left(\theta_k|\mu_\theta, \Sigma_\theta, \Phi, \{Y_{kt}\}_{t\in T}\right) \propto \exp\left(-\frac{1}{2}(\theta_k - \mu_\theta)^T \Sigma_\theta^{-1}(\theta_k - \mu_\theta)\right) L_k\left(\theta_k, \Phi| \{Y_{kt}\}_{t\in T}\right) \tag{G.1}$$

$$\mathbb{P}\left(\Phi|\mu_\theta, \Sigma_\theta, \{\theta_k\}_{k\in K}, Y\right) \propto \exp\left(-\frac{1}{2}(\Phi - \mu_\Phi)^T \Sigma_\Phi^{-1}(\Phi - \mu_\Phi)\right) L\left(\{\theta_k\}_{k\in K}, \Phi|Y\right) \tag{G.2}$$

$$\mu_\theta \sim \mathcal{N}(\mu_n, V_n) \tag{G.3}$$

$$\Sigma_\theta^{-1} \sim W(df_1, S_1) \tag{G.4}$$

where:

$$V_n = \left(V_0^{-1} + K\Sigma_\theta^{-1}\right)^{-1} \tag{G.5}$$

$$\mu_n = V_n\left(\mu_0 V_0^{-1} + K\bar{\theta}\Sigma_\theta^{-1}\right) \tag{G.6}$$

$$S_1 = \left(S_0^{-1} + \sum_{k=1}^{K}(\theta_k - \mu_\theta)(\theta_k - \mu_\theta)^T\right)^{-1} \tag{G.7}$$

$$\bar{\theta} = \frac{1}{K}\sum_{k=1}^{K}\theta_k \tag{G.8}$$

MCMC procedure generates draws from these distribution using a Gaussian random-walk Metropolis Hastings algorithm with an adaptive step, described in Appendix C, with an acceptance ratio of approximately 20% (see Appendix D).

We set uninformative prior hyperparameters as follows:

$$\begin{aligned}
\mu_0 &= \begin{bmatrix} 0.69 & 0.33 & -0.69 & 0.33 & -0.69 & 0.33 & 0 & \dots & 0 \end{bmatrix} \\
V_0 &= \frac{1}{5}\mathbb{I}_{n_\theta} \\
df_0 &= K + 5 \\
S_0 &= (df_0 - K - 1)\mathbb{I}_{n_\theta} \\
\mu_\Phi &= \begin{bmatrix} 0 & -4 & -4 & 0 & \dots & 0 \end{bmatrix} \\
\Sigma_\Phi &= \mathbb{I}_{n_\Phi}
\end{aligned}$$

G.2 Number of Components

For running a MOGP model the number of components M of the Mixture of Gaussians prior must be set. The correct way of selecting the number of components is to run the model with different values of M and choose the one that performs better (in terms of criteria such as LDM, DIC and MSC); but it is too expensive in terms of computer time and number of models that have to be run.

An alternative option is to use the results of the model with fixed number of states but a single Gaussian as the prior. Using the posterior of θ_k, the number of components M is set fitting Mixture of Gaussians with different values of M and choosing the one that fits better to the posterior distribution of the individual-level parameters. The following tables show which number of components fit best for the 2-state HMM and for the 3-state HMM.

Number of Components	**BIC**	**ΔBIC**
1	-373.37	–
2	**-378.30**	**-4.93**
3	-355.07	23.22
4	-367.91	-12.84
5	-311.20	56.71

Table G.1: Number of Components Selection for 2-State HMM

Number of Components	**BIC**	**ΔBIC**
1	-4588.7	–
2	**-6291.9**	**-1703.2**
3	-6243.1	48.8
4	-6170.7	72.4
5	-6070.3	100.4

Table G.2: Number of Components Selection for 3-State HMM

Table G.1 shows that a Mixture of Gaussians with 2 components is what best fits the individual level parameter posterior distribution estimated by the 2-state HMM and Table G.2 shows that a Mixture of Gaussians with 2 components is what best fits the individual level parameter posterior distribution estimated by the 3-state HMM, both according to the Bayesian Information Criterion (BIC[1]).

[1] $BIC = N_p \ln(k) - 2 \cdot LL$, where N_p is the number of parameters of the distribution, k is the number of observations, and LL is the log-likelihood function.

G.3 Number of States Selection

Table G.3 shows LDM, DIC and MSC for Mixture of Gaussians as a Prior Model with 2 and 3 states. The best model was the 3-state model.

Model	**LMD**	**DIC**	**MSC**
2 states	-71274.45	142634.49	152205.05
3 states	**-30317.22**	**60671.90**	**70311.26**

Table G.3: Model Comparison

Appendix H

Markov chain Monte Carlo algorithm for Latent Class HMM

The Bayesian approach used to estimate the parameters is a hierarchical Bayesian Markov chain Monte Carlo (MCMC) based on the algorithm described in Appendix C. The main difference in this algorithm is the update of parameters Φ^m per class in parallel.

Each iteration of the algorithm consists in updating the parameters value to obtain draws from the posterior distribution, in four major steps:

1. Update Φ^m (Metropolis-Hastings acceptance criterion)
2. Update Σ_θ (Gibbs move from full conditionals)
3. Update μ_θ (Gibbs move from full conditionals)
4. Update $\theta_k = \{\theta_k^m\}_{m=1}^M$ for each k separately (Metropolis-Hastings acceptance criterion)

Given that Φ^m is a parameter across customers, the draws of Φ^m have to be made separately to ensure that every model is improving their parameters. In this model is important because a HMM with more states than other is not only a different model but also is a more general model so the draws could be accepting the moves only if the parameters of the model that is converging faster is improving, and letting the other models behind.

This is not done for θ_k because the movements are made at the individual level are more flexible to adapt to each customer.[1]

A more detailed explanation of each of the steps of the algorithm follows next.

Consider the $(i+1)$'th iteration, and the parameters obtained in the i'th iteration $(\left\{(\theta_k^m)^i\right\}, (\Phi^m)^i, (\mu_\theta)^i, (\Sigma_\theta)^i, (\alpha)^i)$.

[1] A future improvement of this algorithm could be to draw θ_k separately for model

For the following equations, recall that θ_k^m is a $n_\theta(m) = n_m(n_m - 1) \times 1$ vector (where n_m is the fixed number of states of model m) so μ_k is $N_\Theta = \sum_{m=1}^{M} n_\theta(m)$ vector, Φ^m is a $n_m \times 1$ vector and Σ_θ is a $N_\Theta \times N_\Theta$ matrix.

1. Update Φ^m (Metropolis-Hastings acceptance criterion) Let $x^i = \left\{\left\{(\theta_k)^i\right\}_{k\in K}, (\Phi)^i\right\}$ the point where the MCMC is, before updating Φ.
   Let $\left(\Phi^{m'}\right)^{new} = \left(\Phi^{m'}\right)^i \quad \forall m' = 1\dots M$ (i.e. $\Phi^{new} = \Phi^i$).
   For each model m the algorithm computes the following:

   The proposed candidate Φ^C is computed using a random walk as following:

$$\left(\Phi^{m'}\right)^C = \begin{cases} \left(\Phi^{m'}\right)^{new} & \text{if } m' \neq m \\ \left(\Phi^{m'}\right)^{new} + \sigma_{i\phi}^m \cdot z & \text{if } m' = m \end{cases}$$

   with $z \sim \mathcal{N}\left(0, \Lambda_{i\phi}^m\right)$ Note that the movement is only on the parameters of model m, so Φ^{new} and Φ^C only differs on those components. The choices of $\sigma_{i\phi}^m$ and $\Lambda_{i\phi}^m$ are made by the algorithm using the adaptive part described in Appendix D, but using this time every model m is adapted separately.

   Then the acceptance probability is $\alpha_{im\Phi} = \min\left\{1, \alpha_\Phi\left(\Phi^{new}, \Phi^C\right)\right\}$ where:

$$\alpha_\Phi\left(\Phi^{new}, \Phi^C\right) = \frac{L\left(\left\{\theta_k^i\right\}_{k\in K}, \Phi^C | Y\right) e^{\frac{1}{2}\left(\Phi^C - \mu_\Phi\right)^T \Sigma_\Phi^{-1}\left(\Phi^C - \mu_\Phi\right)}}{L\left(\left\{\theta_k^i\right\}_{k\in K}, \Phi^i | Y\right) e^{\frac{1}{2}\left(\Phi^{new} - \mu_\Phi\right)^T \Sigma_\Phi^{-1}\left(\Phi^{new} - \mu_\Phi\right)}}$$

   The likelihood function is the described in Equation 5.3. Let $u \sim \mathcal{U}(0,1)$. If $u < \alpha_{im\Phi}$ then the candidate is accepted, i.e., $\Phi^{new} = \Phi^C$, otherwise is rejected, i.e., $\Phi^{new} = \Phi^{new}$.

   Finally when the movements of all models m are either accpeted or rejected the new value of Φ is stored $\Phi^{i+1} = \Phi^{new}$. Also $x_{new}^i = \left\{\left\{\theta_k^i\right\}_{k\in K}, \Phi^{i+1}\right\}$.$x_{new}^i$ is not the $(i+1)$'th draw of the MCMC given that the movements on θ_k still have to be considered.

2. Update Σ_θ (Gibbs move from full conditionals)
   Let:

$$f_1 = f_0 + n_k$$
$$S_1 = \sum_{k\in K}\left(\theta_k^i - \mu_\theta^i\right)\left(\theta_k^i - \mu_\theta^i\right)^T + {S_0}^{-1}$$

   Then the algorithm just draw from a Wishart distribution with parameters $\left(f_1, {S_1}^{-1}\right)$:

$$\left(\Sigma_\theta^{i+1}\right)^{-1} \sim W\left(f_1, {S_1}^{-1}\right)$$

   (i.e., $\Sigma_\theta^{i+1} \sim IW\left(f_1, S_1\right)$).

3. Update μ_θ (Gibbs move from full conditionals)

   Let:

$$\bar{\theta} = \frac{1}{n_k} \sum_{k \in K} \theta_k^i$$

$$V_1 = \left[V_0^{-1} + n_k \cdot \left(\Sigma_\theta^{i+1} \right)^{-1} \right]^{-1}$$

$$\mu_1 = V_1 \left[V_0^{-1} \cdot \mu_0 + n_k \cdot \left(\Sigma_\theta^{i+1} \right)^{-1} \cdot \bar{\theta} \right]$$

   Then the algorithm just draw from a Multivariate Normal distribution with parameters (μ_1, V_1):

$$\mu_\theta^{i+1} \sim \mathcal{N}(\mu_1, V_1)$$

4. Update θ_k for each k separately (Metropolis-Hastings acceptance criterion)

   For each $k \in K$ the algorithm does the following:

   Propose a candidate using a random walk aproach:

$$\theta_k^c = \theta_k^i + \sigma_{i\theta_k} \cdot z$$

   with $z \sim \mathcal{N}(0, \Lambda_{i\theta_k})$

   The choices of $\sigma_{i\theta_k}$ and $\Lambda_{i\theta_k}$ are made by the algorithm using the adaptive part described in Atchadé 2006.

   Then the acceptance probability is $\alpha_{i\theta_k} = \min\left\{1, \alpha_{\theta_k}\left(\theta_k^i, \theta_k^c\right)\right\}$ where:

$$\alpha_{\theta_k}\left(\theta_k^i, \theta_k^c\right\} = \frac{L_k\left(\theta_k^c, \Phi^{i+1} | \left\{Y_{kt}\right\}_{t \in T}\right) e^{\frac{1}{2}\left(\theta_k^c - \mu_\theta^{i+1}\right)^T \left(\Sigma_\theta^{i+1}\right)^{-1} \left(\theta_k^c - \mu_\theta^{i+1}\right)}}{L_k\left(\theta_k^i, \Phi^{i+1} | \left\{Y_{kt}\right\}_{t \in T}\right) e^{\frac{1}{2}\left(\theta_k^i - \mu_\theta^{i+1}\right)^T \left(\Sigma_\theta^{i+1}\right)^{-1} \left(\theta_k^i - \mu_\theta^{i+1}\right)}}$$

   The individual level likelihood function is the described in Equation 5.2. Let $u \sim \mathcal{U}(0,1)$. If $u < \alpha_{i\theta_k}$ then the candidate is accepted, i.e., $\theta_k^{i+1} = \theta_k^c$, otherwise is rejected, i.e., $\theta_k^{i+1} = \theta_k^i$.

   Finally, when the movements for each k are made, accepted or rejected, the algorithm generated the $(i+1)$'th draw of the posterior distribution: $x^{i+1} = \left\{\left\{\theta_k^{i+1}\right\}_{k \in K}, \Phi^{i+1}\right\}$, μ_θ^{i+1} and Σ_θ^{i+1}.