**Fast Approximate Bayesian Estimation Algorithm for Discrete and Continuous Time Models in Intensive Longitudinal Data**

By

Azizur Rahman

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

College of Community and Global Health

Rady Faculty of Health Sciences

University of Manitoba

Winnipeg

## ABSTRACT

**Introduction:** Intensive longitudinal data (ILD) involves collecting repeated measurements in a frequent, natural, and minimally invasive manner. Analyzing ILD requires advanced statistical models capable of handling both discrete-time (DT) and continuous-time (CT) frameworks. The Multilevel Threshold Autoregressive (ML-TAR) model reveals regime-switching behaviors, which are sudden transitions between distinct psychological, physiological, or behavioral states over time. This provides valuable insights into the dynamic nature of ILD. Meanwhile the Continuous-Time Multivariate Mixed Effects Hidden Markov Model (CT-mMixHMM) effectively addresses complex hidden states in a continuous-time setting, further enriching our understanding of temporal dynamics. Inference in these models is difficult because traditional frequentist methods, such as Maximum Likelihood estimation, require integrating likelihood function over random effects, which is often not feasible. Bayesian estimation methods are particularly suited for these analyses due to their flexibility with complex model structures and their ability to incorporate prior information, making them ideal for handling ILD challenges. However, traditional Bayesian methods, such as Markov Chain Monte Carlo (MCMC), tend to be computationally intensive for both ML-TAR and CT-mMixHMM, presenting significant computational challenges. This research aims to develop and evaluate computationally efficient approximate Bayesian estimation algorithms tailored for DT and CT models in ILD. The objectives are to: (1) construct a Mean-Field Variational Bayes (MFVB) algorithm for estimating parameters in the ML-TAR model to improve its effectiveness in capturing regime-switching dynamics, and (2) develop a Sequential Variational Bayes (SVB) algorithm for the newly introduced CT-mMixHMM, enabling analysis of irregularly spaced observations.

**Methods:** We used optimization-based techniques to approximate posterior distributions. For Objective 1, the MFVB algorithm is developed to estimate regime-switching dynamics in ML-TAR

models using synthetic datasets and is applied to real-world affect regulation data (emotional management under stress). For Objective 2, the SVB algorithm is tailored for CT-mMixHMM and applied to hypertension data from the Manitoba Follow-Up Study (MFUS). The performance of both algorithms are assessed for prediction accuracy (e.g., ability to predict future state switches, hidden state classification, and mean absolute percentage error), Bayesian posterior accuracy (bias, confidence intervals, expected lower bound convergence), scalability, and computational efficiency through simulation studies. The results from both simulation studies and real-world data application were compared against MCMC methods to assess robustness and validity.

**Results:** The first study demonstrated that in case of simulation results our MFVB algorithm for ML-TAR (2,1,1) model is significantly faster than the standard Markov Chain Monte Carlo (MCMC) approach. Additionally, the simulation results revealed that increasing the number of individuals or the number of time points improves the accuracy of parameter estimates using the MFVB method, suggesting that sufficient data are critical for accurate parameter estimation, especially for complex models like multilevel threshold autoregressive models. When the algorithm was applied to real-world data, it significantly outperformed MCMC in speed and accuracy. The findings of the second study indicated that the SVB method for CT-mMixHMM demonstrated superior performance over MCMC in correctly identifying full and partial absorbing states and in estimating transition probabilities, with computational efficiency and stable convergence in both simulated and real data. Across simulations, SVB reliably recovered the true six-state solution across different initializations and sample sizes, with improved parameter accuracy and lower Deviance Information Criterion (DIC) values as sample size increased. In the MFUS hypertension data, age effects feature a stepwise progression through hypertension stages with increasing treatment and mortality risk, emphasizing the importance of early intervention. Individual disease progression profiles varied, with rapid progression linked to higher mortality and effective medication use associated with survival.

**Significance:** This research develops fast approximate Bayesian estimation algorithms for ML-TAR and CT-mMixHMM models. These algorithms efficiently handle complex and large datasets, thereby facilitating the accurate analysis of intensive longitudinal data. By solving computational issues in such models, the study has provided valuable tools for interdisciplinary applications, such as in psychology and healthcare. These advancements support data-driven decision-making, personalized and adaptive interventions, real-time monitoring of dynamic processes, and improved understanding of complex longitudinal patterns across a wide range of applications in health and social sciences.

## ACKNOWLEDGEMENTS

The completion of this doctoral thesis would not have been possible without the invaluable support, guidance, and encouragement of numerous individuals to whom I owe my deepest gratitude.

First and foremost, I extend my sincerest appreciation to my advisor and mentor, Dr. Depeng Jiang, for his unwavering support, exceptional guidance, and steadfast commitment to academic excellence. His mentorship has been instrumental in shaping this research and my growth as a scholar.

I am profoundly grateful to my thesis committee members, Dr. Lisa Lix and Dr. Po Yang, for their insightful feedback, constructive critiques, and invaluable mentorship throughout this journey. Their expertise and encouragement greatly enhanced the quality of this work.

I extend my sincere gratitude to the faculty, staff, and fellow students at the George & Fay Yee Centre for Healthcare Innovation for their support and enriching academic environment. Special thanks to: Dr. Robert Tate and Dr. Phil St. John for their tremendous support in using MFUS data and encouragement; Dr. Monica for her mentorship during my VADA internship at Social Innovation Office; Josie for her administrative support; Yixiu Liu and Md Ashiqul Haque for their help in my candidacy exam; Noor Breik, Muditha Lakmali, Chendong Li, and Shi Zhang for their camaraderie.

This research was supported by the VADA Program and Dr. Depeng Jiang's NSERC Discovery Grant, for which I am deeply grateful. Compute Canada- now the Digital Research Alliance of Canada at Simon Fraser University-supplied high-performance computers for my simulation.

Finally, I thank my wife and son, and friends for their unwavering love and encouragement. To all who contributed, thank you.

## DEDICATION

*To my parents, Mr. Shafiqur Rahman and Mrs. Afia Rahman, whose support and encouragement guided me toward pursuing graduate studies.*

*To my wife, Mariam Akter, whose unwavering emotional and financial support has allowed me to pursue my PhD journey in Canada with confidence and stability.*

*To my son, Fahim Muntasir Rahman, whose constant inspiration motivates me to strive toward my dreams.*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

## 1.1 Overview

Intensive Longitudinal Data (ILD) involves collecting repeated measurements in more intensive, frequent, natural, and less invasive manner. It occupies a unique space between panel data and time series data. Typically, ILD may include fewer individuals (around 50) but with more time points (30 or more) [1]. However, these figures are flexible, as one study, like [2], have used 15 to 20 occasions and up to two hundred subjects. ILD are often encountered in a variety of disciplines, including psychology and healthcare. The richness of ILD makes it particularly suitable for exploring nonlinear dynamics and disease transition dynamics, critical for understanding complex biological and psychological phenomena. Analyzing ILD requires models that can handle both discrete and continuous data while addressing computational challenges due to the data's high dimensionality and complexity.

The discrete-time (DT) framework treats time in discrete steps, suitable for regular time intervals, while the continuous-time (CT) framework models time as a continuous variable, accommodating irregular or sparse observations. For example, DT models have been applied in psychology and behavioral studies to analyze daily mood fluctuations or emotions in individuals over several weeks [2], and CT models have been applied in healthcare and medicine such as in modeling the transition of blood pressure levels over time [3].

This thesis concentrated on two advanced modeling approaches for Intensive Longitudinal Data (ILD): the multilevel threshold autoregressive model (ML-TAR) and the multivariate mixed-effects hidden Markov models (mMixHMMs). First, the ML-TAR was explored under discrete-time (DT) framework. The threshold autoregressive (TAR) model captured regime-switching behaviors, representing sudden switches between distinct psychological, physiological, or behavioral states over time [4]. The model identified threshold effects, where the relationship between variables changes

depending on the level of a threshold variable. In a TAR model, a "threshold" refers to a specific value of a chosen variable, often a lagged value of the dependent variable, that determines which set of model parameters applies, thus delineating different dynamic behaviors or "regimes". For instance, a threshold could be a particular level of stress that, when exceeded, drastically alters mood regulation [2]. The ML-TAR model extends traditional TAR models by incorporating multiple hierarchical levels, such as individual-level and group-level effects, to capture complex behavioral processes like affect regulation [2]. However, determining the optimal threshold value can be computationally intensive.

The second focus was on multivariate mixed effects hidden Markov models (mMixHMMs) under CT framework for ILD. These CT-mMixHMMs extend traditional Hidden Markov Models (HMMs) [5] by modeling transitions in continuous time instead of DT steps, incorporating both fixed effects and random effects to account for individual-level heterogeneity in both transition and observation processes. These models facilitated the study of hidden states, such as unobserved health conditions, making them ideal for tracking latent conditions like hypertension severity. The importance of such a model lies in its ability to handle biological processes, where measurements are collected at irregular intervals, making discrete-time HMMs impractical [4]. Estimating CT transition rates and random effects in CT-mMixHMMs requires sophisticated numerical estimation methods.

Inference in models such as ML-TAR and CT-mMixHMM under frequentist methods (e.g., maximum likelihood) is challenging because the likelihood requires the integration over random effects, which is often intractable. Moreover, these approaches provide point estimates but do not provide full posterior distributions over the parameters. By contrast, Bayesian estimation provides full posterior distributions by delivering a complete quantification of parameter uncertainty. Moreover, it manages complex data structures with intricate dependencies between repeated measurements that usually occur in ILD. Finally, Bayesian methods allow flexible priors on model parameters, reducing the reliance

solely on the observed data.

However, existing Bayesian estimation techniques, such as Markov Chain Monte Carlo (MCMC), face computational inefficiencies for large-scale ILD data and for complex models such as ML-TAR and CT-mMixHMM. To address these challenges, this research proposed approximate Bayesian estimation algorithms for ML-TAR and CT-mMixHMM. The aim is to achieve computational efficiency without sacrificing accuracy, providing a more feasible and insightful approach to modeling ILD.

### 1.2 Purpose and Objectives

The purpose of this research was to develop and evaluate computationally efficient approximate Bayesian estimation algorithms for DT and CT models of ILD. This research pursued two objectives:

(1) To develop and evaluate an algorithm for estimating parameters of ML-TAR models in discrete-time ILD to address computational issues and compare its performance with existing estimation methods; and

(2) To construct and evaluate an SVB algorithm for estimating parameters of CT-mMixHMM in continuous-time ILD and compare its performance with existing estimation methods.

### 1.3 Organization of Thesis

This thesis has included two related manuscripts addressing the stated objectives. Chapter 2 reviewed discrete and continuous time models for intensive longitudinal data (ILD), detailing their features and current estimation methods.

Chapter 3 contains Manuscript 1, "*Efficient and accurate variational inference for multilevel threshold autoregressive model in intensive longitudinal data*." It assessed the performance of the mean field variational Bayes (MFVB) algorithm for ML-TAR model parameter estimation, comparing it with

MCMC on simulated and real data. The analysis found that MFVB algorithm demonstrated improved performance over traditional MCMC methods. The key findings revealed a substantial improvement in the coverage rates of parameter estimates- across both simulated and real-data applications - when using the MFVB algorithm. These gains became more pronounced as sample size and the number of measurement occasions increased, with MFVB outperforming the MCMC method under comparable conditions. Furthermore, the MFVB algorithm demonstrated a significant computational advantage, operating approximately 21 times faster than MCMC when applied to large-scale datasets.

Chapter 4 contains Manuscript 2, "*Approximate Bayesian Estimation Algorithm for Continuous Time Multivariate Mixed Hidden Markov Model in Intensive Longitudinal Data*". It has introduced and evaluated a Sequential Variational Bayes (SVB) algorithm, specifically the recursive variational Gaussian approximation for dependent mixtures (R-VGADM). The study assessed the performance of the proposed R-VGADM algorithm for parameter estimation in CT-mMixHMM models, benchmarking it against Markov Chain Monte Carlo (MCMC) methods using both simulated data and the Manitoba Follow Up Study (MFUS) dataset. In MFUS, systolic and diastolic blood pressure measurements were used to identify latent hypertensive states, with age included as a covariate to examine its effect on transition rates at both the population and individual levels. Findings demonstrate that the R-VGADM algorithm for CT-mMixHMM exhibited superior performance compared to MCMC, in identifying both full and partial absorbing states and in modeling transition probabilities, while also demonstrating greater computational efficiency and stable convergence in both simulated and real-world data analyses. To better reflect real-world data scenarios, we generated simulated datasets using the CT-mMixHMM model with six states. In simulation studies, the SVB approach consistently recovered the correct six-state solution across varying initializations and sample sizes, yielding improved parameter accuracy and lower Deviance Information Criterion (DIC) values as sample size increased. In the real data application,

age was associated with a stepwise progression through hypertension stages, implying increased treatment and mortality risk and underscoring the importance of early intervention. Individual disease progression profiles varied, with rapid progression linked to higher mortality and effective medication use associated with survival.

Chapter 5 summarizes the key research findings, discusses strengths and limitations, and outlines the main conclusions and directions for further research.

**References**


1. L. Zhou, M. Wang, and Z. Zhang, "Intensive longitudinal data analyses with dynamic structural equation modeling," *Organizational Research Methods*, vol. 24, no. 2, pp. 219–250, 2021, doi: 10.1177/1094428119833164.
2. S. De Haan-Rietdijk, J. M. Gottman, C. S. Bergeman, and E. L. Hamaker, "Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation," *Psychometrika*, vol. 81, no. 1, pp. 217–241, 2016, doi: 10.1007/s11336-014-9417-x.
3. X. Zheng, J. Xiang,, Y. Zhang, et al., "Multistate Markov Model application for blood-pressure transition among the Chinese elderly population: a quantitative longitudinal study," BMJ Open, 12:e059805, 2022, , doi:10.1130/bmjopen-2021-059805.
4. R. S. Tsay, "Testing and modeling threshold autoregressive processes," *Journal of the American Statistical Association*, vol. 84, pp. 231–240, 1989.
5. D. J. Raff and J. A. Dubin, "Multivariate longitudinal analysis with mixed effect HMMs," *Biometrics*, 2015, doi: 10.1111/biom.12096.

# CHAPTER 2: LITERATURE REVIEW OF DISCRETE AND CONTINUOUS TIME MODELS AND THEIR ESTIMATION METHODS FOR INTENSIVE LONGITUDINAL DATA

## 2. Overview

This chapter describes the background and literature review of existing work on statistical methods and models of intensive longitudinal data under discrete and continuous time framework.

ILD are commonly encountered in psychology, epidemiology, biostatistics, public health, and the social sciences, where researchers seek to understand dynamic processes such as symptom evolution, behavioral change, disease progression, and treatment response at the individual level [1,2,3]. ILD poses a challenge to the traditional assumption of independent observations, potentially leading to inaccuracies when using conventional methods such as naïve regression, fixed effects regression and regression of summary measures. These methods either underestimate or overestimate the standard error of parameter estimates, resulting in inflated type I or type II error rates [3], particularly when there is substantial outcome dependency within the same individual [4]. While growth curve models and mixed effect models are commonly preferred for analyzing longitudinal panel data in various fields, they may not be suitable for ILD due to several reasons [5]. Firstly, growth curve models and mixed effect models may become computationally burdensome for longer longitudinal data due to the nonlinearity and multiple random effects. Secondly, with the intensive longitudinal design, the modeling focus has shifted from stable processes to dynamic processes [6]. Therefore, there is a need to enhance statistical models for ILD analysis.

Broadly, methodological approaches for ILD can be categorized into discrete-time (DT) and continuous-time (CT) modeling frameworks. Discrete-time models assume that the underlying process evolves in discrete steps indexed by observation occasions, whereas continuous-time models

conceptualize the data-generating mechanism as evolving continuously over time, independent of observation schedules. Each framework offers distinct advantages and limitations, and the choice between them has important implications for inference, interpretation, and generalizability. This section reviews the major developments in DT and CT models for ILD, highlighting their theoretical foundations, extensions, and applications.

## 2.1 Discrete time models for intensive longitudinal data

### 2.1.1 *Conceptual Framework*

Discrete-time models treat time as a sequence of ordered measurement occasions and assume that changes in the outcome occur only between these discrete points. This framework is intuitive and aligns naturally with many data collection designs, particularly when observations are approximately equally spaced. As a result, DT models have historically dominated the analysis of ILD [3].

### 2.1.2 *Autoregressive and Time-Series Models*

The analysis of intensive longitudinal data (ILD) collected at discrete time framework requires specialized statistical models that can capture the dynamic nature of the data, such as time series analysis, multilevel modeling, and dynamic system modeling. Autoregressive (AR) models and their multivariate extensions, such as vector autoregressive (VAR) models, are among the most widely used DT approaches for modeling temporal dependence in ILD [3]. These models capture within-person dynamics by relating current observations to lagged values of the same or other variables [7]. In ILD contexts, AR and VAR models are often extended to multilevel formulations to distinguish within-person temporal effects from between-person differences [8]. Dynamic multilevel models and dynamic structural equation models (DSEM) integrate time-series components within hierarchical modeling frameworks, allowing researchers to model individual-specific dynamics while borrowing strength across subjects [52]. These models have been particularly influential in psychological and behavioral

research using EMA data [2].

Autoregressive models typically assume a fixed autoregressive parameter commonly known as “inertia” for an individual, but this assumption is often invalid due to the model's failure to consider evolutions between different states of a process, known as "regime-effect" [10, 11]. Non-linear dynamics, which involve behavior changing based on context or state, are prevalent across various research fields. For example, a study on affect regulations discovered that the autoregressive parameter, a metric of affect regulation, could fluctuate depending on the current affect level [12]. This research focused on DT models designed to identify dynamic and non-linear patterns within individual outcome processes.

### 2.1.3 *Threshold Autoregressive Model*

The TAR model introduced by [13] was a pioneering attempt to explain non-linear oscillations (such as harmonic or other periodic variation) and jump phenomena (such as discontinuities in system response) in time series. TAR models incorporate threshold points or turning points that act as a switching mechanism. The developmental process comprises multiple segments, delineated by these turning points. Each segment before or after a turning point is associated with an autoregressive (AR) model. In the TAR model, if a previous observation of the outcome process falls within the $m^{th}$ segment, the $m^{th}$ AR model generates the current outcome observation.

Since [13]'s seminal work in 1980, there have been significant advancements in threshold autoregressive (TAR) models and its estimation process. [10] outlines various special cases of TAR models, with the most used being the self-exciting TAR (SETAR) model. The SETAR model is characterized by two key parameters: a positive integer *d*, referred to as the delay parameter, and a real threshold value *r*. Most existing approaches focus on individual time series TAR models, limiting the generalizability of results to a broader population [16, 17]. Conversely, ILD collected from multiple

individuals can provide insights into both intra-individual dynamics and inter-individual variations. To achieve this, it is essential to extend the TAR model to a multilevel framework, allowing TAR model parameters to vary across individuals. For example, [14] introduced the multilevel extension of the SETAR model to investigate affect regulation. Several authors have delved into the parameter estimation and model specifications of TAR models [15-21]. [15] utilized arranged autoregression to generate predictive residuals and visually identified threshold values through scatter plots. Despite several developments in TAR models, available Bayesian estimation methods in this direction are still limited and under development.

A key limitation of DT models is their dependence on the chosen time scale. Parameter estimates and interpretations can change substantially with different discretization of time, complicating comparisons across studies or individuals with different observation schedules [53]. Moreover, DT models may yield biased estimates when observations are irregularly spaced or when the underlying process evolves continuously [54]. These limitations have motivated increased interest in continuous-time modeling frameworks for ILD.

## 2.2 Continuous time models for intensive longitudinal data

### 2.2.1 *Conceptual Framework*

Continuous-time models assume that the underlying process evolves continuously over time, regardless of when observations are collected. Observations are viewed as snapshots of this latent process, taken at possibly irregular time points. This perspective is particularly well suited for ILD, where measurement occasions often vary across individuals and are not under full experimental control.

### 2.2.2 *Stochastic Differential Equation Models*

Many CT models are formulated using stochastic differential equations (SDEs), which describe instantaneous rates of change in the latent process. Classical examples include the Ornstein–Uhlenbeck

process, which forms the basis of continuous-time autoregressive models [54,55]. These models allow parameters to be interpreted as drift and diffusion rates, offering time-scale–invariant descriptions of dynamics. Continuous-time structural equation models (CT-SEM) extend this framework to multivariate latent variable systems, enabling the modeling of dynamic relationships among multiple constructs [55]. CT-SEM has been increasingly applied to ILD in psychology and social sciences due to its ability to handle irregular sampling and disentangle temporal dynamics from measurement design [55].

### 2.2.3 *Continuous-Time Hidden Markov Models*

Many population studies now include longitudinal follow-up and mortality data, allowing estimation of transitions from healthy to unhealthy states before death [22,31]. For example, high blood pressure is a significant risk factor and leading cause of death from cardiovascular disease [31]. The measurement in systolic blood pressure (SBP) greater than 140mmHg or diastolic blood pressure greater than 90mmHg is often used as the cut point for identifying not hypertensive vs hypertensive condition of a patient. Instead of using thresholds or cutoffs to define disease status, it may be more efficient to model the data as generated from these hidden or latent disease condition.

This research focused on CT models designed to simultaneously reveal the dynamic association and serial heterogeneity of an ILD process. To account for longitudinal data features, including serial heterogeneity, parametric models known as Hidden Markov models (HMMs) [16] are well suited alternative approaches. HMMs describe the relationship between two stochastic processes, namely, an observed outcome process and an unobservable finite-state transition process. They assume that an unobserved stochastic sequence governs the observations, and the assumptions of the Markov property are imposed on the unobserved sequence. The application of HMMs is justified by their versatility and mathematical tractability: the availability of all moments; the likelihood computation is linear in the number of observations; the marginal distributions are easy to determine, and missing observations can

be handled with minor effort; the conditional distributions are available, outlier identification is possible and forecast distributions can be calculated [17]. HMMs have attracted significant attention from various disciplines [18- 23]. However, the preceding works on HMMs have focused on longitudinal responses that are observed at regular times. In many cases during observational and clinical trials, observation times are irregular or continuous.

Continuous-time Hidden Markov Models (CT-HMMs), also known as continuous-time multi-state models, generalize DT-HMMs by allowing latent state transitions and the arrival of observations to occur at arbitrary time points [24, 25]. Transitions are governed by an intensity (or generator) matrix, which specifies instantaneous transition rates between states [55]. CT-HMMs are widely used in medical and public health research to model disease progression, treatment pathways, and survival outcomes [46-51,55]. These models naturally accommodate intermittent observation times and enable meaningful comparisons across individuals and studies with different follow-up schedules [26]. Extensions incorporating covariates, random effects, and misclassification further enhance their applicability to complex ILD settings [54].

CT models offer several advantages over DT models, including invariance to observation schedules, improved interpretability of temporal parameters, and reduced discretization bias [55]. However, these benefits come at the cost of increased mathematical and computational complexity. Likelihood evaluation often requires matrix exponentiation or numerical integration, and estimation can be challenging for high-dimensional or nonlinear models [55].

Despite several early advancements in continuous-time HMMs, available estimation methods in this direction are still limited and under progress. [27, 28] used a time-binning step for irregularly spaced longitudinal data, leading to artificially equidistant data that could be modeled with discrete time HMMs. [29] conducted estimation for stochastic volatility models using structured HMMs and showed that

discrete HMMs can be used to approximate their continuous counterpart. Efficient learning methods that can scale to large state spaces are still limited and under development for CT-HMMs. [30] studied efficient learning method of a continuous time HMM for disease progression where they proposed EM-based learning methods. Their models allowed a large number of states to be defined roughly by some observed surrogate covariates, which may not be the case in practice. Recently, [31] proposed an efficient maximum likelihood (ML) based learning method in their study that utilized the formula for differentiating and integrating exponential matrices of hidden Markov process-based transition pattern with an unknown number of states for CT-HMMs. Much of these works on continuous-time HMMs have been devoted to frequentist approaches, with inference via the expectation-maximization (EM) algorithm [32, 33]. All these studies considered CT-HMMs for individual outcome process and did not explore unit-specific heterogeneity as a random variable which is a special aspect of modelling ILD. However, [20] developed a discrete time HMM for multiple processes and considered the generalized linear mixed effect model (GLMM) framework, making it possible to incorporate covariates associated with each observation as fixed and random effects. Although several methods have been proposed to model between individual variability, the estimation of individual specific random effects on both components of the CT-HMMs via traditional maximum likelihood methods tends to be computationally intractable, possibly limiting a more widespread utilization of these models.

**2.3 Bayesian Estimation for DT model of ILD**

[34] and [35] proposed Bayesian analysis of two-regime TAR models in the determination of threshold values. [36] introduced a Bayesian estimation method for identifying threshold values for TAR models using a posterior probability plot. [37] proposed the Gibbs sampling technique to derive the marginal densities of threshold values by transforming the TAR model into a change point problem through arranged regression. [38] presented a Bayesian stochastic search selection approach for multiple

threshold issues, employing a hybrid Markov Chain Monte Carlo (MCMC) method to estimate all model parameters. Recently [14], proposed Bayesian estimation approach under MCMC algorithm for ML-TAR model to study the affect regulation among adults but did not thoroughly explore the computational challenges associated with their models.

### 2.4 Bayesian Estimation for CT model of ILD

Few studies have explored the Bayesian approach. Recently, [39] applied a Bayesian estimation approach with MCMC for CT-HMMs to study dementia progression, using restrictions on the parameter space to identify the parameters in the HMM. However, they did not address the estimation process with a large number of unknown hidden states due to high computational demands. A fully Bayesian CT-HMMs using different missing data likelihood formulations for the underlying Markov chain by constructing reversible jump MCMC algorithm was developed to determine the progression of COPD patients in greater Montreal area, Canada through identifying and modeling the transition of different states [40]. More recently, [41] developed Bayesian CT hidden Markov model with covariate selection with measurement error via MCMC approach to identify risk factors of incentivizing abstinence from cigarette smoking among socioeconomically disadvantaged adults. [42] extended the study of [20] in case of DT mixed effects hidden Markov models (HMMs) under fully Bayesian framework for analyzing multivariate longitudinal data in the context of smoking cessation trials.

### 2.5 Gaps in the literature

Reviews of optimization-based approximate Bayesian estimation studies on DT and CT models that include both fixed effects and random effects, specifically focusing on the nonlinear dynamics and state transition dynamics of ILD datasets, are scarce in the literature.

Unlike traditional Bayesian methods, optimization-based approaches such as variational Bayes (VB) do not rely on sampling. Instead, they assume independence among latent variables and iteratively

optimize posterior distributions. This methodology makes them more scalable than approaches like MCMC and exact Bayesian inference, offering a computationally efficient alternative for complex models [43]. Scalability here refers to an algorithm's ability to maintain computational efficiency as data size or model complexity increases. This can be measured through factors such as computational complexity, convergence speed (e.g., seconds per iteration), memory usage, and runtime growth when dataset size increases. Based on data processing, two types of algorithms are commonly used in variational Bayes studies: the mean field variational Bayes (MFVB) algorithm and the sequential variational Bayes (SVB) algorithm. Common limitations of the studies developed by these algorithms are the use of linear functional forms and fixed effects terms in DT models for ILD data sources. For example, a mean field variational Bayes (MFVB) algorithm was developed for DT models such as Generalized Autoregressive (GAR) models, employing a Gaussian mixture noise framework for robust estimation of coefficients [45]. It demonstrated that the VB algorithm prevents overfitting, provides model order selection for both AR and noise model orders, and offers superior performance compared to MCMC approach in capturing linear dynamics, with applications shown in EEG data [45]. To our knowledge there are no studies in literature that take advantage of using VB approaches in DT models such as ML-TAR model and capturing nonlinear dynamics for ILD datasets.

An approximate Bayesian approach was developed to ease the computation for CT-HMMs and demonstrated its performance in the disease progression dynamics in real-world data application like coronary allograft vasculopathy [46]. Moreover, studies that have developed optimization algorithms for CT-HMMs with fixed effect terms have used batch-based data processing. For example, a batch-based MFVB algorithm has been developed for large-scale inference of unknown hidden states, with applications in active learning tasks that require processing all data simultaneously [44]. To the best of our knowledge, no studies in literature have introduced CT-HMM models incorporating both fixed and

random effects to account for unit-specific heterogeneity in the state transition dynamics of outcome process using ILD datasets. Moreover, research has yet to address continuous-time mixed effects hidden Markov models (MHMMs) for multivariate data within the generalized linear mixed effects model (GLMM) framework. Furthermore, the development of a sequential Bayesian estimation algorithm-which could enable large-scale state inference and address the computational complexities and dependency issues associated with MHMM model parameters-has yet to be explored. To date, there appears to be no published research addressing the sequential variational Bayesian approach for continuous-time multivariate mixed-effects hidden Markov models (CT-mMixHMM), making this an area yet to be explored in the literature.

## 2.6 Conclusions

In conclusion, chapter two of this literature review highlights the significant advancements and ongoing challenges in modeling ILD for health research, particularly with respect to disease progression and behavioral risk factors. The literature discusses the Multi-Level Threshold Autoregressive (ML-TAR) model, which has been explored using Bayesian estimation methods under MCMC algorithms, such as those proposed in [14] for studying affect regulation among adults. However, these studies have not thoroughly examined the computational challenges inherent in such models. The ML-TAR framework provides another avenue for modeling nonlinear dynamics and regime changes in ILD datasets, further highlighting the importance of developing efficient computational techniques.

Moreover, the literature revealed notable gaps in the development of algorithms that address unit-specific heterogeneity, nonlinear dynamics, and large-scale inference for ILD. While Hidden Markov Models (HMMs) and their continuous-time extensions (CT-HMMs) have shown great promise in handling irregularly sampled data and capturing dynamic associations in ILD, much of the current research remains focused on discrete-time frameworks and frequentist methodologies. Bayesian

approaches, including MCMC and optimization-based methods like variational Bayes, offer scalable alternatives but are still underutilized in models that account for both fixed and random effects, especially in continuous-time settings. Recent efforts have begun to explore Bayesian estimation and approximate methods, yet comprehensive solutions for multivariate mixed effects continuous-time HMMs remain an open area for exploration. This chapter identified these gaps and argued for developing sequential approximate Bayesian methods for CT-mMixHMMs, which have the potential to advance the scalability and applicability of ILD modeling in health databases and beyond.

By tackling the computational challenges associated with parameter estimation in both DT and CT models, this research can develop a comprehensive methodological toolkit that significantly advances the statistical analysis of ILD.

## 2.7 Reference


1. L. M. Collins, "Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model," *Annual Review of Psychology*, vol. 57, pp. 505–528, 2006.
2. N. Bolger,, and J. P. Laurenceau, *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press, 2013
3. T. A. Walls and J. L. Schafer, *Models for Intensive Longitudinal Data*. Oxford, U.K.: Oxford University Press, 2012, doi: 10.1093/acprof:oso/9780195173444.001.0001.
4. M. Moerbeek, G. J. P. Van Breukelen, and M. P. F. Berger, "A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies," *Journal of Clinical Epidemiology*, vol. 56, pp. 341–350, 2003, doi: 10.1016/S0895-4356(03)00007-6.
5. B. E. Wampold and R. C. Serlin, "The consequence of ignoring a nested factor on measures of effect size in analysis of variance," *Psychological Methods*, vol. 5, no. 4, pp. 425–433, 2000.
6. K. J. Grimm, P. Davoudzadeh, and N. Ram, "Developments in the analysis of longitudinal data," *Monographs of the Society for Research in Child Development*, vol. 82, no. 2, pp. 46–66, 2017, doi: 10.1111/mono.12298.
7. P. C. M. Molenaar, K. O. Sinclair, M. J. Rovine, N. Ram, and S. E. Corneal, "Analyzing developmental processes on an individual level using nonstationary time series modeling," *Developmental Psychology*, vol. 45, no. 1, pp. 260–271, 2009, doi: 10.1037/a0014170.

8. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
9. E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. New York, NY, USA: Wiley, 1988.
10. H. Tong, "Threshold models in time series analysis: 30 years on," *Statistics and Its Interface*, vol. 4, pp. 107–118, 2011.
11. K. S. Chan and H. Tong, "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, vol. 7, no. 3, pp. 179–190, 1986, doi: 10.1111/j.1467-9892.1986.tb00501.x.
12. E. Hamaker and M. Wichers, "No time like the present: Discovering the hidden dynamics in intensive longitudinal data," *Current Directions in Psychological Science*, vol. 26, no. 1, pp. 10–15, 2017.
13. H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society: Series B*, vol. 42, no. 3, pp. 245–292, 1980.
14. S. De Haan-Rietdijk, J. M. Gottman, C. S. Bergeman, and E. L. Hamaker, "Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation," *Psychometrika*, vol. 81, no. 1, pp. 217–241, 2016, doi: 10.1007/s11336-014-9417-x.
15. R. S. Tsay, "Testing and modeling threshold autoregressive processes," *Journal of the American Statistical Association*, vol. 84, pp. 231–240, 1989.
16. W. Zucchini and I. L. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2009.

17. A. Maruotti, “Mixed hidden Markov models for longitudinal data: An overview,” *International Statistical Review*, vol. 79, no. 3, pp. 427–454, 2011, doi: 10.1111/j.1751-5823.2011.00160.x.
18. P. S. Albert, H. F. McFarland, M. E. Smith, and J. A. Frank, “Time series for modelling counts from a relapsing–remitting disease,” *Statistics in Medicine*, vol. 13, no. 5, pp. 453–466, 1994, doi: 10.1002/sim.4780130509.
19. O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, 2nd ed. New York, NY, USA: Springer, 2005.
20. R. M. Altman, “Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 201–210, 2007.
21. E. H. Ip, W. J. Rejeski, T. B. Harris, and S. B. Kritchevsky, “Partially ordered hidden mixed Markov model for the disablement process of older adults,” *Journal of the American Statistical Association*, vol. 108, pp. 370–384, 2013.
22. F. Bortolucci and A. Farcomeni, “A shared-parameter continuous-time hidden Markov and survival model,” *Statistical Methods in Medical Research*, vol. 38, pp. 1056–1073, 2019.
23. K. Kang, J. Cai, X. Song, and H. Zhu, “Bayesian hidden Markov model for Alzheimer disease,” *Statistical Methods in Medical Research*, vol. 28, pp. 2112–2124, 2019.
24. D. R. Cox and H. D. Müller, *The Theory of Stochastic Processes*. London, U.K.: Chapman & Hall, 1965.
25. C. H. Jackson, “Multi-state models for panel data: The msm package,” *Journal of Statistical Software*, vol. 38, no. 8, pp. 1–29, 2011, doi: 10.18637/jss.v038.i08.

26. N. Bartolomeo, P. Trerotoli, and G. Serio, “Progression of liver cirrhosis to HCC,” *BMC Medical Research Methodology*, vol. 11, Art. no. 38, 2011.
27. F. Bartolucci and F. Pennoni, “Latent Markov models for capture–recapture data,” *Biometrics*, vol. 63, pp. 568–578, 2007.
28. F. Bartolucci, G. E. Montanari, and S. Pandolfi, “Dimensionality of latent structure via multidimensional IRT,” *Psychometrika*, vol. 77, pp. 782–802, 2012.
29. R. Langrock, I. L. McDonald, and W. Zucchini, “Nonstandard stochastic volatility models,” *Journal of Empirical Finance*, vol. 19, pp. 147–161, 2012.
30. M. Liu et al., “Transportability of smoking status detection modules,” *AMIA Annual Symposium Proceedings*, pp. 577–586, 2012.
31. X. Zhou, K. Kang, and X. Song, “Two-part hidden Markov models,” *Statistical Methods in Medical Research*, vol. 39, pp. 1801–1816, 2020.
32. C. H. Jackson and L. D. Sharples, “Hidden Markov model for bronchiolitis obliterans syndrome,” *Statistics in Medicine*, vol. 21, pp. 113–128, 2002.
33. J. Lange et al., “Joint multi-state disease models,” *Biometrics*, vol. 71, pp. 821–831, 2015.
34. P. Cook and L. D. Broemeling, “Bayesian analysis of threshold autoregressions,” *Advances in Econometrics*, vol. 11, pp. 89–108, 1996.
35. J. Geweke and N. Terui, “Bayesian threshold autoregressive models,” *Journal of Time Series Analysis*, vol. 14, no. 5, pp. 441–454, 1993.
36. R. E. McCulloch and R. S. Tsay, “Bayesian analysis of autoregressive time series,” *Journal of Time Series Analysis*, vol. 15, no. 2, pp. 235–250, 1994.
37. C. W. S. Chen and J. C. Lee, “Bayesian inference of threshold autoregressive models,” *Journal of Time Series Analysis*, vol. 16, no. 5, pp. 483–492, 1995.

38. J. Pan, Q. Xia, and J. Liu, "Bayesian analysis of multiple thresholds autoregressive model," *Computational Statistics*, vol. 32, pp. 219–237, 2012.

39. J. P. Williams et al., "Bayesian multi-state hidden Markov models," *Journal of the American Statistical Association*, 2019.

40. J. R. Stephen and D. P. Will, "Variational Bayes for Generalized Autoregressive Model," *IEEE Transaction on Signal Processing*, vol. 50, no. 9, pp. 2245-2257, 2002.

41. Y. Luo et al., "Bayesian latent multi-state modeling," *Biometrics*, vol. 77, pp. 78–90, 2020.

42. M. Liang et al., "Bayesian continuous-time hidden Markov models," *Psychological Methods*, vol. 27, no. 6, pp. 1007–1024, 2022.

43. D. J. Raff and J. A. Dubin, "Multivariate longitudinal analysis with mixed effect HMMs," *Biometrics*, 2015, doi: 10.1111/biom.12096.

44. T. P. Minka, *A Family of Algorithms for Approximate Bayesian Inference*, Ph.D. dissertation, MIT, 2001.

45. S. Ji, B. Krishnapuram, and L. Carin, "Variational Bayes for continuous hidden Markov models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 522–532, 2006.

46. A. Tancredi, "Approximate Bayesian inference for discretely observed continuous-time multi-state models," *Biometrics*, vol. 75, no. 3, pp. 966–977, 2019, doi: 10.1111/biom.13019.

47. E. Hamaker, L., Kuiper, and R. P. Grasman, "A critique of the cross-lagged panel model," *Psychological Methods*, **20**(1), 102–116,2015.

48. J. H. L. Oud, and M. J. M. H. Delsing, "Continuous-time modeling of panel data by means of SEM,". *Developmental Psychology*, **46**(1), 233–246,2010,

49. A. R. Bergstrom, "Continuous time stochastic models and issues of aggregation over time," *Handbook of Econometrics*, **2**, 1145–1212,1984.

50. M. C. Voelkle, and J. H. L. Oud, "Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes," *British Journal of Mathematics and Statistics in Psychology*. 66, 103–126, 2013, doi: 10.1111/j.2044-8317.2012.02043.x.

51. C. C. Driver, J. H. L. Oud, and . C. Voelkle, "Continuous time structural equation modeling with R package ctsem,". *Journal of Statistical Software*, **77**(5), 1–35, 2017.

52. C. H. Jackson, "Multi-state models for panel data: The msm package for R," *Journal of Statistical Software*, **38**(8), 1–28, (2011).

53. C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto, "Multistate models for disease progression with classification error,". *Journal of the Royal Statistical Society: Series D*, **52**(2), 193–209, (2011).

54. J. D. Kalbfleisch, and J. F. Lawless, "The analysis of panel data under a Markov assumption,". *Journal of the American Statistical Association*, **80**(392), 863–871,1985.

55. C. Proust-Lima, V. Philipps, and B. Liquet, "Estimation of extended mixed models using latent classes,". *Journal of Computational and Graphical Statistics*, **26**(2), 231–244, (2017).

# CHAPTER 3: EFFICIENT AND ACCURATE VARAIATIONAL INFERENCE FOR MULTILEVEL THRESHOLD AUTOREGRESSIVE MODEL IN INTENSIVE LONGITUDINAL DATA

## 3.1 Overview

This chapter provides an algorithm for obtaining efficient and computationally fast parameter estimates of the selected specialized discrete time model for ILD. In particular, we considered multilevel threshold autoregressive model (ML-TAR) that could capture the "regime-effect" or phase change from one process state to another and construct mean field variational Bayes (MFVB) approximation algorithm in estimating the parameters of ML-TAR model with substantial computational savings in comparison with standard Markov chain Monte Carlo (MCMC) approach under Bayesian framework. This study evaluated the proposed estimation algorithm through proper simulation conditions, and its efficacy has been shown by an application to a dataset on Physical Activity and Affect Regulation Intensive Longitudinal Study [1].

The methodological innovation lies in the comparative evaluation of MCMC and MFVB algorithms both for simulated and real-world data application, offering insights into performance trade-offs across bias, coverage rates and computation time (in seconds). By demonstrating improved accuracy of parameter estimates using MFVB algorithm relative to MCMC in case of increased sample size and time points for simulated data, this work contributes to a growing body of literature seeking to large scale inference in discrete time models of ILD. When the proposed algorithm was applied to Physical Activity and Affect Regulation Intensive Longitudinal Study [1], it significantly outperformed MCMC in speed and accuracy of estimating affect regulation parameters. In addition, this study found individuals with higher levels of physical activity exhibit a lower threshold for negative affect score. In summary, this study offers a notably faster and consistent alternative algorithm to MCMC for intricate and large-scale inference in ILD models,

emphasizing its potential as a fast and efficient tool for data analysis.

*Publication Details*

A. Rahman, D. Jiang, and L.M. Lix, "Efficient and accurate variational inference for multilevel threshold autoregressive model in intensive longitudinal data," *British Journal of Mathematical and Statistical Psychology*, 78(3):785-803, 2025; doi: 10.1111/bmsp.12381.

**3.2 Abstract**

Recent technological advancements have facilitated the collection of ILD, comprising repeated measurements from the same individual. The threshold autoregressive (TAR) model is commonly utilized to capture the dynamic outcome process in ILD, with autoregressive parameters varying based on outcome variable levels. For ILD data from multiple individuals, multilevel TAR models have been proposed, with a Bayesian approach recommended for parameter estimation. However, fitting multilevel TAR models can present computational challenges. In this study, we introduce a mean field variational Bayes (MFVB) algorithm as an alternative to traditional Bayesian inference. By optimizing to approximate posterior densities, variational Bayes aims to find the best approximation within a defined set of distributions rather than asymptotically converging to the true posterior. Simulation results demonstrated that our MFVB algorithm is significantly faster than the standard Markov Chain Monte Carlo (MCMC) approach. Additionally, the simulation results reveal that increasing the number of individuals or the number of time points improves the accuracy of parameter estimates using the MFVB method, suggesting that sufficient data are critical for accurate parameter estimation, especially for complex models like multilevel threshold autoregressive models. When the algorithm was applied to real-world data, it significantly outperformed MCMC in speed and accuracy. In summary, the MFVB algorithm offers a notably faster and consistent alternative to MCMC for intricate and large-scale inference

in ILD models, underscoring its potential as a swift and efficient tool for data analysis.

### 3.3 Introduction

Recent advances in data collection methods like smartphone apps, wearables, and EMA have made Intensive Longitudinal Design (ILD) more common. These tools allow frequent data capture-often 15 to 20 or more observations per person. ILD provides an opportunity to examine intra-individual change variability, and relationship over time [2, 3].

However, the richness and density of ILD bring unique statistical challenges. Unlike traditional longitudinal data, ILD often violates the assumption of independent observations, a fundamental assumption of many conventional statistical methods. Using traditional methods such as naïve regression, fixed effect regression, or the regression of summary measure, this lack of independence can lead to significant biases in parameter estimates. These biases can either underestimate or overestimate the standard error of parameter estimates, resulting in inflated rates of Type I or II errors, especially where there is substantial dependency within an individual's repeated measurements [4, 5]. While growth curve models and mixed effect models are commonly used for analyzing longitudinal panel data in various fields, they may not be well-suited for ILD for several reasons [6]. Firstly, growth curve and mixed effect models can become computationally intensive when applied to longer and more frequent longitudinal data due to the complexity induced by nonlinearity and multiple random effects. Secondly, the shift towards intensive longitudinal designs reflect a broader methodological shift from focusing on stable processes (e.g., in psychology, a stable emotional process might be one where an individual quickly returns to their baseline mood after experiencing stress) to examining dynamic processes (e.g., human psychological and physical development is dynamic, with ongoing changes throughout a person's life, influenced by biological, environmental and social factors) [7]. This shift necessitates the

development of more advanced statistical models that can capture the unique features of ILD, including dynamic and often non-linear patterns.

Many real-world processes exhibit non-linear dynamics where behaviors change based on context or state, often referred to as "regime-effect" [8, 9]. For example, in affect regulation research, the autoregressive parameter, a measure of affect regulation, has been shown to vary based on the current level of affect [10]. To account for these intra-individual dynamic and non-linear patterns, Tong and Lim [11] introduced the threshold autoregressive (TAR) model. This was the first model to explain non-linear oscillations, such as harmonic or other periodic variations, and abrupt changes or discontinuities in time series data. TAR models introduce threshold points that act as a switching mechanism within the model. These thresholds segment the time series into different regimes, each of which can be modeled using different autoregressive processes. This allows the model to capture the complexity of systems where behavior can change dramatically based on past states.

ILD, which includes data from multiple individuals, allows for the examination of both inter-individual and intra-individual dynamics. This calls for the extension of TAR models to a multilevel framework, accommodating individual variation in TAR parameters. An example of such a model is the multilevel self-exciting threshold autoregressive (SETAR) model introduced by [12], which investigated affect regulation across individuals.

Many researchers have proposed methods for parameter estimation and model specification in TAR models. [13] introduced a Bayesian method for identifying threshold values using a posterior probability plot. [14] presented a Bayesian stochastic search selection approach for multiple threshold issues, employing a hybrid Markov Chain Monte Carlo (MCMC) method to estimate all model parameters.

The parameter estimation using a Bayesian approach with the Markov Chain Monte Carlo (MCMC) algorithm is computationally demanding. [12] reported that for datasets of moderate to large sizes (with 130 and 160 individuals and 120 and 56 time points, respectively), it took over three hours for two outcome variables and an hour and a half for one outcome variable on a system with 16 GB RAM, running on a single core with a processing speed of 3.40 GHz, using *OpenBugs* with two parallel MCMC chains. Consequently, the computational complexity of incorporating additional parameters in such models typically hinders the simultaneous modeling of multiple outcomes in moderate to large datasets, where the number of individuals ranges from 50 to 150 and the number of repeated measurements ranges from 30 to 150.

Despite these enhancements, the computational challenges associated with multilevel TAR models remain underexplored. To overcome the gap, our research advocates the use of the Mean Field Variational Bayes (MFVB) algorithm as a more efficient alternative to traditional Bayesian inference methods. Variational Bayes (VB) inference, a popular technique in machine learning approximates posterior densities in Bayesian models through optimization rather than simulation. Unlike standard Markov Chain Monte Carlo (MCMC) methods, which rely on iterative sampling techniques to approximate the posterior distribution, VB methods attempt to find closest approximation from a predefined distribution class, resulting in quicker convergence to complex posteriors.

The subsequent sections are structured as follows: Section 3.4.1 introduces the notations and specifications of the two-regime multilevel threshold autoregressive (ML-TAR) model. Section 3.4.2 provides a concise overview of the Mean Field Variational Bayes (MFVB) method and outlines the computations required to develop an algorithm for MFVB estimation of the ML-TAR model. The performance of our MFVB algorithm is evaluated via a series of simulation

studies against the MCMC approach in Section 3.5.1. In Section 3.5.2, we demonstrated the effectiveness of our approach using a real-world dataset on physical activity and affect regulation, previously examined by [1]. Finally, Section 6 includes a discussion and potential extensions of this work.

## 3.4 Methods

### 3.4.1 *Multilevel Threshold Autoregressive Model (ML-TAR)*

We first introduced the threshold autoregressive model (TAR) for an individual, as described in [12]. Their study focuses on the self-exciting TAR (2,1,1), comprising two alternating AR processes, each with first order auto-regression at lag 1. This model was denoted as TAR (2,1,1).

$$Y_t = \begin{cases} \tau + \phi_1 (Y_{t-1} - \tau) + \epsilon_t, & Y_{t-1} < \tau \\ \tau + \phi_2 (Y_{t-1} - \tau) + \epsilon_t, & Y_{t-1} \geq \tau \end{cases}$$

where, $\tau$ refers to the threshold value, $\phi_1$ and $\phi_2$ are autoregressive coefficients (or inertias) associated with two AR processes, and $Y_{t-1}$ is the lagged values of the outcome process $Y_t$.

[12] introduced the multilevel threshold autoregressive (ML-TAR) model for ILD for multiple individuals, which is appropriate to study both inter and intra-individual variation. This model can be expressed as follows:

$$\text{Level 1:} \quad Y_{t,i} = \begin{cases} \tau_i + \phi_{1,i} \left(Y_{t-1,i} - \tau_i\right) + \epsilon_{t,i}, & \text{if } Y_{t-1,i} < \tau_i \\ \tau_i + \phi_{2,i} \left(Y_{t-1,i} - \tau_i\right) + \epsilon_{t,i}, & \text{if } Y_{t-1,i} \geq \tau_i \end{cases} \tag{1}$$

$$\text{Level 2:} \quad \phi_{1,i} = \gamma_{\phi_1} + u_{\phi_1,i}$$
$$\phi_{2,i} = \gamma_{\phi_2} + u_{\phi_2,i}$$
$$\tau_i = \gamma_\tau + u_{\tau,i}$$

where $\begin{bmatrix} u_{\phi_1,i} \\ u_{\phi_2,i} \\ u_{\tau,i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_R = \begin{bmatrix} \sigma^2_{\phi_1} & \sigma_{\phi_1\phi_2} & \sigma_{\phi_1\tau} \\ \sigma_{\phi_2\phi_1} & \sigma^2_{\phi_2} & \sigma_{\phi_2\tau} \\ \sigma_{\tau\phi_1} & \sigma_{\tau\phi_2} & \sigma^2_{\tau} \end{bmatrix}\right)$, and $\epsilon_{t,i} \sim N(0, \sigma^2_\epsilon)$.

Here $i = 1,2,\cdots,m$ denotes the individuals, $t = 1,2,\cdots,n_i$ denotes the occasions or measurements, the fixed (or stable) effects $\gamma$ $(\gamma_{\phi_1}, \gamma_{\phi_2})$ represent the average of inertias $(\phi_{1,i}, \phi_{2,i})$, and $\gamma_\tau$ represents the average of threshold $(\tau_i)$. The random effects $(u_{\phi_1,i}, u_{\phi_2,i}, u_{\tau,i})$ can be correlated. The threshold $\tau_i$ for each person signifies the equilibrium point that partitions the space of the threshold variable $Y_{t-1,i}$. The residual term $\epsilon_{t,i}$ follows an independent and identically (iid) normal distribution with mean 0 and variance $\sigma_\epsilon^2$. Expressing the ML-TAR (2,1,1) model in matrix notation we have $\boldsymbol{y}|\,\boldsymbol{\gamma}, \boldsymbol{u},\ \mathbf{G}, \boldsymbol{\Sigma_R} \sim \boldsymbol{N}(\boldsymbol{X\gamma} + \boldsymbol{u}, \boldsymbol{G})|$, as defined in equation (1), where $\boldsymbol{X}$ is design matrices incorporating variables dependent on the threshold values (here , $Y_{t-1,i}$) and $\boldsymbol{G} = \boldsymbol{\Sigma}_\in$ defined in section 3.4.2.

This model extends the traditional threshold autoregressive model (TAR) of order 1 [11] to a multilevel framework, representing a piecewise AR model in the space of $Y_{t-d,i}$, where the delay parameter $d = 1$ determines the time lag at which threshold variable has a considerable effect on the changes in the time series structure of $Y_{t,i}$. More specifically, the delay parameter in a threshold autoregressive model determines how far back in time the model reviews to decide which regime (set of autoregressive parameters) to apply. This parameter is essential for capturing delayed effects in the time series and for accurately modeling the non-linear dynamics that the TAR model is designed to address. The delay parameter is often chosen based on domain knowledge, empirical testing, or through model selection criteria. In some cases, it is treated as a hyper-parameter and selected using techniques like cross-validation. Regarding inertia parameters in a TAR model, where the behavior of the series changes depending on the regime (e.g., depending on whether a threshold is crossed), high inertia suggests that within each regime, the series exhibits strong autoregressive behavior. This means that the past values heavily influence the current state, leading to a slower adjustment to new shocks or changes. High inertia also

suggests that the series is less likely to switch regimes frequently. For example, if the series is in one regime, it may take a larger shock or a more extended period of change to push the series into a different regime.

### 3.4.2 P*roposed algorithm for ML-TAR (2,1,1) model estimation*

We adopted a standard Bayesian estimation approach to infer the ML-TAR model by using the posterior distribution of unknown parameters. The Stan software, an open-source program for Bayesian model estimation, maintains the straightforward model specifications while enabling the simultaneous estimation of all parameters. Bayesian estimation was chosen due to challenges with classical approaches for the multilevel TAR model [10, 12, 15]. Previous studies have also applied Bayesian approach using the MCMC algorithm to estimate the unknown parameters in the TAR model to simplify statistical inference ([15, 16, 17]. It is worth noting that posterior sampling with MCMC can be computationally intensive, highlighting the need for new techniques to enhance the feasibility of inference in complex models.

### 3.4.3 *Overview of optimization-based approach*

In Bayesian inference, our objective was to find the posterior distribution for the model parameters of interest, denoted as θ. However, in nonlinear or non-Gaussian models, estimating the posterior can be challenging due to the intractable integrals involved in the posterior function or normalization component. To address this issue, we proposed the (naïve) mean field variational Bayes (MFVB) algorithm. This algorithm allows for efficient inference in moderate to large intensive longitudinal datasets, typically with 50 to 150 individuals and 30 to 150 repeated measures, as well as in multilevel models like the MLTAR model. The MFVB algorithm approximates $p(\theta \mid y)$ by using a simpler density function q(θ), especially in situation where a full MCMC sampling procedure would be computationally burdensome.

### 3.4.4 *General idea of (naïve) mean field variational Bayes (MFVB)*

In the mean field variational Bayes (MFVB) framework, the latent variables (i.e., parameters) are assumed to be mutually independent, with each being influenced by a distinct factor in the variational density. Specifically, MFVB achieves substantial computational efficiency by imposing a product restriction as shown in Equation (2).

$$q(\theta) = \prod_{k=1}^{M} q_k\,(\theta_k) \qquad (2)$$

where $\{\theta_1\,, \cdots, \theta_M\}$ represents a partition of $\theta$, and M denotes the number of partitions in the entire parameter space of $\theta$. In our case, $\theta$ contains fixed effects, random effects, the covariance matrix of random effects, innovation variance, and threshold parameters. This restriction is known as the mean field restriction, hence the optimal solution, $q^*(\theta)$ (from all possible distributions of $Q$ ) is known as the MFVB approximation to the actual posterior distribution $p(\theta\,|y)$. The main challenge of MFVB lies in choosing an optimal $q^*(\theta)$ that closely approximates the true posterior distribution in terms of Kullback-Leibler (KL) divergence [24]. To justify this approach, let's first examine the joint posterior vector given the observed data,

$$p(\theta\,|\boldsymbol{y}) = \frac{p(\boldsymbol{y}, \theta)}{p(\boldsymbol{y})}$$

Here, $\boldsymbol{y}$ denote all the longitudinal observations. As explained by [24], algebraic manipulations show that the logarithm of the marginal likelihood satisfies.

$$\log p(\boldsymbol{y}) = \int q(\theta) \log\left\{\frac{p(\boldsymbol{y},\theta)}{q(\theta)}\right\} \mathrm{d}\theta + \int q(\theta) \log\left\{\frac{q(\theta)}{p(\theta|\boldsymbol{y})}\right\} \mathrm{d}\theta\,,$$
$$= \log \underline{p}(\boldsymbol{y}, q) + KL\{q(\theta)||p(\theta|\boldsymbol{y})\}.$$

Since, a KL divergence is always non-negative ($KL \geq 0$), we have that $p(\boldsymbol{y}) \geq \underline{p}(\boldsymbol{y}, q)$, where $\underline{p}(\boldsymbol{y}, q)$ is often more tractable and known as evidence lower bound (ELBO) and so

minimizing the KL divergence is equivalent to maximizing $\underline{p}(\boldsymbol{y}, q)$. See Figure 1 for a simple graphical illustration. The posterior is approximated with optimal member $q^*(\theta)$ of a nice family of posterior $q(\theta)$ , which is called a *variational posterior*. Statistical inference is then based on the variational posterior $q^*(\theta)$ .

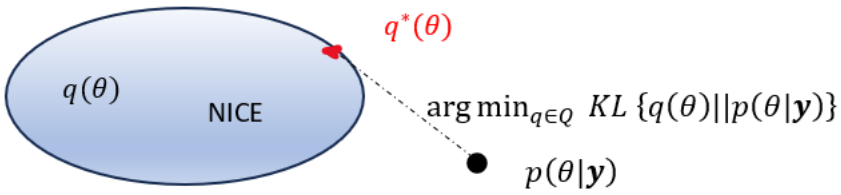



**Figure 3.1** An illustration of mean field variational Bayesian approach

In Figure 3.1, a **NICE** family of posterior distribution in the MFVB methods refers to a family of variational distributions that are natural (in terms of natural parameters), independent (factorized across different latent variables), and conditionally exponential-family distribution. The black dot in Figure 3.1 represents the true posterior distribution. The red dot represents the optimal member $q^*(\theta)$ of a NICE family of posterior $q(\theta)$, and the arrow pointing to the red dot indicates that MFVB aims to find the optimal $q^*(\theta)$ that closely approximate the true posterior $p(\theta|\boldsymbol{y})$ in terms of Kullback-Leibler (KL) divergence criteria.

Hence, we have

$$q^*(\theta) = \arg\min_{q\in Q}\ \mathrm{KL}\ \{q(\theta)||p(\theta|\boldsymbol{y})\} = \arg\max_{q\in Q}\ \underline{p}(\boldsymbol{y}, q)\ .$$

Under the mean field product restriction, the optimal $q$ density functions satisfy.

$$q_k^*(\theta_k) \propto \exp\{E_{-\theta_k}[\log p(\boldsymbol{y}, \theta)]\} \propto \exp\{E_{-\theta_k}[\log p(\theta_k| -\theta_k, \boldsymbol{y})]\}, \qquad (3)$$

The expectation $E_{-\theta_k}$in Equation (3) is with respect to all parameters in the model except for those in partition $\theta_k$, which are referred to as the rest. Because of the Mean-field family assumption, that all the latent variables (parameters) are independent, the expectations on the right-hand side do not

involve the $k^{th}$variational factor. Thus, this is a valid coordinate update, and Equation (3) underlies the coordinate ascent Mean-field variational inference (CAVI) algorithm as defined by [25] and illustrated in **Figure 3.2**.

**Input**: A model $p(\boldsymbol{y}, \theta)$ , a data set $\boldsymbol{y}$

**Output:** A variational density $q(\theta) = \prod_{k=1}^{M} q_k\ (\theta_k)$

**Initialize:** Variational factors $q_k\ (\theta_k)$

**While** *the* ELBO *has not converged* **do**

**for** $k \in \{1, \cdots, M\}$ **do**

Set $q_k^*(\theta_k) \propto \exp\{E_{-\theta_k}[\log p(\theta_k|rest)]\}$

**end**

Compute ELBO$(q)$ = E $[\log p(\boldsymbol{y}, \theta)]$ − E $[\log q(\theta)]$

**end**

**return** $q(\theta)$

**Figure 3.2** *Algorithm 1* Coordinate ascent variational inference (CAVI)

CAVI algorithm proceeds by determining optimal forms for each partition of $\theta$. These expressions can be iteratively updated until there is negligible increase in $\log \underline{p}(\boldsymbol{y}, q)$.

### 3.4.5 *Mean-field variational Bayes for ML-TAR (2,1,1) model*

Next, we will introduce the MFVB approximation for the ML-TAR model as outlined in Section 3.4.4. The Bayesian framework for the ML-TAR (2,1,1) by Equation (1) can be represented in a general form as below:

$$\boldsymbol{y}|\ \boldsymbol{\gamma}, \boldsymbol{u}, , \mathbf{G}, \boldsymbol{\Sigma_R} \sim N(\boldsymbol{X\gamma} + \boldsymbol{u}, \boldsymbol{G}), \qquad \boldsymbol{u}|\ \boldsymbol{\Sigma_R} \sim N(\mathbf{0},\ \boldsymbol{I_m} \otimes \boldsymbol{\Sigma_R}) \tag{4}$$

Here, $\boldsymbol{y} = (\boldsymbol{y_1}, \cdots, \boldsymbol{y_m})$ denotes the vector of observed measurements for all individuals and $\boldsymbol{y_i} = \left(y_{i1,} \cdots, y_{in_i}\right)$ denotes the vector of observed measurements for the $i^{th}$ individual $(i = 1, \cdots, m)$ at $t = (1, \cdots, n_i)$ time points. Furthermore, $\boldsymbol{\gamma}$ represents the vector of fixed effects and $\boldsymbol{u}$ denotes the $mq$ vector of individual random effects in our model. We also define covariance matrices

$$\boldsymbol{G} = \sigma_{\epsilon}^2 \boldsymbol{I} = \boldsymbol{\Sigma}_{\epsilon} \text{ and } \boldsymbol{\Sigma_R} = \begin{bmatrix} \sigma_{\phi_1}^2 & \sigma_{\phi_1\phi_2} & \sigma_{\phi_1\tau} \\ \sigma_{\phi_2\phi_1} & \sigma_{\phi_2}^2 & \sigma_{\phi_2\tau} \\ \sigma_{\tau\phi_1} & \sigma_{\tau\phi_2} & \sigma_{\tau}^2 \end{bmatrix}.$$

In addition, $\boldsymbol{X}$ represent the corresponding design matrix, including variables that are dependent on threshold variable, and $\boldsymbol{I_m}$ denotes the identity matrix**.**

We seek an approximation to the full posterior as follows:

$$p(\boldsymbol{\gamma}, \boldsymbol{u}, \Sigma_R, a_1, \cdots, a_k, a_{\in}, \sigma_{\epsilon}^2 | \boldsymbol{y}) \approx q(\boldsymbol{\gamma}, \boldsymbol{u}, \Sigma_R, a_1, \cdots, a_k, a_{\in}, \sigma_{\epsilon}^2) = q(\boldsymbol{\theta})$$

Subject to the $q$ product density restriction

$$q(\boldsymbol{\theta}) = q(\boldsymbol{\gamma}, \boldsymbol{u}, \boldsymbol{\Sigma_R}, a_1, \cdots, a_k, a_{\in}, \sigma_{\epsilon}^2) = q(\boldsymbol{\gamma}, \boldsymbol{u}, a_1, \cdots, a_k, a_{\in})\, q(\boldsymbol{\Sigma_R}, \sigma_{\epsilon}^2),$$

$$= q(\boldsymbol{\gamma}, \boldsymbol{u})\, q(\boldsymbol{\Sigma_R}) q(a_{\in}, \sigma_{\epsilon}^2) \prod_{k=1}^{q=3} q(a_k)$$

Here, in our model, the number of fixed effects $p = 3$ (two inertias and one threshold) and $\boldsymbol{u}$ denoting the $mq$ vector of individual random effects. We assume that the random effects in our model jointly follow a multivariate normal distribution with mean vector $\mathbf{0}$ and unstructured covariance matrix $\boldsymbol{\Sigma_R}$ as defined in Section 3.4.4. The remaining terms in our model are:

$$\boldsymbol{\Sigma_R} | a_1, \cdots, a_l \sim IW\left(v + q - 1, 2v\text{diag}\left\{1/a_1, \cdots, 1/a_q\right\}\right), \qquad a_k \sim IG(1/2\,, A_k^{-2})$$

$$\sigma_{\epsilon}^2 \sim IG(1/2\,, a_{\in}^{-1}), \quad a_{\in} \sim IG(1/2\,, A_{\in}^{-2}) \quad \gamma \sim N\left(0, \sigma_{\gamma}^2 I_p\right) \tag{5}$$

In Equation (5), we specify an Inverse-Wishart prior for the random effects covariance matrix $\boldsymbol{\Sigma}_R$ , and inverse gamma priors for the residual variances $\sigma_{\epsilon}^2$ of the level 1 units. The inclusion of auxiliary variables $a_l$ $(l = 1, \cdots, q)$ and $a_{\in}$ follows the extension of [18] to have

weakly informative priors (the Half-Cauchy distributions) on the covariance terms in $\boldsymbol{\Sigma}_R$, which is a special case of the conditionally-conjugate folded-noncentral -*t* family of prior distribution proposed by [19]. With $\upsilon$=2, the standard deviation terms in $\boldsymbol{\Sigma}_R$ have Half-*t* distributions, whilst the correlation parameters follow a uniform distribution (-1,1). Each auxiliary variable is assumed to independently follow an inverse-gamma distribution as specified in Equation (5). A normal prior with mean 0 and variance $\sigma_\gamma^2 I_p$ is placed on the fixed effects parameters $\gamma$ . Note that the second restrictions are induced due to assumed independence in the model.

The optimal $q$ – densities can be calculated using Equation (3). The updates for $q(\boldsymbol{\Sigma_R}), q(a_\epsilon, \sigma_\epsilon^2)$ and $q(a_l)$ involve standard calculations, resulting in inverse gamma (IG) distribution. Similarly, $q^*(\boldsymbol{\Sigma_R})$ follows an inverse Wishart distribution, and $q^*(\gamma, \boldsymbol{u})$ is a multivariate normal distribution. The derivation of all these q-densities is given in the Supplementary materials 3.9. The MFVB algorithm for the ML-TAR (2,1,1) model is presented in Algorithm 2.

---

**Algorithm 2** (Naïve) Mean-field variational Bayes algorithm for Multilevel Threshold Autoregressive Model (MLTAR)

---

0. **Data Input:** $\boldsymbol{y_i}$ **,** $i = 1, \cdots, m$ vectors of length *m*.
1. **Hyperparameter Input:** $\nu \in \mathbb{R}^+$, $A_\epsilon, A_{q_r,} \in \mathbb{R}^+$ and $c_0, b_0 \in \mathbb{R}^+$.
2. **Initialize:** $M_{q(\boldsymbol{\Sigma_R}^{-1})}$ a $q_r \times q_r$ positive definite matrix, $\mu_{q(1/\sigma_\epsilon^2)}, \mu_{q(1/a_\epsilon)}$, $and\ \mu_{q(1/a_{q_r})}, \mu_{q(\lambda_r)}, q_{rt}$, $c_0, b_0 \in \mathbb{R}^+$ for all **,** $i = 1, \cdots, m$ ; **,** $r = 1, \cdots, R, q = 1, \cdots, q_r$ and **,** $t = 1, \cdots, n_i$
3. **Cycle through updates:**

**for** $i = 1, \cdots, m$ **do**

**M-step:** for all $q(\lambda_r)$ fixed, and for **,** $r = 1, \cdots, R$ and **,** $t = 1, \cdots, n_i$

- **i.** recompute each $\xi_{r,t}$ for fixed $\zeta_t$ via equation contains ${\xi_{r,t}}^2$.
- **ii.** recompute each $\zeta_t$ for fixed $\xi_{r,t}$ via equation contains $\zeta_t$.

**End**

**E-Step:** for all fixed $\xi_{r,t}$ and $\zeta_t$; **for** $r = 1, \cdots, R$ **do**

**a)** $\boldsymbol{\Sigma_{q(\gamma_r, u_r)}} \leftarrow \left(\mu_{q(1/\sigma_\epsilon^2)}\left[\sum_{t=1}^{n_i} q_{rt}\, C_{r,t}{}^T C_{r,t}\right] + \begin{bmatrix} \sigma_{\gamma_r}^{-2} I_p & 0 \\ 0 & I_m \otimes \Sigma_{R_r}^{-1} \end{bmatrix}\right)^{-1}$

**b)** $\boldsymbol{\mu_{q(\gamma_r, u_r)}} \leftarrow \Sigma_{\boldsymbol{q(\gamma_r, u_r)}} \left(\mu_{q(1/\sigma_\epsilon^2)}\left[\sum_{t=1}^{n_i} q_{rt}\, C_{r,t}{}^T y_{r,t}\right]\right)$

**c)** $\mu_{\lambda_r} \leftarrow \Sigma_{\lambda_r} \left(\sum_{t=1}^{n_i}\left[q_{rt} - \frac{1}{2} + 2\zeta_t \rho(\xi_{r,t})\right] C_{r,t}\right)$

**d)** $\Sigma_{\lambda_r}{}^{-1} \leftarrow \mu_{\vartheta_r} + \sum_{t=1}^{n_i} 2\ \rho(\xi_{r,t}) C_{rt}\, C_{rt}^T$

**e)** $q_{rt} \leftarrow \frac{\exp \eta_r^T}{\sum_{r=1}^{R} \exp(\eta_r^T)} \qquad \eta_r \rightarrow \mu_{\lambda_r}{}^T C_{r,t} - \frac{1}{2}\, \mu_{q(1/\sigma_\epsilon^2)} \cdot \left(y_{r,t}^2 + C_{r,t}{}^T\left(\Sigma_{\boldsymbol{q(\gamma_r, u_r)}} + \mu_{\boldsymbol{q(\gamma_r, u_r)}} \mu_{\boldsymbol{q(\gamma_r, u_r)}}{}^T\right) C_{r,t} - 2 y_{r,t}\, \mu_{\boldsymbol{q(\gamma_r, u_r)}}{}^T C_{r,t}\right)$

**f)** $\mu_{q(1/a_\epsilon)} \leftarrow \frac{1}{\mu_{q(1/\sigma_\epsilon^2)} + A_\epsilon^{-2}} \qquad B_{q(a_\epsilon)} \leftarrow \mu_{q(1/\sigma_\epsilon^2)} + A_\epsilon^{-2}$

**g)** $\mu_{q(1/\sigma_\epsilon^2)} \leftarrow \frac{\frac{1}{2}\left(\sum_{i=1}^{m} n_i + 1\right)}{B_{q(\sigma_\epsilon^2)}}$

$$B_{q(\sigma_\epsilon^2)} \leftarrow \frac{1}{2}\sum_{r=1}^{R} q_{rt}\left\{\left(\left\|\boldsymbol{y_r} - \boldsymbol{C_r}\mu_{\boldsymbol{q(\gamma_r, u_r)}}\right\|^2 + tr\left[\boldsymbol{C_r}{}^T \boldsymbol{C_r}\, \Sigma_{\boldsymbol{q(\gamma_r, u_r)}}\right]\right) + \mu_{q(1/a_\epsilon)}\right\}$$

**for** $q = 1, \cdots, q_r$ **do**

**h)** $M_{q(\Sigma_{R_r}{}^{-1})} \leftarrow (\nu + q_r + m - 1)\, B_{q(\Sigma_{R_r})}{}^{-1}$

$$B_{q(\Sigma_{R_r})} \leftarrow \left\{\sum_{i=1}^{m}\left(\boldsymbol{\mu_{q(u_{R_r,i})}} \boldsymbol{\mu_{q(u_{R_r,i})}}{}^{\boldsymbol{T}} + \boldsymbol{\Sigma}_{q(u_{R_r,i})}\right) + 2\,\upsilon \mathrm{diag}\left(\mu_{q(1/a_1)} \cdots \mu_{q(1/a_{q_r})}\right)\right\}$$

**i)** $\mu_{q(1/a_{q_r})} \leftarrow \frac{\frac{\nu + q_r}{2}}{B_{q(a_{q_r})}} \qquad B_{q(a_{q_r})} \leftarrow \upsilon\, M_{q(\Sigma_{R_r}{}^{-1}) q_r q_r} + A_{q_r}^{-2}$

**Stop** if $F[q(I), q(\boldsymbol{\Theta})]$ converged, else repeat step 3.

---

### 3.4.6 Simulation Study

#### 3.4.6.1 *Data Generation Process*

The population multilevel threshold autoregressive model for data generation, specifically ML-TAR (2,1,1), is described below.

$$\text{Level 1: } Y_{t,i} = \begin{cases} \tau_i + \phi_{1,i}\left(Y_{t-1,i} - \tau_i\right) + \epsilon_{t,i}, & \text{if } Y_{t-1,i} < \tau_i \\ \tau_i + \phi_{2,i}\left(Y_{t-1,i} - \tau_i\right) + \epsilon_{t,i}, & \text{if } Y_{t-1,i} \geq \tau_i \end{cases}$$

$$\text{Level 2:} \qquad \phi_{1,i} = \gamma_{\phi_1} + u_{\phi_1,i}\,,$$

$$\phi_{2,i} = \gamma_{\phi_2} + u_{\phi_2,i}\,,$$

$$\tau_i = \gamma_\tau + u_{\tau,i}\,,$$

$$\text{With} \quad \begin{bmatrix} u_{\phi_1,i} \\ u_{\phi_2,i} \\ u_{\tau,i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma_R = \begin{bmatrix} \sigma^2_{\phi_1} & \sigma_{\phi_1\phi_2} & \sigma_{\phi_1\tau} \\ \sigma_{\phi_2\phi_1} & \sigma^2_{\phi_2} & \sigma_{\phi_2\tau} \\ \sigma_{\tau\phi_1} & \sigma_{\tau\phi_2} & \sigma^2_{\tau} \end{bmatrix}\right), \text{ and } \quad \epsilon_{t,i} \sim N(0, \sigma^2_{\epsilon}), \sigma_{\tau}.$$

In this simulation study, random samples were generated from the ML-TAR model with varying sample sizes. The number of person (*m*) ranged from 50 to 150 in increments of 50, and the number of observations (or time points, *T*) per person was set to 30, 50, 100 or 150, resulting in 12 conditions. For each condition, 5000 samples were generated after discarding the initial 1000 samples as burn-in and thinning the remaining the remaining samples by a factor of 20 [20]. Convergence of the MCMC samples was assessed using trace plots and autocorrelation functions. In the ML-TAR models, the fixed effects $\left[\gamma_{\phi_1}, \gamma_{\phi_2}\right]$ were set to [0.3, 0.5] and [0.2, 0.8] with standard deviations of 0.1. The threshold for those models was normally distributed with a mean 15 and a standard deviation 2. The correlation between the inertias was set to 0.5, and the level 1 innovation variance was set to 4 in the MLTAR models. We chose these parameters values based on [12] as they reflect the unique interpretation and scale of the inertia parameters in the TAR model. The stopping criteria for our algorithm was the relative change in the log lower bound, falling below $10^{-5}$ or a maximum of 300 iterations. We determined the maximum iteration value by iteratively checking the convergence of the parameter distribution, starting with 100 iterations and adjusting as needed through trial and error. While 100 simulation replications are generally considered a modest amount for Monte Carlo studies, we defined a replication as an independently

generated simulated dataset. To determine if our results were influenced by the number of replications, we performed extra simulations with 200 and 300 replications. Our findings revealed that performance patterns remained consistent regardless of the increased replication count. Thus, we determined that raising the number of replications past 100 did not significantly impact our outcomes and was unnecessary. Additionally, the estimation in Stan for this complex non-linear model using both MCMC and MFVB was time-consuming for 300 or more iterations.

For the stopping criteria, in MCMC we have considered trace plot and the autocorrelation function to check whether the chains have mixed well and reached stationarity. Since we observed that the plots converged and exhibited no clear trends, the algorithm was allowed to stop. In contrast, the stopping criteria in MFVB focus on the optimization process rather than on sampling convergence like MCMC. The most common stopping criterion in MFVB is based on the Evidence Lower Bound (ELBO), which quantifies how well the approximate distribution matches the true posterior. The algorithm typically stops when the change in the ELBO between iterations falls below a small tolerance level (in our case we set $10^{-5}$), indicating that further optimization is unlikely to yield significant improvement.

### 3.4.7 *Model Performance Measures*

We evaluated the performance of our proposed Mean-Field Variational Bayes (MFVB) algorithm compared to standard MCMC in terms of speed gains and accuracy. Specifically, we examined how accurately the MFVB estimates reproduced the true parameters (the values used to generate the data) by MFVB and compared these accuracy scores to those obtained with the MCMC method. To assess accuracy, we considered both the bias (the difference between the estimates and the true values) of the point estimates (i.e., the posterior means) for the average inertias and threshold, as well as the coverage of their 95% credible intervals. This evaluation was

based on 100 fitted models per simulation conditions. The actual coverage of their 95% credible intervals is defined as follows:

$$\text{Coverage Rate} = \frac{1}{S}\sum_{s=1}^{S} I(\theta_{true} \in CI_s)$$

Where S is the number of simulations and $I(\theta_{true} \in CI_s)$ is an indicator function, where $I = 1$ if the true parameter $\theta_{true}$ is within the credible interval and $I = 0$ otherwise.

### 3.4.8 *Real-world Data Application*

We showcased the application of the MFVB algorithm to fit the MLTAR (2,1,1) model in real-world scenarios. The data used in this demonstration is sourced from [1] and is openly accessible on the Harvard Dataverse Network database at http://thedata.harvard.edu/dvn/dv/healthbeahv. The study conducted by [1] explores the correlation between physical activity and affect regulation among university students during a demanding examination period through an intensive longitudinal study design.

The study involved first-year psychology students at the University of Basel, comprising a younger cohort aged 19 and above. A group of 82 students were selected to complete online questionnaires over 32 consecutive days during the examination period from May to July 2011. The primary focus was on analyzing the daily fluctuations in negative affect. Specifically, the study hypothesized that the regulation of negative affect on a daily basis was influenced by the individual's current state. The estimation of the ML-TAR model was restricted to students whose negative affect scores exhibited a standard deviation of 0.3 or higher. This criterion was applied to exclude participants with the most stable score patterns, while avoiding subjective decisions regarding their inclusion.

To further refine our selection criteria, we only included students with no more than five missing scores on the negative affect variable, ensuring a minimum of 25 scores per student. Given that the dataset also includes measurements of physical activity, we aimed to explore the correlation between physical activity and affect regulation. Specifically, we expected to observe a more positive disparity in inertia between two phases ($\phi_2 - \phi_1$) for highly active students compared to less active ones. Additionally, we hypothesized that highly active students would exhibit a lower equilibrium (threshold) of negative affect than their less active counterparts. To evaluate these hypotheses, we incorporated physical activity as a centered level 2 predictor of both the inertia difference and threshold. Out of the initial 82 students with adequate data, 75 individuals (91.4%) met the criteria for sufficient variance in negative affect scores and were included in the analysis.

***Study Measures***

Physical activity was evaluated using the Godin Leisure-Time Exercise Questionnaire (GLTEQ) and served as a covariate in our model. The daily minutes of mild, moderate, and intense exercise were converted into metabolic equivalents and combined to calculate a total daily leisure activity score. A higher score indicates greater physical activity levels. Daily positive and negative affect (PA and NA) were assessed using the Positive and Negative Affect Schedule (PANAS) on a 7-point Likert Scale, ranging from 1 (very little or no negative affect) to 7 (very intense negative affect) in 0.3 increments [21]. A 0.3 increment on a 7-point Likert scale for negative affect score suggests a small increase in the intensity or frequency of negative emotions. This increment provides a finer resolution for capturing changes in affective states, which can be important for tracking subtle shifts in mood or emotional well-being. While a 0.3 change is relatively small, in aggregate (e.g., across a population), such shifts could indicate meaningful trends or responses to

an intervention. For an individual, it could suggest a slight worsening or improvement in mood. In our analysis, we focused on daily NA scores as the outcome variable.

## 3.5 Results

### 3.5.1 *Results of simulation study*

Table 3.1 displays the average time required to fit each model across various simulation scenarios. The MFVB routine demonstrated significantly faster performance compared to the MCMC procedure. This speed gain is particularly evident when dealing with a large number of individuals and time points. For instance, in the scenario involving 150 individuals with 200 time points, the MCMC model necessitates over 50 minutes for fitting, whereas the MFVB completes the task in under 3 minutes on a system equipped with 32 GB RAM, operating on a single core with a processing speed of 4.70 GHz.

For small sample $m$ and $T$ (e.g., $m$=50, $T$=30), MCMC is about three times slower than our proposed approach (MFVB). However, as $T$ increases to 150 for $m$=50, MFVB is nine times faster than MCMC. For large models (e.g., $m$=150), the computational time for MCMC increases even more drastically. At $T$=150, MFVB is over 21 times faster than MCMC. MFVB is substantially more efficient than MCMC, especially as the model size and time points increase. This makes MFVB preferable for large-scale data or models with many parameters, as it offers similar accuracy with much faster computation time.

Comparing the speeds of different approaches can be somewhat subjective. When fitting a MLTAR (2,1,1) model, both the MCMC and MFVB approaches have different stopping criteria, as detailed in section 3.4.6.1. We have developed *RStan* code for fitting MLTAR (2,1,1) using both MCMC and MFVB with identical settings for the number of samples, burn-in, and thinning.

**Table 3.1** Average computing time in seconds for MFVB and MCMC approaches in 100 simulated data sets from MLTAR (2,1,1) model.

| | | **T=30** | **T=50** | **T=100** | **T=150** |
|---|---|---|---|---|---|
| **m=50** | MCMC | 455.28 | 2155.27 | 3055.27 | 4055.27 |
| | MFVB | 152.30 | 565.20 | 555.20 | 455.20 |
| | **Ratio** | **2.98** | **3.81** | **5.50** | **8.90** |
| **m=100** | MCMC | 465.28 | 2955.27 | 3385.27 | 9545.92 |
| | MFVB | 156.30 | 642.20 | 525.20 | 877.50 |
| | **Ratio** | **2.96** | **5.50** | **6.40** | **10.87** |
| **m=150** | MCMC | 865.28 | 6155.27 | 8645.92 | 25145.82 |
| | MFVB | 156.30 | 755.20 | 764.50 | 1167.50 |
| | **Ratio** | **5.50** | **8.10** | **11.30** | **21.53** |

Note: *m* represents number of person and *T* represents number of time points

It is worth mentioning that other package such as *brms* package [22] could also be used to fit the models considered in this article. Nevertheless, our aim in comparing speeds in this article is to show that MFVB models are substantially faster alternatives to standard MCMC algorithm.

The results from Figure 3.3 indicate that fixed effects and residual standard deviation are estimated accurately, showing minimal discrepancy between the MCMC and MFVB posterior estimates. However, the accuracy of the random effects covariance matrix is slightly lower, a common characteristic of MFVB algorithms as noted in previous studies [23]. Despite this, the accuracy remains above 60% for all covariance parameters. Notably, the accuracy of posterior estimates appears to be generally unaffected by larger sample sizes but ls lower for small sample size conditions such as m=50, T=30 and m=50, T=50).

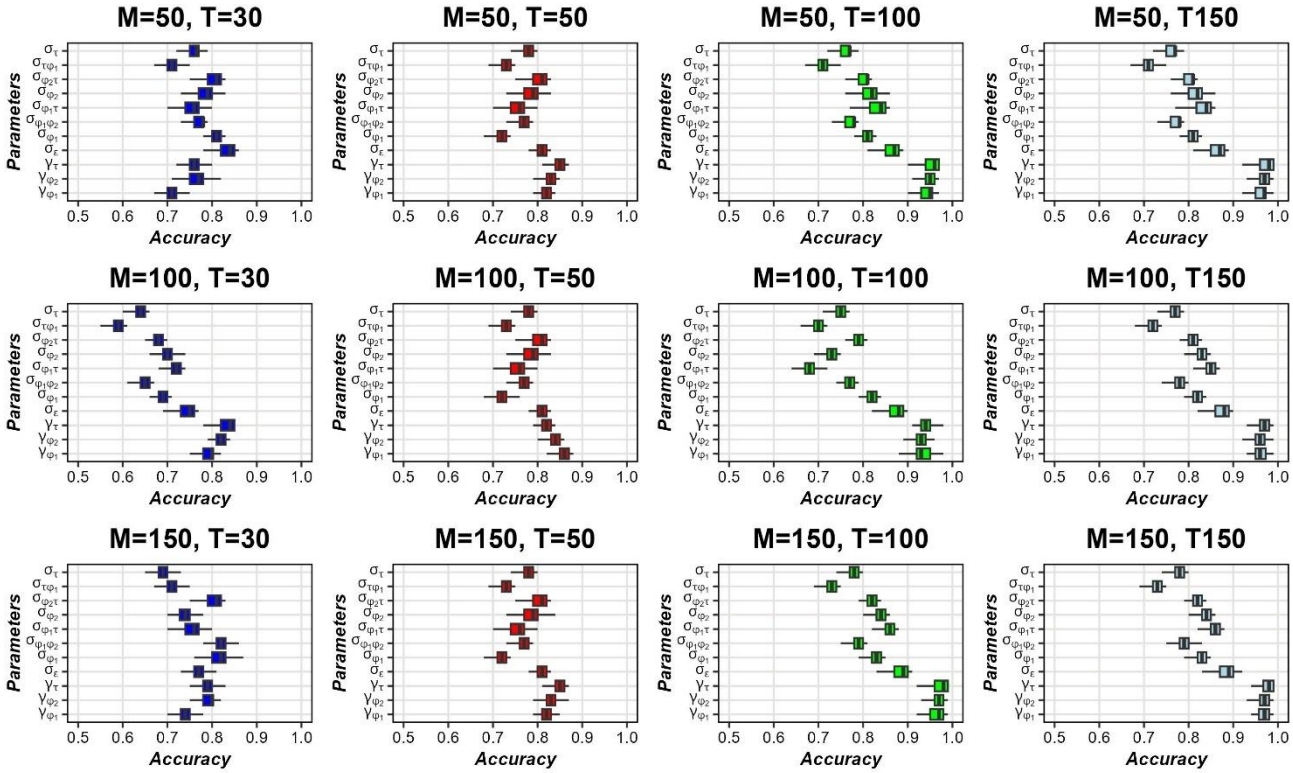



**Figure 3.3** Accuracy score box plots for mean field variational Bayes compared to MCMC for 100 simulated data sets of 12 simulated data generation conditions. *M* refers to the number of units (or individuals) used in the simulation; *T* refers to the number of time points used in the simulation, the parameters being estimated are likely correspond to variance and autoregressive parameters in the context of a threshold autoregressive model and on the x-axis, the accuracy of the parameter estimates is plotted. Higher values represent better accuracy (closer to 1.0), while lower values (closer to 0.6) indicate less accurate parameter estimates. Each boxplot shows the distribution of accuracy estimates for a particular parameter across multiple simulation runs.

Moreover, there are several general observations from Figure 3.3. For example, in terms of accuracy trends across time points, as T (the number of time points) increases from left to right in each panel the accuracy generally improves for most parameters. This is expected because more time points provide more information for estimating the parameters, leading to better accuracy. Additionally, in case of impact of the number of individuals (*M*), comparing across the panels, as

*M* increases, the accuracy of the parameter estimates also generally improves. This demonstrated that increasing the number of individuals help in obtaining more reliable estimates. Finally, regarding parameter-specific accuracy, different parameters exhibit varying levels of accuracy across different simulation conditions. Some parameters show unstable accuracy, such as $\sigma_\tau^2$ and $\sigma_\epsilon^2$ have higher accuracy even with lower values of *M* and *T*, while others may require larger sample sizes to achieve similar levels of accuracy. Overall, increasing *M* and *T* improves the accuracy of parameter estimates for our proposed method. Furthermore, it highlights how the accuracy of MFVB compared to MCMC improves with more data (both individuals and time points), suggesting that sufficient data is critical for accurate parameter estimation, especially for complex models like multilevel threshold autoregressive models.

### 3.5.2 *Empirical results*

Table 3.2 presents the model parameters results obtained from MFVB method. The credible interval of $(\gamma_{\phi_2} - \gamma_{\phi_1})$ in the ML-TAR (2,1,1) model was [0.34,0.84] with a point estimate of 0.30. Since this interval did not include zero, we concluded that the inertia was, on average, state dependent. The posterior mean estimates for the mean inertias ($\gamma_{\phi_1} = 0.38$ , $\gamma_{\phi_2} = 0.68$) indicate that, on average, students had a higher inertia during periods of more intense negative effect. The inertia estimates for the individual pupils are depicted in Figure 3.4. With a threshold value of $\gamma_\tau = 2.55$, these results indicate that many participants in our selected sample reported experiencing moderate negative affect on most days. When faced with more intense negative effects, they took longer to recover. Despite not meeting all TAR model assumptions perfectly, such as strict stationarity of the outcome process, the estimated model parameters offer valuable insights. The relatively high mean inertia parameter for the higher state suggests less stability, indicating that the average student struggled to maintain a lack of negative effect for

prolonged periods, possibly due to the age factor (average age = 22.5 years in our dataset). The effect of physical activity on the equilibrium (i.e., on the threshold $\tau$ ) was -0.06, supporting our hypothesis individuals with higher levels of physical activity exhibit a lower threshold for negative affect. This coefficient suggests that for every 1-hour increase in physical activity, there is a corresponding 0.06-point decrease in the negative affect equilibrium. Furthermore, we noted a significant influence of physical activity on the inertia difference, as evidenced by the 95% credible interval.

Table 3.2 provides the model parameter results derived from the standard MCMC method for comparison. While the empirical results align with those from the MFVB approach, the MFVB algorithm demonstrated a remarkable speed enhancement of approximately 19 times compared to MCMC. This significant speed advantage underscores the efficiency of the MFVB algorithm as a powerful tool for analyzing ILD, offering a compelling alternative to MCMC. Interestingly, the point estimates of the difference between the two auto regressive parameters fall outside of their 95% credible interval, likely due to characteristics of the posterior distribution such as multi-modality. The credible interval is computed based on where the bulk of the posterior probability lies, not necessarily centered on the mean or mode. This explains how this situation may arise.

Figure 3.4 shows the scatter plots of the estimated inertias for phase 1 ($\phi_1$) against those for phase 2 ($\phi_2$). Most data points lie below the diagonal line, indicating that a significant portion of students exhibited weak regulation during periods of heightened negative affect behavior.

**Table 3.2** MFVB vs MCMC point estimate (posterior mean) and 95% credible interval for selected parameters of the ML-TAR (2,1,1) model along with their computational time for our dataset.

| | MFVB | | | MCMC | | |
|---|---|---|---|---|---|---|
| | Mean | 95% CI | Time (Sec) | Mean | 95% CI | Time (Sec) |
| $\gamma_{\phi_1}$ | 0.38 | [0.23, 0.68] | | 0.34 | [0.24, 0.66] | |
| $\gamma_{\phi_2}$ | 0.68 | [0.47, 0.86] | | 0.65 | [0.42, 0.88] | |
| $(\gamma_{\phi_2} - \gamma_{\phi_1})$ | 0.30 | [0.34, 0.84] | 10034.62 | 0.25 | [0.33, 0.89] | 190567.80 |
| $\gamma_\tau$ | 2.55 | [1.33, 5.66] | | 2.15 | [1.31, 5.72] | |
| $\beta$ | -0.06 | [-0.085, -0.034] | | -0.08 | [-0.096, -0.031] | |

This suggests that these students tended to have prolonged episodes of negative affect and struggled to recover swiftly from intense negative emotions.

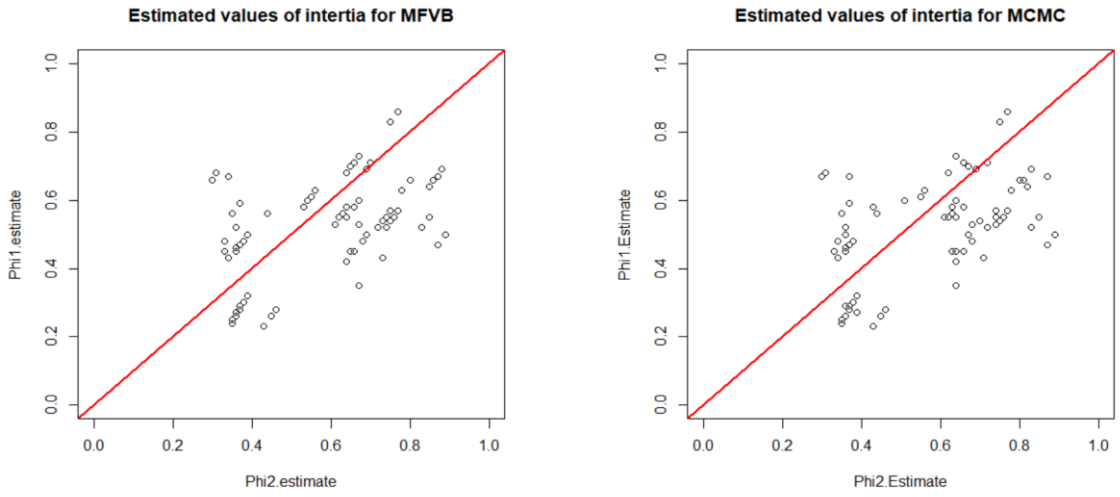



**Figure 3.4** Scatter plots of the estimated inertias between two phases.

Conversely, the points above the line indicate that some students experienced prolonged episodes with only minimal negative affect, and when they did experience more intense negative affect, they recovered quickly.

### 3.6 Discussion

This study introduced a fast approximate Bayesian inference method for analyzing multilevel threshold autoregressive (ML-TAR) models applied to ILD. Our research shows that the Mean Field Variational Bayes (MFVB) approach can efficiently and accurately produce results comparable to those obtained using the computationally demanding Markov Chain Monte Carlo (MCMC) method. By using MFVB, we achieved a good balance of speed and accuracy in estimates, making it a useful option for researchers working with complicated models and large datasets. These findings for MFVB method are consistent with previous studies [23-25].

Our simulation studies showed that the Mean Field Variational Bayes (MFVB) method offers substantial time savings compared to Markov Chain Monte Carlo (MCMC), albeit with slightly less precise covariance estimation in some cases. However, extensive simulations in various scenarios demonstrate that MFVB delivers accurate parameter estimates much faster, with an average speed enhancement of 18 times for moderate-sized samples compared to MCMC. This suggests that MFVB can be a viable alternative in situations where MCMC is computationally impractical which supports the findings of previous study [25].

We also explored the application of the MFVB method for ML-TAR models, which are effective for analyzing data collected repeatedly over a period. We examined how factors like the number of repeated measurements and the number of individuals in the study can affect the performance of the MFVB method. Understanding these factors is crucial for researchers who want to use ML-TAR models in their research, as it helps them adjust the model settings to achieve

the most accurate results. In almost all simulations conditions, we found that the accuracy of the variance and covariance estimates for ML-TAR model with both approaches was less precise. Several reasons justify our focus on less accurate estimates of these parameters: firstly, our primary focus was to obtain a predictive accuracy for fixed and random effects estimates in ML-TAR models rather than precise estimation of variance and covariance parameters. Since the MFVB estimates provide good accuracy, slight inaccuracies in variance and covariance estimates might be considered acceptable; Secondly, MFVB is known to provide relatively good point estimates for means and regression coefficients, as confirmed by our simulation study, but it can be less accurate for variance and covariance estimates. Moreover, computational savings might justify the less accurate estimation of variance and covariance parameters, especially since these parameters were not the focus of our analysis. Finally, the threshold and autoregressive parameters in the ML-TAR model directly influence the behavior of the time series and the indication of regimes, which might be of greater interest to researchers than the exact variance estimates [10,12].

The result of our empirical analysis, focusing on affect regulation data based on the ML-TAR model, showed that regulatory strength varies within individuals with the intensity of their affect, underlying the relevance of our model approach and aligning with the theoretical expectations of existing studies [11, 12]. We highlighted the adaptability of the ML-TAR model for addressing diverse research inquiries. While our empirical investigation was focused on the influence of physical activity on the threshold variable, the model can be expanded to incorporate the effects of other variables such as daily learning goal attainment, age, and sex. It is important to note that the MFVB algorithm may yield imprecise covariance matrix estimates, a recognized limitation. The significance of this issue depends on the researcher's specific requirements. If posterior mean estimates suffice, MFVB can offer accurate outcomes.

Other assumptions such as correct model order selection, no perfect multicollinearity, consideration of the exogeneity of the threshold variable, and choice of the number of thresholds need to be met for the ML-TAR model. However, for simplicity, this study assumes that data are stationary, and the series is driven by a simple autoregressive process, where the most recent observation is the primary determinant of the next observation (simple dynamics). In this study, we subjectively chose model order 1 to ensure simple dynamics of the process.

### 3.6.1 Study Implications

The findings have a number of implications for further research and public health. Overall, this study contributes to the growing body of literature on modeling ILD by highlighting MFVB as a fast and effective alternative to MCMC for scalable Bayesian inference in correlated data. Future research endeavors could build upon this foundation to explore the full potential of MFVB in addressing complex modeling challenges in longitudinal data analysis.

### 3.6.2 Strengths and Limitations

The strengths of this study lie in the innovative application of MFVB algorithm to ILD model under discrete data frameworks. The optimization-based approach enhances parameter accuracy and scalability, offering improved capacity to describe the dynamic characteristics of ILD. We evaluated MFVB and MCMC for computational efficiency using both simulated and real-world data. The proposed algorithm was effectively assessed with real-world sources.

The study faced limitation because the delay parameter lag order was subjectively set to 1 when developing the new algorithm, prioritizing simplicity. However, the choice of values for the delay parameter can significantly impact the efficiency and accuracy of parameter estimation methods like Mean Field Variational Bayes (MFVB). Here's how: in case of model fit and

complexity, an incorrect delay parameter can prevent the TAR model from capturing the true underlying process, leading to poor model fit. This may cause the threshold to misidentify regime changes, thereby diminishing the model's performance. An incorrect delay can result in model misspecification, increasing the potential for bias in parameter estimates and decreasing the accuracy of MFVB. In our study, we used a subjective choice based on literature and for ease of deriving the evidence lower bound (ELBO) for parameter estimation. Nonetheless, identifying the optimal delay parameter could be a focus of future research to improve estimate accuracy. Moreover, in case of convergence of MFVB, if the delay parameter is inaccurate, the MFVB algorithm might struggle to efficiently explore the parameter space. The delay parameter can influence the shape of the posterior distribution, potentially leading to a multimodal or skewed posterior if the delay is inappropriate [27,28]. This complicates the estimation process and interpretation. Although we did not address these issues in our current study, they represent promising avenues for future investigation.

A higher-order lag of the delay parameter would require extensive mathematical integration for our proposed approach. Additionally, we assume that the underlying series shows strong first-order autocorrelation and that the regime switching is driven by the most recent observation. The goal was to maintain model simplicity, avoid overfitting, and ensure ease of interpretation of the TAR model parameters. Although the efficiency and accuracy of the parameter estimation method can be substantially influenced by the choice of delay parameter lag order, we have decided to leave this topic for future research.

### 3.6.3 Future Directions

In future research, we will examine how violating these assumptions affects parameter estimates and will apply model diagnostics, robust statistical methods, and alternative models to

address these issues. Moving forward, it would be valuable for future studies to delve into the application of the mean field variational Bayes (MFVB) method to higher-order threshold autoregressive (TAR) models utilizing clustered ILD. This extension could provide insights into the performance of MFVB in capturing complex temporal dependencies and thresholds in such models. Additionally, exploring the robustness and efficiency of MFVB in handling larger and more intricate TAR structures within clustered data settings could offer further understanding of its applicability in real-world scenarios.

Furthermore, as the literature on modeling ILD continues to expand, further research might also focus on revising and extending the MFVB approach to cover a wider scope of application and more varied structure of data. Investigating the adaptability of MFVB to different types of longitudinal data, such as multivariate time series or longitudinal networks, could open new avenues for utilizing this method in diverse research domains.

### 3.7 Conclusions

By leveraging proposed algorithms and incorporating an intensive longitudinal dataset, our study demonstrated a promising avenue for identifying linear relationships that varies over regimes delineated by threshold values via a fast and effective alternative to MCMC for scalable Bayesian inference. We suggest using a computationally efficient estimation approach for the ML-TAR model to handle regime shifts in the ILD outcome process, rather than depending only on MCMC methods.

### 3.8 References


1. L. Flueckiger, R. Lieb, A. H. Meyer, and J. Mata, "How health behaviors relate to academic performance via affect: An intensive longitudinal study," *PLoS ONE*, vol. 9, no. 10, 2014, doi: 10.1371/journal.pone.0111080.

2. P. C. M. Molenaar, K. O. Sinclair, M. J. Rovine, N. Ram, and S. E. Corneal, "Analyzing developmental processes on an individual level using nonstationary time series modeling," *Developmental Psychology*, vol. 45, no. 1, pp. 260–271, 2009, doi: 10.1037/a0014170.

3. T. A. Walls and J. L. Schafer, *Models for Intensive Longitudinal Data*. Oxford, U.K.: Oxford University Press, 2012, doi: 10.1093/acprof:oso/9780195173444.001.0001.

4. M. Moerbeek, G. J. P. Van Breukelen, and M. P. F. Berger, "A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies," *Journal of Clinical Epidemiology*, vol. 56, no. 4, pp. 341–350, 2003, doi: 10.1016/S0895-4356(03)00007-6.

5. B. E. Wampold and R. C. Serlin, "The consequence of ignoring a nested factor on measures of effect size in analysis of variance," *Psychological Methods*, vol. 5, no. 4, 2000.

6. K. J. Grimm, P. Davoudzadeh, and N. Ram, "Developments in the analysis of longitudinal data," *Monographs of the Society for Research in Child Development*, vol. 82, no. 2, pp. 46–66, 2017, doi: 10.1111/mono.12298.

7. L. Zhou, M. Wang, and Z. Zhang, “Intensive longitudinal data analyses with dynamic structural equation modeling,” *Organizational Research Methods*, vol. 24, no. 2, pp. 219–250, 2021, doi: 10.1177/1094428119833164.

8. H. Tong, “Threshold models in time series analysis: 30 years on,” *Statistics and Its Interface*, vol. 4, 2011.

9. K. S. Chan and H. Tong, “On estimating thresholds in autoregressive models,” *Journal of Time Series Analysis*, vol. 7, no. 3, pp. 179–190, 1986, doi: 10.1111/j.1467-9892-1986.tb00501-x.

10. E. L. Hamaker, “Using information criteria to determine the number of regimes in threshold autoregressive models,” *Journal of Mathematical Psychology*, vol. 53, no. 6, pp. 518–529, 2009, doi: 10.1016/j.jmp.2009.07.006.

11. H. Tong and K. S. Lim, “Threshold autoregression, limit cycles and cyclical data,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 42, no. 3, 1980.

12. S. De Haan-Rietdijk, J. M. Gottman, C. S. Bergeman, and E. L. Hamaker, “Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation,” *Psychometrika*, vol. 81, no. 1, pp. 217–241, 2016, doi: 10.1007/s11336-014-9417-x.

13. R. E. McCulloch and R. S. Tsay, “Bayesian analysis of autoregressive time series via the Gibbs sampler,” *Journal of Time Series Analysis*, vol. 15, no. 2, pp. 235–250, 1994, doi: 10.1111/j.1467-9892-1994.tb00188-x.

14. J. Pan, Q. Xia, and J. Liu, “Bayesian analysis of multiple thresholds autoregressive model,” *Computational Statistics*, vol. 32, pp. 219–237, 2012.

15. C. W. S. Chen, "A Bayesian analysis of generalized threshold autoregressive models," *Statistics & Probability Letters*, vol. 40, 1998.

16. J. Geweke and N. Terui, "Bayesian threshold autoregressive models for nonlinear time series," *Journal of Time Series Analysis*, vol. 14, no. 5, pp. 441–454, 1993, doi: 10.1111/j.1467-9892-1993.tb00156-x.

17. C. W. S. Chen and J. C. Lee, "Bayesian inference of threshold autoregressive models," *Journal of Time Series Analysis*, vol. 16, no. 5, pp. 483–492, 1995, doi: 10.1111/j.1467-9892-1995.tb00248-x.

18. A. Huang and M. P. Wand, "Simple marginally noninformative prior distributions for covariance matrices," *Bayesian Analysis*, vol. 8, no. 2, pp. 439–452, 2013, doi: 10.1214/13-BA815.

19. A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian Analysis*, vol. 1, no. 3, 2006. [Online]. Available: http://www.stat.columbia.edu/~gelman/

20. R. L. Harms and A. Roebroeck, "Robust and fast Markov chain Monte Carlo sampling of diffusion MRI microstructure models," *Frontiers in Neuroinformatics*, vol. 12, Art. no. 97, 2018, doi: 10.3389/frinf.2018.00097.

21. D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, Jun. 1988, doi: 10.1037/0022-3514.54.6.1063.

22. P. C. Bürkner, “brms: An R package for Bayesian multilevel models using Stan,” *Journal of Statistical Software*, vol. 80, 2017, doi: 10.18637/jss.v080.i01.

23. J. Luts and M. P. Wand, “Variational inference for count response semiparametric regression,” *Bayesian Analysis*, vol. 10, no. 4, pp. 991–1023, 2015.

24. J. T. Ormerod and M.P. Wand, “Explaining Variational Approximation”, *The American Statistician,* 64(2), 140-153, https://doi.org/10.1198/tast.2010.09058.

25. D. M. . Hughes, M. García-Fiñana, and M.P. Wand. “Fast approximate inference for multivariate longitudinal data,” *Biostatistics.* Dec 12;24(1):177-192, 2022. doi: 10.1093/biostatistics/kxab021. PMID: 33991420; PMCID: PMC9748580.

26. N. Bolger, and J. P. Laurenceau, *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press, 2013.

27. T. T. L. Chong, H. Chen, TN. Wong, and IKM. Yan, “Estimation and inference of threshold regression models with measurement errors”, Stud Nonlinear Dyn Econom. 22(2):1–27,2017.

28. R. J. Carroll, D. Ruppert, LA. Stefanski, and CM. Crainiceanu, “Measurement Error in Nonlinear Models: A Modern Perspective”. 2nd ed. Boca Raton: CRC Press; 2006.

## 3.9 Supplementary Materials

### 3.9.1 Derivation of MFVB approximation for ML-TAR (2,1,1) model

This supplementary document elaborates on the derivation of the mean field variational Bayes (MFVB) approximation for the multilevel threshold autoregressive (ML-TAR) model applied to ILD. It outlines the alternative form of the ML-TAR model, derives the optimal form of each q-density, and presents an expression for the log-lower bound on the marginal log-likelihood. Furthermore, additional simulation results are included for comprehensive illustration. The two-regime multilevel threshold autoregressive model can be represented as the following linear mixed model with the help of indicator function.

L1: $Y_{t,i} = \tau_i + \left[\phi_{1,i}\left(Y_{t-1,i} - \tau_i\right) I\left(W_{t-1,i} < \tau_i\right)\right] + \left[\phi_{2,i}\left(Y_{t-1,i} - \tau_i\right) I\left(W_{t-1,i} \geq \tau_i\right)\right] + \epsilon_{t,i}$

L2:
$$\phi_{1,i} = \gamma_{\phi_1} + u_{\phi_1,i}\,, \tag{1}$$
$$\phi_{2,i} = \gamma_{\phi_2} + u_{\phi_2,i}\,,$$
$$\tau_i = \gamma_{\tau} + u_{\tau,i}\,,$$

Let $X_{t,i} = \left((Y_{t-1,i} - \tau_i)\right)'$ and $X_t(\tau_i) = \left(1,\ X_{t,i}{}'\, I\left(W_{t-1,i} < \tau_i\right)\ X_{t,i}{}'\, I\left(W_{t-1,i} \geq \tau_i\right)\right)'$

$$Y_{t,i} = \boldsymbol{\gamma_i}\, \boldsymbol{X_t}(\boldsymbol{\tau_i}) + \epsilon_{t,i}$$

Let, $\boldsymbol{\gamma_i} = \left(\tau_i\,, \phi_{1,i}, \phi_{2,i}\right)'$ be a vector of individual coefficients comprising both fixed and random effects, denoted as $\boldsymbol{\gamma} = \left(\gamma_{\phi_1}, \gamma_{\phi_2}, \gamma_{\tau}\right)'$, $I$ is an indicator function, and $\boldsymbol{u_i} = \left(u_{\phi_1,i}, u_{\phi_2,i}, u_{\tau,i}\right)'$, respectively. Expressing the ML-TAR model in matrix notation we have $\boldsymbol{y|\ \gamma, u,\ G, \Sigma_R \sim N(X\gamma + u, G)}$|. as defined in equation (1), where $\boldsymbol{X}$ is design matrices incorporating variables dependent on the threshold values (here $W_{t-1,i} = Y_{t-1,i}$). The above model is a multilevel version of the self-exiting threshold autoregressive (ML-SETAR) model or the non-Markov (regime) switching

probabilistic mixture of autoregressive (AR) models. For simplicity, we will focus on two AR models of order 1 corresponding to two regimes. Since our ML-TAR (2,1,1) model is a non-Markov switching probabilistic mixture of AR models, we we can rewrite equation (1) in line with a general mixture of expert (ME) autoregressive model outlined in Nisar and Mark (2011) as follows,

$$Y_{t,i} = \sum_{r=1}^{R} P\left(I_{t,i} = r | W_{t-1,i}, \Lambda\right) \left[\left(X_{t,i}{}' \, \phi_{1,i} \, I\left(W_{t-1,i} < \tau_i\right) \; + \; X_{t,i}{}' \, \phi_{2,i} I\left(W_{t-1,i} \geq \tau_i\right)\right)\right.$$

$$\left. + \epsilon_{t,i}\right] . \qquad (1.1)$$

The hidden variable $\boldsymbol{I_i} = \left\{I_{1,i}, \cdots, I_{T,i}\right\}$ is the identity of identification data pairs indicating which $r^{th}$ (regime) local model generates the data pairs for the *i*th individual and $\boldsymbol{I} = \{\boldsymbol{I_1}, \cdots, \boldsymbol{I_m}\}$ is the total number of data points. Given the number of local models based on the number of regimes, a set of parameters needs to be estimated that includes the local model parameters defined in $\boldsymbol{\Theta}$. Additionally, a model weighting function $P\left(I_{t,i} = r | W_{t-1,i}, \Lambda\right) = \alpha_r$ is introduced to signify the importance of each local model, with a set of mixing coefficients $\alpha = \{\alpha_1, \cdots, \alpha_R\}$, subject to the constraint that

$$\sum_{r=1}^{R} \alpha_r = 1 \qquad (1.1a).$$

Now let us define the unobserved $r \times n_i$ binary level matrix $B$ with elements $b_{rt} = 1$ if $I_{t,i} = r$ and $b_{rt} = 0$ otherwise for each observed data collected from the *i*th individual. Direct continuous optimization methods using non-linear least squares (NLS) (Taguchi et al., 2009) or maximum likelihood (MLE) (Carvalho & Tanner, 2006) can be used to estimate the mixing coefficients and model parameters with fixed delay and autoregressive model orders. However, both NLS and MLE are sensitive to outliers, making them vulnerable to severe overfitting (Roberts & Penny, 2002).

In the Variational Bayes (VB) algorithm, the posterior probability distributions over parameters are not evaluated directly. Instead, free distributions or a suitable family of distributions, such as exponential family of density distribution, are used to approximate the intractable true posterior distributions. In this study, the following free distributions will be introduced for each individual:

$$q(\boldsymbol{I}) = \prod_{i=1}^{m} \prod_{t=1}^{n_i} q\left(I_{t,i}\right), \tag{1.2}$$

$$q(\boldsymbol{\Theta}) = \prod_{r=1}^{R} q(\boldsymbol{\Theta}_r). \tag{1.3}$$

Thus, the logarithm of the marginal likelihood can be expressed as

$$\ln p(Y|\boldsymbol{\Psi}) = \mathcal{L} + KL(q||p),$$

where the lower bound is

$$F[q(\boldsymbol{I}), q(\boldsymbol{\Theta})] = \mathcal{L}$$

$$= \sum_{I} \int q(\boldsymbol{I}) q(\boldsymbol{\Theta}) \ln \frac{p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta}|\boldsymbol{\Psi})}{q(\boldsymbol{\Theta}) q(\boldsymbol{I})} \, d\boldsymbol{\Theta} \,. \tag{1.4}$$

Here, $\boldsymbol{\Psi} = \{r, l_u\}$ and $l_u$ is fixed output lag and $r$ is the number of regimes such that $r \geq 1$.

**A. Priors**

- The prior probability of $P\left(I_{t,i} = r|W_{t-1,i}, \Lambda\right)$ is defined via the SoftMax function in Taguci et al. (2009) as follows:

$$P\left(I_{t,i} = r|W_{t-1,i}, \Lambda\right) = \frac{\exp\left(\lambda_r^T W_{t-1,i}\right)}{\sum_{r=1}^{R} \exp\left(\lambda_r^T W_{t-1,i}\right)}, \qquad for\ each\ i = 1,2,\cdots,m$$

- Observed output $Y_{t,i}$ at time $t$ for individual $i$, with regime ($r$) specific conditional pdf $p\left(Y_{t,i}|W_{t-1,i}, I_{t,i}=r\right)\sim N\left(\boldsymbol{f_r}(\boldsymbol{W_{t-1,i}}), \boldsymbol{G}\right)$ where, $\boldsymbol{f_r}\left(\boldsymbol{W_{t-1,i}}\right) = \boldsymbol{X\gamma} + \boldsymbol{u} = \boldsymbol{C\Theta_r}$ and $\boldsymbol{G} = \sigma_{\epsilon}^2\boldsymbol{I} = \boldsymbol{\Sigma_{\epsilon}}$**., where,** $\boldsymbol{C} = [\boldsymbol{1}\ \boldsymbol{X}]$**.**
- The priors for $\boldsymbol{\Lambda} = [\lambda_1, \cdots, \lambda_r]$ is $p(\boldsymbol{\Lambda}|\boldsymbol{\vartheta}) = \prod_{r=1}^{R} p(\lambda_r|\vartheta_r)$ . Here, $p(\lambda_r|\vartheta_r) = N_{\lambda_r}\left(0, {\vartheta_r}^{-1}I\right)$ with unknown precision parameter ${\vartheta_r}^{-1}$. Here, $p(\vartheta_r;\ c_0, b_0) = Gamma\ (c_0, b_0)$ with known $c_0, b_0$ as defined in Nisar et al. (2011).
- The priors for the set of all unknown parameters $\boldsymbol{\Theta_r} = (\boldsymbol{\gamma^T}, \boldsymbol{u^T})^{\boldsymbol{T}}$ in $\boldsymbol{f_r}\left(\boldsymbol{W_{t-1,i}}\right)$ have been defined previously.

**1. Derivation of optimal $\boldsymbol{q^*}$ densities**

To find the optimal densities for each part of the factorization of $q(\boldsymbol{\Theta})$ we use the result:

$$q_j^*(\theta_j) \propto exp\left\{E_{-\theta_j}[log\ p(\boldsymbol{y}, \theta)]\right\} \propto exp\left\{E_{-\theta_j}\left[log\ p\left(\theta_j| -\theta_j, \boldsymbol{y}\ rest\right)\right]\right\}$$

If we denote $\boldsymbol{\Theta} = \left[\sigma_{\epsilon}^2, a_{\in}, \boldsymbol{\gamma}, \boldsymbol{u}, \boldsymbol{\Sigma_R}, a_q, \Lambda, \vartheta, B\right]$ to be the set of all latent variables, then the joint pdf of $[Y, \boldsymbol{\Theta}]$ given $\boldsymbol{\Psi}$ is:

$$p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta}|\boldsymbol{\Psi})$$

$$= p(\sigma_{\epsilon}^2|a_{\in})\ p(\ a_{\in}) \prod_{r=1}^{R} p\left(\boldsymbol{\gamma_r}, \boldsymbol{u_r}|\ \boldsymbol{\Sigma_{R_r}}\right)\ p\left(\ \boldsymbol{\Sigma_{R_r}}|a_1, \cdots, a_{q_r}\right)\ p\left(a_{q_r}\right)\ p(\lambda_r)p(\vartheta_r)$$

$$\times \prod_{t=1}^{n_i}\prod_{r=1}^{R}\left[p\left(y_{t,i}|w_{t,i}, I_t = r, \boldsymbol{\Theta}, \Lambda\right)\ p(I_t = r|\Lambda)\right]^{b_{rt}}$$

Now, for ML-TAR(2,1,1) learning, the variational approximate posterior distribution $q$ is specified via the typical "mean field" approximation, which is the product of optimal densities $q^*$ of each factorized term of $\boldsymbol{\Theta} = \left[\sigma_{\epsilon}^2, a_{\in}, \boldsymbol{\gamma}, \boldsymbol{u}, \boldsymbol{\Sigma_R}, a_q, \Lambda, \vartheta, B\right]$ , as follows:

$$q(\boldsymbol{\Theta}) = q^*(\sigma_\epsilon^2|a_\in)\, q^*(a_\in) \prod_{r=1}^{R} q^*\big(\boldsymbol{\gamma_r}, \boldsymbol{u_r}|\, \boldsymbol{\Sigma_{R_r}}\big)\, q^*\big(\, \boldsymbol{\Sigma_{R_r}}|a_1, \cdots, a_{q_r}\big)\, q^*\big(a_{q_r}\big)\, q^*(\lambda_r)\, q^*(\vartheta_r) \prod_{t=1}^{n_i} \prod_{r=1}^{R} q(B_t = r)$$

Now, the logarithm of the joint probability distribution of all the variables for our models is defined as:

$$E_{-\boldsymbol{\Theta_j}}[\log p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})]$$

$$= E_{-\boldsymbol{\Theta_j}}\left[\sum_{t=1}^{n_i}\sum_{r=1}^{R} \log\{p\big(y_{t,i}|w_{t,i}, I_t, \boldsymbol{\Theta}, \Lambda\big)\, p(I_t = r|\Lambda)\}\right] + E_{-\boldsymbol{\Theta_j}}[\log p(\, \sigma_\epsilon^2|a_\in)]$$

$$+ E_{-\boldsymbol{\Theta_j}}[\log p(\, a_\in)]$$

$$+ \sum_{r=1}^{R}\Big\{E_{-\boldsymbol{\Theta_j}}\big[\log p(\, \boldsymbol{\gamma_r}, \boldsymbol{u_r}\, |\boldsymbol{\Sigma}_{R_r})\big] + E_{-\boldsymbol{\Theta_j}}\big[\log p(\, \boldsymbol{\Sigma_{R_r}}\, |a_1, \cdots, a_{q_r}\, )\big]$$

$$+ E_{-\boldsymbol{\Theta_j}}\big[\log p(\, a_{q_r})\big] + E_{-\boldsymbol{\Theta_j}}[\log p(\, \lambda_r|\vartheta_r\, )] + E_{-\boldsymbol{\Theta_j}}[\log p(\vartheta_r\, )]\Big\}$$

**D (1.1) Derivation of optimal $\boldsymbol{q^*}\ (I_t)$ densities**

The optimal density for $q(I_t)$ is

$$q^*(I_t) \propto exp\big\{E_{-q(\boldsymbol{I_t})}[\log p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta}|\Psi)]\big\} \propto exp\big\{E_{-q(\boldsymbol{I_t})}\big[\log\{p\big(y_{t,i}|w_{t,i}, I_{t,i}, \boldsymbol{\Theta}, \Lambda\big)\, p\big(I_{t,i}|\Lambda\big)\}\big]\big\}$$

Since $Y_{t,i}|W_{t,i}, I_{t,i}, \boldsymbol{\Theta} \sim N\big(\boldsymbol{f_r}(\boldsymbol{W_{t-1,i}}), \boldsymbol{G}\big)$ and $p(I_t = r|\Lambda) = \frac{\exp(\lambda_r^T w_{t-1,i})}{\sum_{r=1}^{R}\exp(\lambda_r^T w_{t-1,i})}$, therefore, after some calculation and applying approximation of Softmax function according to Bouchard (2007) and based on the expected log-likelihood, the approximated posterior for $I_t$ is:

$$E(b_{rt}) = q^*(I_t = r) = \frac{\exp \eta_r^T}{\sum_{r=1}^{R}\exp(\eta_r^T)}$$

Where, $\eta_r = E[\lambda_r]^T w_{t-1,i} - \frac{1}{2}\, E[\, \boldsymbol{\Sigma}_\in] \cdot \big(y_{t,i}^2 + {w_{t-1,i}}^T E[\, \boldsymbol{\Theta_r}\, \boldsymbol{\Theta_r}^{\,T}] w_{t-1,i} - 2y_{t,i} E[\boldsymbol{\Theta_r}]^T w_{t-1,i}\big)$

for $E[\, \boldsymbol{\Theta_r}\, \boldsymbol{\Theta_r}^{\,T}] = \Sigma_{\Theta_r} + \mu_{\boldsymbol{\Theta_r}}\, \mu_{\boldsymbol{\Theta_r}}^T$ . Clearly, the optimal density depends on the distribution of $\boldsymbol{\Theta_r}$ , $\boldsymbol{\Sigma}_\in$ **, and** $\lambda_r$ .

**D (1.2) Derivation of optimal $q^*(a_\epsilon)$ density**

In this case, the functional derivation is the partial derivative with respect to $q^*(a_\epsilon)$ of what is inside the integral of $F[q(\boldsymbol{I}), q(\boldsymbol{\Theta})]$ and solving for the form of $q^*(a_\epsilon)$ we have:

$$q^*(a_\epsilon) \propto exp\{E_{-a_\epsilon}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta}|\boldsymbol{\Psi})]\}$$

To find the optimal form of $q^*(a_\epsilon)$ density we have:

$$E_{-a_\epsilon}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})] = E_{-a_\epsilon}[log\, p(\, \sigma_\epsilon^2|a_\epsilon)] + E_{-a_\epsilon}[log\, p(\, a_\epsilon)] + constant$$

$$= E_{-a_\epsilon}\left[-\frac{1}{2}\log a_\epsilon - \log\Gamma\,(1/2) - \frac{3}{2}\log\sigma_\epsilon^2 - \frac{{a_\epsilon}^{-1}}{\sigma_\epsilon^2}\right]$$

$$+ E_{-a_\epsilon}\left[-\log A_\epsilon - \log\Gamma\,(1/2) - \frac{3}{2}\log a_\epsilon - \frac{{A_\epsilon}^{-2}}{a_\epsilon}\right] + contant$$

$$= -2\log a_\epsilon - \frac{\left(E_{-a_\epsilon}\left[\frac{1}{\sigma_\epsilon^2}\right] + A_\epsilon^{-2}\right)}{a_\epsilon} + constant_1$$

Where we absorb all terms not involving $a_\epsilon$ into the "constant" terms and we recognize that this is an inverse-Gamma $\left(1, B_{q(a_\epsilon)}\right)$ distribution with

$$B_{q(a_\epsilon)} = E_{-a_\epsilon}\left[\frac{1}{\sigma_\epsilon^2}\right] + A_\epsilon^{-2} = \mu_{q(1/\sigma_\epsilon^2)} + A_\epsilon^{-2}$$

Clearly, the optimal density of $q^*(a_\epsilon)$ depends on distribution of $\sigma_\epsilon^2$.

**D (1.3) Derivation of optimal $q^*(\boldsymbol{\sigma_\epsilon^2})$ density**

To find the optimal form of $q^*(\sigma_\epsilon^2)$ density, we need

$$E_{-\sigma_\epsilon^2}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})] = E_{-\sigma_\epsilon^2}[log\, p(\, \sigma_\epsilon^2|a_\epsilon)] + E_{-\sigma_\epsilon^2}[log\, p(\, Y|\, W, \boldsymbol{I}, \boldsymbol{\Theta})] + constant$$

$$= -\frac{1}{2}\, E_{-\sigma_\epsilon^2}\log|\, \boldsymbol{\Sigma}_\epsilon| - \frac{1}{2}\, E_{-\sigma_\epsilon^2}\left\{\left([\boldsymbol{y} - \boldsymbol{C\Theta_r}]^T\, {\boldsymbol{\Sigma}_\epsilon}^{-1}[\boldsymbol{y} - \boldsymbol{C\Theta_r}]\right)\right\} + E_{-\sigma_\epsilon^2}[\lambda_r^T w_{t-1.i} - \phi_{t-1}]$$

$$- \frac{3}{2}\log\sigma_\epsilon^2 - \frac{1}{\sigma_\epsilon^2}\, E_{-\sigma_\epsilon^2}[1/a_\epsilon] + constant_{\,1}$$

$$= \left\{-\frac{1}{2}\left(\sum_{i=1}^{m} n_i + 1\right) - 1\right\} \log \sigma_{\epsilon}^2$$

$$- \left( \frac{\frac{1}{2}\sum_{r=1}^{R} E[b_{rt}] \left\{ \left( \left\| \boldsymbol{y} - \boldsymbol{C}\mu_{q(\boldsymbol{\Theta_r})} \right\|^2 + tr\left[ \boldsymbol{C}^T \boldsymbol{C}\, \boldsymbol{\Sigma}_{q(\boldsymbol{\Theta_r})} \right] \right) + \mu_{q(1/a_{\epsilon})} \right\}}{\sigma_{\epsilon}^2} \right) + const_{\,2}$$

Where we absorb all terms not involving $\sigma_{\epsilon}^2$ into the constant terms, we have an inverse Gamma $\left(\frac{1}{2}(\sum_{i=1}^{m} n_i + 1), B_{q(\sigma_{\epsilon}^2)}\right)$ distribution with

$$B_{q(\sigma_{\epsilon}^2)} = \frac{1}{2}\sum_{r=1}^{R} E[b_{rt}] \left\{ \left( \left\| \boldsymbol{y} - \boldsymbol{C}\mu_{q(\boldsymbol{\Theta_r})} \right\|^2 + tr\left[ \boldsymbol{C}^T \boldsymbol{C}\, \boldsymbol{\Sigma}_{q(\boldsymbol{\Theta_r})} \right] \right) + \mu_{q(1/a_{\epsilon})} \right\}$$

The optimal density of $q^*(\sigma_{\epsilon}^2)$ depends on the distribution of $q(\boldsymbol{\Theta_r})$ and $q(I_t)$.

**D (1.4) Derivation of optimal $q^*(a_{q_r})$ density**

To find the optimal form of $q^*(a_{q_r})$ density we would have,

$$E_{-a_{q_r}}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})] = E_{-a_{q_r}}\left[log\, p(\, \boldsymbol{\Sigma_{R_r}} \,|a_1, \cdots, a_{q_r})\right] + E_{-a_{q_r}}\left[log\, p(\, a_{q_r})\right] + C^*$$

$$= E_{-\boldsymbol{a_{q_r}}}\left[ \frac{-\upsilon + q_r - 1}{2} \log 2\upsilon \mathrm{diag}\left\{1/a_1, \cdots, 1/a_{q_r}\right\} - \frac{\upsilon + q_r + m}{2} \log\left(|\, \boldsymbol{\Sigma_{R_r}}|\right) \right.$$

$$\left. - tr\left(2\upsilon \mathrm{diag}\left\{1/a_1, \cdots, 1/a_{q_r}\right\}.\, {\boldsymbol{\Sigma_{R_r}}}^{-1}\right) \right]$$

$$+ E_{-\boldsymbol{a_{q_r}}}\left[ -\log A_{q_r} - \log \Gamma\,(1/2) - \frac{3}{2}\log a_{q_r} - \frac{{A_{q_r}}^{-2}}{a_{q_r}} \right] + C^*$$

$$= -\frac{-\upsilon + q + 2}{2} \log a_{q_r} - \frac{E_{-a_{q_r}}\left[\, \boldsymbol{\Sigma_{R_r, q_r q_r}}\right] + A_{q_r}^{-2}}{a_{q_r}} + C^*$$

Where we absorb all terms not involving $a_{q_r}$ into the "$C^*$" terms and we recognize that this is an inverse-Gamma $\left(1/2(\upsilon + q_r), B_{q(a_{q_r})}\right)$ distribution with

$$B_{q(a_{q_r})} = \upsilon\, M_{q({\Sigma_{R_r}}^{-1})q_r q_r} + A_{q_r}^{-2}$$

$M_{q(\Sigma_{R_r}{}^{-1})q_r q_r}$ involves shape parameter of $q^*(\boldsymbol{\Sigma_{R_r}})$. The optimal density of $q^*(a_{q_r})$ depends on the distribution of $q^*(\boldsymbol{\Sigma_{R_r}})$.

**D (1.5) Derivation of optimal $q^*(\boldsymbol{\Sigma_{R_r}})$ density**

To find the optimal form of $q^*(\boldsymbol{\Sigma}_{R_r})$ density we would like,

$$E_{-\boldsymbol{\Sigma}_{R_r}}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})] = E_{-\boldsymbol{\Sigma}_{R_r}}\left[log\, p(\boldsymbol{u}|\, \boldsymbol{\Sigma}_{R_r})\right] + E_{-\boldsymbol{\Sigma}_{R_r}}\left[log\, p(\boldsymbol{\Sigma}_{R_r}|a_1, \cdots, a_{q_r})\right] + C^*$$

$$= E_{-\boldsymbol{\Sigma}_{R_r}}\left[-\frac{1}{2}\log|\boldsymbol{\Sigma}_{R_r}| - \frac{1}{2}\left[\boldsymbol{u_r}^{\boldsymbol{T}}(I_m \otimes \boldsymbol{\Sigma}_{R_r}^{-1})\boldsymbol{u_r}\right]\right]$$

$$+ E_{-\boldsymbol{\Sigma}_{R_r}}\left[-\frac{\upsilon + q_r + m}{2}\log|\boldsymbol{\Sigma}_{R_r}| - tr\,(2\upsilon \mathrm{diag}\{1/a_1, \cdots, 1/a_{q_r}\}.\, \boldsymbol{\Sigma_{R_r}}^{-1})\right]$$

$$+ constant$$

$$= -\frac{\upsilon + q_r + m - 1}{2}\log|\boldsymbol{\Sigma}_{R_r}| - \frac{1}{2}E_{-\boldsymbol{\Sigma}_{R_r}}\left[\boldsymbol{u_r}^{\boldsymbol{T}}(I_m \otimes \boldsymbol{\Sigma}_{R_r}^{-1})\boldsymbol{u_r}\right]$$

$$- E_{-\boldsymbol{\Sigma}_{R_r}}[\, tr\,(2\upsilon \mathrm{diag}\{1/a_1, \cdots, 1/a_{q_r}\}.\, \boldsymbol{\Sigma}_{R_r}{}^{-1}] + C^*$$

$$= -\frac{1}{2}\, tr\,\left(\left\{\sum_{i=1}^{m}\left(\mu_{q(u_{r,i})}\mu_{q(u_{r,i})}{}^{T} + \boldsymbol{\Sigma}_{q(u_{r,i})}\right) + 2\,\upsilon \mathrm{diag}\left(\mu_{q(1/a_1)} \cdots \mu_{q(1/a_{q_r})}\right)\right\}\boldsymbol{\Sigma}_{R_r}{}^{-1}\right)$$

$$- \frac{\upsilon + q_r + m - 1}{2}\log|\boldsymbol{\Sigma}_{R_r}| + C^*$$

Where we absorb all terms not involving $\boldsymbol{\Sigma}_{R_r}$ into the "$C^*$" terms and we recognize that this is an inverse-Wishert $\left((\upsilon + q_r + m - 1), B_{q(\boldsymbol{\Sigma}_{R_r})}\right)$ distribution with

$$B_{q(\boldsymbol{\Sigma}_{R_r})} = \left\{\sum_{i=1}^{m}\left(\mu_{q(u_{r,i})}\mu_{q(u_{r,i})}{}^{T} + \boldsymbol{\Sigma}_{q(u_{r,i})}\right) + 2\,\upsilon \mathrm{diag}\left(\mu_{q(1/a_1)} \cdots \mu_{q(1/a_{q_r})}\right)\right\}$$

Here, $\mu_{q(u_{r,i})}$ is the sub-vector of mean vector $\boldsymbol{\mu}_{q(\boldsymbol{\gamma},\boldsymbol{u})}$ and $\boldsymbol{\Sigma}_{q(u_{r,i})}$ is the sub-matrix of covariance matrix $\boldsymbol{\Sigma}_{q(\boldsymbol{\gamma},\boldsymbol{u})}$ for the $i$th individual random effects vector $\boldsymbol{\mu}_i$ in the $r^{th}$ regime. The optimal density of $q^*(a_{q_r})$ depends on the distribution of random effects part of $q^*(\boldsymbol{\Theta_r})$ and $\boldsymbol{q^*(a_{q_r})}$.

**D (1.6) Derivation of optimal $q^*(\boldsymbol{\Theta}_r)$ density**

For continuous ILD, it is relatively easy to show that the optimal form of $\boldsymbol{q}^*(\boldsymbol{\Theta}_r)$ density is multivariate normal with mean vector $\boldsymbol{\mu}_{q(\boldsymbol{\gamma},\boldsymbol{u})}$ and covariance matrix $\boldsymbol{\Sigma}_{q(\boldsymbol{\gamma},\boldsymbol{u})}$.

$$E_{-\boldsymbol{\Theta}_r=(\boldsymbol{\gamma},\boldsymbol{u})}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})]$$

$$= E_{-(\boldsymbol{\gamma},\boldsymbol{u})}[log\, p(\boldsymbol{Y}|\, \boldsymbol{W}, I, \boldsymbol{\Theta})] + E_{-(\boldsymbol{\gamma},\boldsymbol{u})}\left[log\, p\left(\boldsymbol{\gamma}_r, \boldsymbol{u}_r\, |\boldsymbol{\Sigma}_{R_r}\right)\right] + constant$$

$$= -\frac{1}{2}\, E_{-(\boldsymbol{\gamma},\boldsymbol{u})} \log|\, \boldsymbol{\Sigma}_{\in}| - \frac{1}{2}\, E_{-(\boldsymbol{\gamma},\boldsymbol{u})}\left\{\left([\boldsymbol{y} - \boldsymbol{C}\boldsymbol{\Theta}_r]^T\, \boldsymbol{\Sigma}_{\in}{}^{-1}[\boldsymbol{y} - \boldsymbol{C}\boldsymbol{\Theta}_r]\right)\right\} + E_{-(\boldsymbol{\gamma},\boldsymbol{u})}[\lambda_r^T w_{t-1.i} - \phi_{t-1}]$$

$$-\frac{(p + mq_r)}{2}\log 2\pi - \frac{1}{2}E_{-(\boldsymbol{\gamma},\boldsymbol{u})}\left[\log\left|\begin{matrix}\sigma_{\boldsymbol{\gamma}_r}^{-2} I_p & 0 \\ 0 & I_m \otimes \Sigma_{R_r}^{-1}\end{matrix}\right|\right]$$

$$-\frac{1}{2}\left\{\mu_{q(\boldsymbol{\gamma},\boldsymbol{u})}^T \begin{bmatrix}\sigma_{\boldsymbol{\gamma}_r}^{-2} I_p & 0 \\ 0 & I_m \otimes \Sigma_{R_r}^{-1}\end{bmatrix} \mu_{q(\boldsymbol{\gamma},\boldsymbol{u})} + tr\left(\begin{bmatrix}\sigma_{\boldsymbol{\gamma}_r}^{-2} I_p & 0 \\ 0 & I_m \otimes \Sigma_{R_r}^{-1}\end{bmatrix} \boldsymbol{\Sigma}_{q(\boldsymbol{\gamma},\boldsymbol{u})}\right)\right\}$$

Where we absorb all terms not involving $(\boldsymbol{\gamma}_r, \boldsymbol{u}_r)$ into the "constant" terms and we recognize that this is a multivariate normal density function: $N\left(\mu_{q(\boldsymbol{\Theta}_r)}, \boldsymbol{\Sigma}_{q(\boldsymbol{\Theta}_r)}\right)$ with

$$\boldsymbol{\Sigma}_{q(\boldsymbol{\Theta}_r)} = \left(\mu_{q\left(1/\sigma_\epsilon^2\right)}\left[\sum_{t=1}^{n_i} E[b_{tr}] C^T C\right] + \begin{bmatrix}\sigma_\gamma^{-2} I_p & 0 \\ 0 & I_m \otimes \Sigma_{R_r}^{-1}\end{bmatrix}\right)^{-1}$$

$$\mu_{q(\boldsymbol{\Theta}_r)} = \boldsymbol{\Sigma}_{q(\boldsymbol{\Theta}_r)}\left(\mu_{q\left(1/\sigma_\epsilon^2\right)}\left[\sum_{t=1}^{n_i} E[b_{tr}] C^T \boldsymbol{y}\right]\right)$$

**D (1.7) Derivation of optimal $q^*(\boldsymbol{\lambda}_r)$ density**

To find the optimal form of $q^*(\boldsymbol{\lambda}_r)$ density we would like,

$$E_{-\,\boldsymbol{\lambda}_r}[log\, p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})]$$

$$= E_{-\,\boldsymbol{\lambda}_r}[log\, p(\boldsymbol{I}\, |\boldsymbol{\Lambda})] + E_{-\,\boldsymbol{\lambda}_r}[log\, p(\, \boldsymbol{\Lambda}|\boldsymbol{\vartheta})] + E_{-\,\boldsymbol{\lambda}_r}[log\, p(\, \boldsymbol{\vartheta})] + constant$$

$$= E_{-\,\boldsymbol{\lambda}_r}\left[log\, P\left(I_{t,i} = r|W_{t-1,i},\, \lambda_r\right)\right] + E_{-\,\boldsymbol{\lambda}_r}[log\, P(\, \lambda_r|\vartheta_r)] + constant$$

Here, *constant* term comprises all terms unrelated to $\lambda_r$ (even terms involving only $\vartheta_r$

because it does not change with respect to $\lambda_r$). After plugging the corresponding distributions, and taking the expectations on other parameters not involving $\lambda_r$, we recognize that it resembles a normal density function with

$$\Sigma_{\lambda_r}^{-1} = E[\vartheta_r].I + \sum_{t=1}^{n_i} 2\ \rho(\xi_{r,t}) C_t\ C_t^T$$

$$E(\lambda_r) = \mu_{\lambda_r} = \Sigma_{\lambda_r} \left( \sum_{t=1}^{n_i} \left[ E[b_{rt}] - \frac{1}{2} + 2\zeta_t \rho(\xi_{r,t}) \right] C_t \right)$$

However, Bouchard (2007) shows that $\xi_{rt}$ and $\zeta_t$ obtained from the following formula:

$$\xi_{rt}{}^2 = C_t^T E[\lambda_r \lambda_r^T] C_t + \zeta_t{}^2 - 2\,\zeta_t\, \mu_{\lambda_r}{}^T C_t$$

And

$$\zeta_t = \frac{\frac{1}{2}\left(\frac{R}{2}-1\right) + \sum_{r=1}^{R} \rho(\xi_{r,t})\, \mu_{\lambda_r}{}^T C_t}{\sum_{r=1}^{R} \rho(\xi_{r,t})}$$

With

$$\rho(\xi_{r,t}) = \frac{1}{2\xi_{r,t}} \left[ \frac{1}{1+\exp(-\xi_{r,t})} - \frac{1}{2} \right]$$

The optimal density of $q^*(\lambda_r)$ depends on distribution of $q^*(\boldsymbol{I_t})$ and $q^*(\vartheta_r)$ .

**D (1.8) Derivation of optimal $q^*(\vartheta_r)$ density**

To find the optimal form of $q^*(\vartheta_r)$ density, we would like

$$E_{-\vartheta_r}[log\ p(Y, W, \boldsymbol{I}, \boldsymbol{\Theta})] = E_{-\vartheta_r}[\log p(\boldsymbol{\Lambda}|\boldsymbol{\vartheta})] + E_{-\vartheta_r}[\log p(\boldsymbol{\vartheta})] + C^*$$

$$= E_{-\vartheta_r}[\log P(\lambda_r|\vartheta_r)] + E_{-\vartheta_r}[\log p(\vartheta_r|c_0, b_0)] + C^*$$

$$= (c_0 - 1)\log \vartheta_r - b_0 \vartheta_r + \frac{n_l}{2}\log \vartheta_r - \frac{\vartheta_r}{2}\, E_{\vartheta_r}[\lambda_r^T \lambda_r] + C^*$$

After plugging the corresponding distributions, and taking expectations on other parameters not involving $\vartheta_r$ , we recognize that it has a form of Gamma density function $G\ (c_r, b_r)$ with

$$c_r = c_0 + \frac{n_l + 1}{2}$$

And $$b_r = b_0 + \frac{1}{2} E[\lambda_r^T \lambda_r] = b_0 + \frac{1}{2} \left[\mathrm{tr}\left(\Sigma_{\lambda_r} + \mu_{\lambda_r}{}^T \mu_{\lambda_r}\right)\right]$$

## 1. Derivation of lower bound on the marginal log-likelihood

We monitor the bound on the marginal log-likelihood of $F[q(\boldsymbol{I}), q(\boldsymbol{\Theta})]$ defined in (1.4) to manage the convergence of our MFVB algorithm. This is calculated as:

$$F[q(\boldsymbol{I}), q(\boldsymbol{\Theta})] = E_q[\log p(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{I}, \boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\Lambda) - \log q(\boldsymbol{I}) - \log q(\boldsymbol{\Theta})]$$

More specifically:

$$\begin{aligned} F[q(I), q(\boldsymbol{\Theta})] = {} & E_{\boldsymbol{q}}\left[\sum_{t=1}^{n_i}\sum_{r=1}^{R} b_{rt}\left\{\log\left\{p\left(y_{t,i}|w_{t,i}, I_t, \boldsymbol{\Theta}, \Lambda\right) p(I_t|\Lambda)\right\}\right\}\right] + E_q[\log p(\sigma_\epsilon^2|a_\epsilon)] \\ & + E_q[\log p(a_\epsilon)] \\ & + \sum_{r=1}^{R}\left\{E_q\left[\log p\left(\boldsymbol{\gamma_r}, \boldsymbol{u_r}|\boldsymbol{\Sigma}_{R_r}\right)\right] + E_q\left[\log p\left(\boldsymbol{\Sigma_{R_r}}|a_1, \cdots, a_{q_r}\right)\right] + E_q\left[\log p\left(a_{q_r}\right)\right]\right. \\ & \left. + E_q[\log p(\lambda_r|\vartheta_r)] + E_q[\log p(\vartheta_r)]\right\} - E_q\left[\sum_{t=1}^{n_i}\sum_{r=1}^{R} b_{rt}\log\{q(I_t)\}\right] \\ & - E_q[\log q(\sigma_\epsilon^2)] - E_q[\log q(a_\epsilon)] \\ & - \sum_{r=1}^{R}\left\{E_q[\log q(\boldsymbol{\gamma_r}, \boldsymbol{u_r})] + E_q\left[\log q\left(\boldsymbol{\Sigma_{R_r}}\right)\right] + E_q\left[\log q\left(a_{q_r}\right)\right]\right. \\ & \left. + E_q[\log q(\lambda_r|\vartheta_r)] + E_q[\log q(\vartheta_r)]\right\} \end{aligned}$$

We will now derive an expression for each part of $F[q(I), q(\boldsymbol{\Theta})]$ and then state the full expression to obtain the lower bound of marginal log-likelihood for our model.

$$E_{\boldsymbol{q}}\left[\sum_{t=1}^{n_i}\sum_{r=1}^{R} b_{rt}\log\{p(y_{t,i}|w_{t,i},I_t,\boldsymbol{\Theta},\Lambda)\,p(I_t=r|\Lambda)\}\right]$$

$$= -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\,E_q\log|\,\boldsymbol{\Sigma}_{\in}| - \frac{1}{2}\,E_q\left\{b_{rt}\left([\boldsymbol{y}-\boldsymbol{C\Theta_r}]^T\,{\boldsymbol{\Sigma}_{\in}}^{-1}[\boldsymbol{y}-\boldsymbol{C\Theta_r}]\right)\right\}$$

$$+ E_q\{b_{rt}\,[\lambda_r^T w_{t-1.i} - \phi_{t-1}]\}$$

$$E_q[\log p(\,\sigma_\epsilon^2|a_\in)] = -\frac{1}{2}\,E_q[\log a_\in] - \log\Gamma\,(1/2) - \frac{3}{2}\,E_q[\log\sigma_\epsilon^2] - E_q[1/a_\in\,\sigma_\epsilon^2\,]$$

$$E_q[\log p(\,a_\in)] = \frac{1}{2}\log(A_\epsilon^{-2}) - \log\Gamma\,(1/2) - \frac{3}{2}\,E_q[\log a_\in] - \,A_\epsilon^{-2}\,E_q[1/a_\in]$$

$$E_q\left[\log p(\,a_{q_r})\right] = \frac{1}{2}\log\left(A_{q_r}^{-2}\right) - \log\Gamma\,(1/2) - \frac{3}{2}\,E_q\left[\log a_{q_r}\right] - \,A_{q_r}^{-2}\,E_q\left[1/a_{q_r}\right]$$

$$E_q[\log p(\vartheta_r\,)] = (c_0-1)\,E_q[\log\vartheta_r] - b_0E_q[\vartheta_r] + c_0\log b_0 - \log\Gamma(c_0)$$

$$E_q[\log p(\,\lambda_r|\vartheta_r\,)] = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\,E_q\log(\vartheta_r) - E_q\left(\frac{\vartheta_r}{2}\,\lambda_r^T\lambda_r\right)$$

$$E_q\left[\log p(\,\boldsymbol{\gamma_r},\boldsymbol{u_r}\,|\boldsymbol{\Sigma}_{R_r})\right]$$

$$= -\frac{(p+mq_r)}{2}\log 2\pi - \frac{1}{2}E_q\left[\log\begin{vmatrix}\sigma_{\boldsymbol{\gamma_r}}^{-2}I_p & 0\\ 0 & I_m\otimes\Sigma_{R_r}^{-1}\end{vmatrix}\right]$$

$$-\frac{1}{2}\,E_q\left\{\boldsymbol{\Theta_r}^T\begin{bmatrix}\sigma_{\boldsymbol{\gamma_r}}^{-2}I_p & 0\\ 0 & I_m\otimes\Sigma_{R_r}^{-1}\end{bmatrix}\boldsymbol{\Theta_r}\right\}$$

$$E_q\left[\log p(\,\boldsymbol{\Sigma_{R_r}}\,|a_1,\cdots,a_{q_r}\,)\right]\sim E_q\left[\log IW\left(\nu+q_r-1,2\nu\,\mathrm{diag}\{1/a_1,\cdots,1/a_{q_r}\}\,\right)\right]$$

And all the terms $E_q\left[\sum_{t=1}^{n_i}\log\{q(I_t)\}\right]$, $E_q[\log q(\,\sigma_\epsilon^2)]$, $E_q[\log q(\,a_\in)]$, $E_q[\log q(\,\boldsymbol{\gamma_r},\boldsymbol{u_r}\,)]$, $E_q\left[\log q(\,\boldsymbol{\Sigma_{R_r}}\,)\right]$, $E_q\left[\log q(\,a_{q_r})\right]$, $E_q[\log q(\,\lambda_r|\vartheta_r\,)]$ and $E_q[\log q(\vartheta_r\,)]$ related to optimal q-densities for latent variables derived in **D (1.1)** to **D (1.8)** .

After evaluating the expectations and cancelling, we arrive at the expression for the lower bound on the marginal likelihood.

$$F[q(I), q(\boldsymbol{\Theta})] = \frac{R(p + mq_r + 1) - N\log 2\pi}{2}$$

$$- \sum_t \sum_r \frac{q_{rt}}{2} \Bigg\{ [D_{rt}^T \, \Sigma_\epsilon^{-1} D_{rt} + tr(\Sigma_\epsilon^{-1} E_{rt})]$$

$$+ n_i \left[ \log B_{q(\sigma_\epsilon^2)} - \psi \left[ \frac{1}{2} \left( \sum_{i=1}^{m} n_i + 1 \right) \right] \right] \Bigg\} - \sum_t \sum_r q_{rt} \log \frac{p_{rt}}{q_{rt}}$$

$$- \frac{1}{2} \left( \sum n_i + 1 \right) \log B_{q(\sigma_\epsilon^2)} - B_{q(a_\epsilon)} \left[ \mu_{q(1/\sigma_\epsilon^2)} + A_\epsilon^{-2} \right] + B_{q(\sigma_\epsilon^2)} \mu_{q(1/\sigma_\epsilon^2)}$$

$$- \log B_{q(a_\epsilon)}$$

$$+ \sum_r \Bigg\{ \log C_{q_r}, \nu + q_r + m - 1 - \log C_{q_r}, \nu + q_r - 1 + \log A_{q_r} + \log \Gamma\,(1/2)$$

$$+ \log \Gamma \left( \frac{1}{2} (\nu + q_r) \right) - \frac{1}{2} \Bigg\}$$

$$+ \left\{ \frac{3\,(\nu + q_r - 2)}{2} \log B_{q(\,a_{q_r})} - \frac{7(\nu + q_r - 1)}{2} \psi \left( \frac{\nu + q_r}{2} \right) \right\}$$

$$+ \Bigg\{ \log \left| B_{q(\,\Sigma_{R_r})} \right| \left( \frac{\nu + 2q_r + 1}{2} \right) - \frac{2\nu + 3q_r + m}{2} \psi(\nu + q_r + m - 1)$$

$$+ B_{q(\,a_{q_r})} \left[ \frac{A_{q_r}^{-2}}{\frac{\nu + q_r}{2}} + \mu_{q(1/a_{q_r})} \right] + \nu M_{q(\,\Sigma_{R_r}{}^{-1})} \mu_{q(1/a_{q_r})} + \frac{1}{2} (\nu + q_r + m - 1) \Bigg\}$$

$$- \frac{1}{2} \left\{ \mathrm{tr} \left\{ \Sigma_{\boldsymbol{\gamma_r}, \boldsymbol{u_r}}{}^{-1} \Sigma_{\boldsymbol{q}(\boldsymbol{\gamma_r}, \boldsymbol{u_r})} \right\} + \mu_{\boldsymbol{q}(\boldsymbol{\gamma_r}, \boldsymbol{u_r})}^T \Sigma_{\boldsymbol{\gamma_r}, \boldsymbol{u_r}}{}^{-1} \mu_{\boldsymbol{q}(\boldsymbol{\gamma_r}, \boldsymbol{u_r})} \right\}$$

$$+ (c_0 - 1/2) \big( \log b_r + \psi(c_r) \big) - \frac{c_r}{b_r} \left[ b_0 + \frac{1}{2} \left( \mu_{\lambda_r}^T \mu_{\lambda_r} \right) \right] + \frac{1}{2} \log \Sigma_{\lambda_r}$$

Where, $\Gamma(\cdot)$ and $\psi(\cdot)$ denote the gamma and digamma functions, respectively.

$$D_{rt} = \left(y_{t,r} - \boldsymbol{C}_{t,r}\mu_{q(\boldsymbol{\gamma_r},\boldsymbol{u_r})}\right)^T$$

$$E_{rt} = \boldsymbol{C}_{t,r}{}^T \boldsymbol{C}_{t,r}\Sigma_{\boldsymbol{q}(\boldsymbol{\gamma_r},\boldsymbol{u_r})}$$

$$\Sigma_{\epsilon}^{-1} = \frac{\frac{1}{2}(\sum_{i=1}^{m} n_i + 1)}{B_{q(\sigma_{\epsilon}^2)}}$$

$$p_{rt} \approx f(\boldsymbol{C_t}, \lambda) = \exp(\lambda_r^T \boldsymbol{C_t}) \exp(-\phi_t)$$

$$\phi_t = \zeta_t + \sum_{r=1}^{R} \frac{\lambda_r^T w_t - \zeta_t - \xi_{r,t}}{2} + \rho(\xi_{r,t})\left[(\lambda_r^T \boldsymbol{C_t} - \zeta_t)^2 - \xi_{r,t}\right] + \log\left(1 + e^{\xi_{r,t}}\right)$$

$$\rho(\xi_{r,t}) = \frac{1}{2(\xi_{r,t})}\left[\frac{1}{1 + exp(-\xi_{r,t})} - \frac{1}{2}\right]$$

$$q_{rt} = \frac{\exp \eta_r^T}{\sum_{r=1}^{R} \exp(\eta_r^T)}$$

The variational hyper parameters such as scale and shape, $\xi_{r,t}$and $\zeta_t$ respectively control the tightness of the bound for $r \in \{1,2,\cdots,R\}$ and $t \in \{1,2,\cdots,n_i\}$ for all $i \in \{1,2,\cdots,m\}$.

### 3.9.2 Additional simulation results

This supplemental file includes the results for additional scenarios.

**Table 3.3** Bias of the point estimates (i.e., posterior means) for the average inertias and threshold, and coverage of 95% credible intervals, based on 100 fitted models per sample size.

| **Gamm Phi [0.3,0.5]** | | | **Bias (estimate - true value)** | | | | **Coverage** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | T | | | | T | | | |
| **Parameters** | Method | M | 30 | 50 | 100 | 150 | **30** | 50 | 100 | **150** |
| **Beta 10** | MCMC | **50** | -0.06 | -0.03 | -0.02 | -0.02 | **73** | 81 | 94 | 96 |
| | | 100 | -0.04 | -0.03 | -0.01 | -0.01 | 72 | 84 | 93 | 95 |
| | | 150 | -0.05 | -0.02 | -0.01 | -0.01 | 73 | 81 | 96 | 96 |
| | MFVB | 50 | -0.06 | -0.02 | -0.02 | -0.01 | **71** | 81 | 95 | 97 |
| | | 100 | -0.04 | -0.02 | -0.02 | -0.01 | 75 | 85 | 96 | 95 |
| | | 150 | -0.03 | -0.02 | -0.01 | -0.01 | 79 | 83 | 96 | 98 |
| **Beta 20** | MCMC | 50 | -0.08 | -0.03 | -0.02 | -0.02 | **77** | 82 | 94 | 96 |
| | | 100 | -0.05 | -0.02 | -0.01 | -0.01 | 78 | 83 | 93 | 96 |
| | | 150 | -0.05 | -0.02 | -0.01 | -0.01 | 80 | 82 | 96 | 96 |
| | MFVB | 50 | -0.05 | -0.02 | -0.02 | -0.01 | **75** | 84 | 95 | 96 |
| | | 100 | -0.03 | -0.02 | -0.02 | -0.01 | 78 | 83 | 93 | 97 |
| | | 150 | -0.03 | -0.02 | -0.01 | -0.01 | 81 | 83 | 96 | 98 |
| **Tau** | MCMC | 50 | -0.07 | -0.03 | -0.02 | -0.02 | **79** | 82 | 94 | 96 |
| | | 100 | -0.07 | -0.02 | 0.01 | 0.01 | 76 | 81 | 94 | 95 |
| | | 150 | -0.04 | -0.02 | 0.02 | 0.01 | 76 | 82 | 95 | 95 |
| | MFVB | 50 | -0.05 | -0.02 | 0.02 | 0.01 | **75** | 83 | 95 | 97 |
| | | 100 | -0.04 | -0.01 | 0.01 | 0.01 | 78 | 81 | 94 | 96 |
| | | **150** | -0.02 | -0.01 | 0.01 | 0.00 | 79 | 83 | 96 | 97 |

**Continuations of Table 3.3**

| **Gamm Phi [0.2,0.8]** | | | **Bias (estimate - true value)** | | | | **Coverage** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | T | | | | T | | | |
| **Parameters** | Method | M | 30 | 50 | 100 | 150 | **30** | 50 | 100 | **150** |
| **Beta 10** | MCMC | **50** | -0.08 | -0.03 | -0.02 | -0.02 | **74** | 81 | 94 | 96 |
| | | 100 | -0.07 | -0.03 | -0.01 | -0.01 | 75 | 86 | 94 | 95 |
| | | 150 | -0.05 | -0.02 | -0.01 | -0.01 | 74 | 81 | 96 | 96 |
| | MFVB | 50 | -0.07 | -0.02 | -0.02 | -0.01 | **72** | 81 | 95 | 97 |
| | | 100 | -0.05 | -0.02 | -0.02 | -0.01 | 73 | 85 | 95 | 97 |
| | | 150 | -0.03 | -0.02 | -0.01 | -0.01 | 79 | 85 | 97 | **98** |
| **Beta 20** | MCMC | 50 | -0.06 | -0.03 | -0.02 | -0.02 | **73** | 82 | 94 | 96 |
| | | 100 | -0.05 | -0.02 | -0.01 | -0.01 | 78 | 85 | 95 | 96 |
| | | 150 | -0.05 | -0.02 | -0.01 | -0.01 | 80 | 83 | 96 | 96 |
| | MFVB | 50 | -0.06 | -0.02 | -0.02 | -0.01 | **71** | 82 | 95 | 97 |
| | | 100 | -0.05 | -0.02 | -0.02 | -0.01 | 78 | 83 | 96 | 97 |
| | | 150 | -0.03 | -0.02 | -0.01 | -0.01 | 81 | 85 | 97 | **98** |
| **Tau** | MCMC | 50 | -0.06 | -0.03 | 0.02 | 0.02 | **74** | 82 | 93 | 96 |
| | | 100 | 0.05 | 0.02 | 0.01 | 0.01 | 76 | 81 | 93 | 95 |
| | | 150 | 0.04 | 0.02 | 0.02 | 0.01 | 76 | 82 | 95 | 96 |
| | MFVB | 50 | -0.05 | -0.02 | 0.02 | 0.01 | **71** | 83 | 95 | 97 |
| | | 100 | 0.04 | 0.01 | 0.01 | **0.0** | 78 | 83 | 96 | 97 |
| | | **150** | 0.02 | 0.01 | 0.01 | **0.0** | 79 | 83 | 97 | **98** |

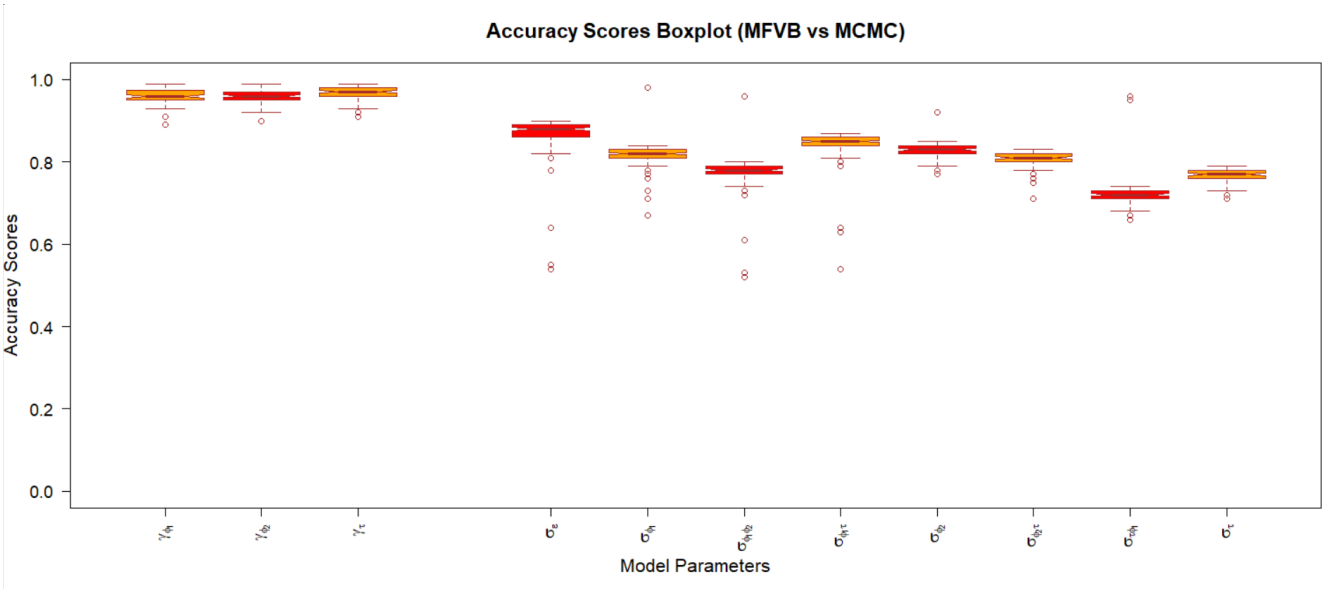



**Figure 3.5** The box plots of accuracy for 100 simulated data sets under the scenario where m=50 and T=100 data points are generated.

**Note:** The accuracy scores are based on the integrated absolute difference in posterior density between MCMC and MFVB.

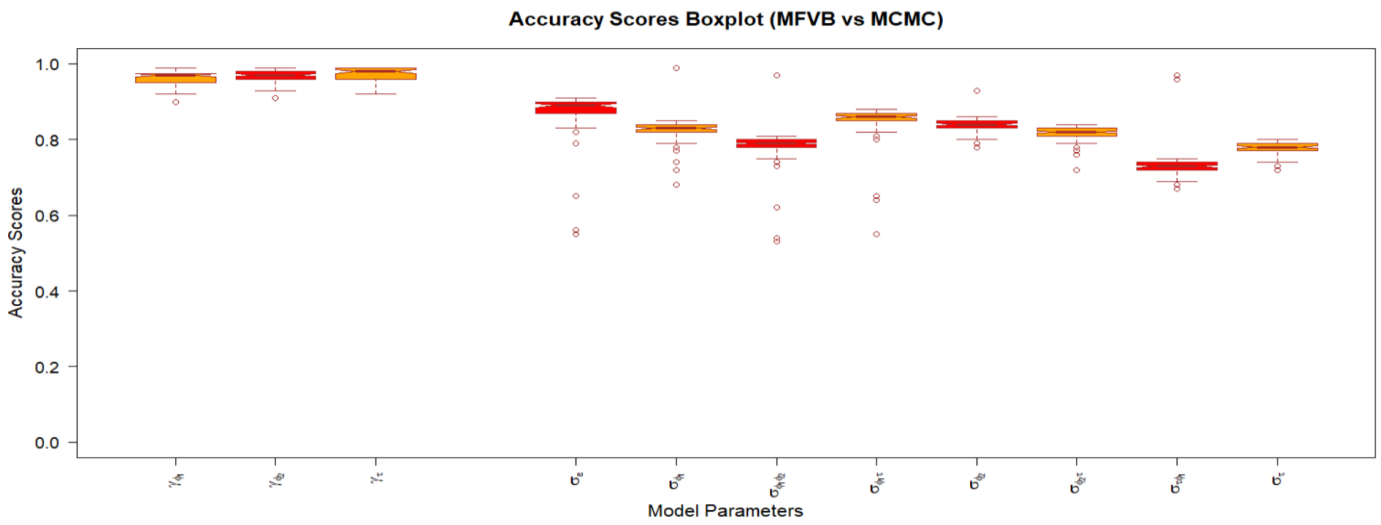



**Figure 3.6** The box plots of accuracy for 100 simulated data sets under the scenario where m=150 and T=150 data points are generated.

**Note:** The accuracy scores are based on the integrated absolute difference in posterior density between MCMC and MFVB.

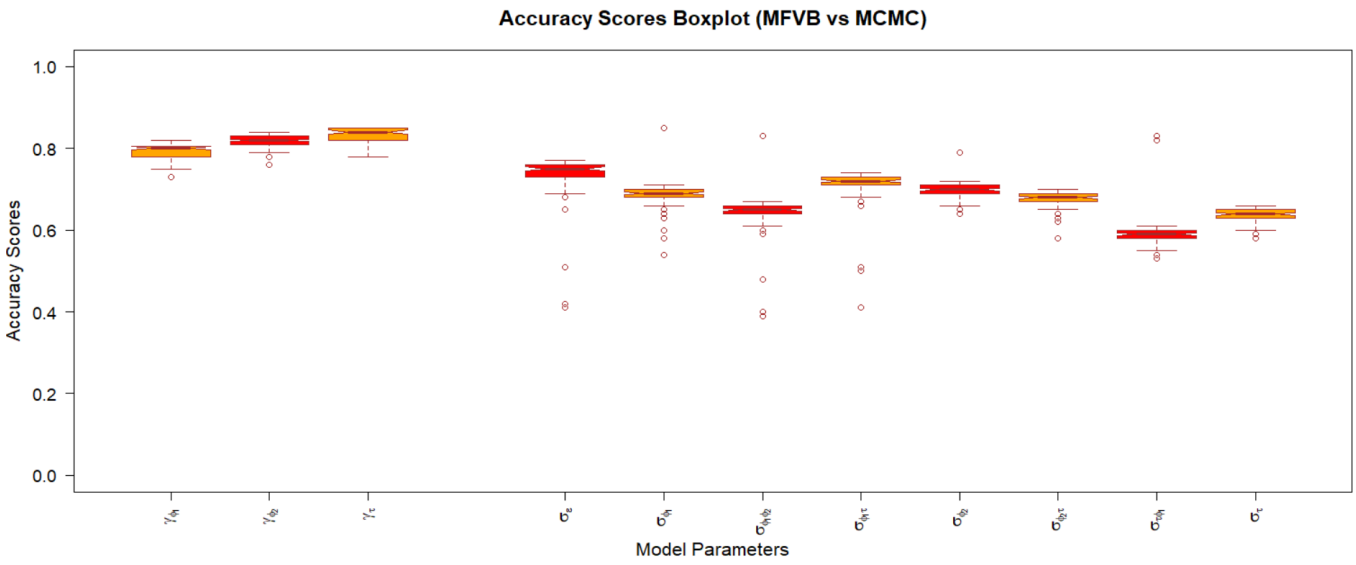



**Figure 3.7** The box plots of accuracy for 100 simulated data sets under the scenario where m=100 and T=30 data points are generated.

**Note:** The accuracy scores are based on the integrated absolute difference in posterior density between MCMC and MFVB.

# CHAPTER 4: APPROXIMATE BAYESIAN ESTIMATION ALGORITHM FOR CONTINUOUS TIME MULTIVARIATE MIXED HIDDEN MARKOV MODEL IN INTENSIVE LONGITUDINAL DATA

## 4.1 Overview

The previous chapter focused on the discrete-time model for ILD, presenting an efficient estimation method for the ML-TAR (2,1,1) model and comparing the proposed MFVB algorithm with MCMC in identifying the optimal threshold for state-dependent relationships between responses and covariates. However, the discrete-time (DT) model assumes longitudinal responses are observed at regular times, while the continuous-time (CT) models consider time as a continuous variable, accommodating irregular or sparse observations in longitudinal responses.

Existing research indicated that continuous time models outperform discrete-time models in uncovering dynamic associations and heterogeneity in ILD [1-4]. Often, longitudinal responses are characterized by a set of underlying hidden states processes and Hidden Markov Models are well suited to describe the relationship between observed responses and latent states. Much work on continuous time models of HMM (CT-HMM) has been devoted to frequentist approach with inference via the expectation maximization and traditional MCMC approach [5-14]. However, computational complication arises due to the model complexity and large-scale inference of such models.

This chapter addresses computational challenges of the CT-mMixHMM model by developing a sequential variational Bayes (SVB) algorithm and comparing its precision relative to MCMC. The findings demonstrated that the SVB method outperformed MCMC in identifying absorbing states and estimating transition probabilities, offering faster computation and stable convergence. In case of empirical application, age effects showed clear progression through hypertension stages, with increased treatment needs and higher mortality risk, underscoring the value of early intervention. The CT-mMixHMM model's fitted state sequences from SVB closely matched observed data for individual MFUS

participants, capturing general trends effectively and supporting the model's validity for longitudinal blood pressure analysis. This study contributes to ongoing efforts to improve the estimation strategies of continuous time models of ILD. It highlights the significance of proposed algorithm performance and suggests that SVB can be a valuable alternative tool for mitigating computational issues in large scale health data analysis.

*Publication Details*

A. Rahman, D. Jiang, L.M. Lix & P. Yang. "Approximate Bayesian Estimation Algorithm for Continuous Time Multivariate Mixed Hidden Markov Model in Intensive Longitudinal Data," *Manuscript in preparation. 2026.*

### 4.2 Abstract

**Background:** Intensive longitudinal data (ILD) involves collecting repeated measurements in a frequent, natural, and minimally invasive manner. Analyzing ILD requires advanced statistical models capable of handling continuous-time (CT) data frameworks. The continuous-time multivariate mixed hidden Markov model (CT-mMixHMM) effectively addresses complex hidden states of outcome variable(s) in a continuous-time setting, enriching our understanding of temporal dynamics. Bayesian estimation methods are particularly suited for these analyses due to their flexibility with complex model structures and their ability to incorporate prior information, making them ideal for handling ILD challenges. However, traditional Bayesian methods, such as Markov Chain Monte Carlo (MCMC), tend to be computationally intensive CT-mMixHMM, presenting significant computational challenges.

**Purpose and Objectives:** This study aims to develop and evaluate computationally efficient approximate Bayesian estimation algorithms tailored for CT models in ILD. Specifically, we develop a Sequential Variational Bayes (SVB) algorithm for CT-mMixHMMto handle irregularly spaced observations and compare its performance with MCMC in simulated and real-data scenarios.

**Methods:** The proposed algorithm used optimization-based techniques to approximate posterior distributions. The SVB algorithm for CT-mMixHMM was developed and first evaluated on simulated data, then applied to hypertension progression data from the Manitoba Follow-Up Study (MFUS). The performance of the algorithm was assessed using prediction accuracy (through metrics such as the ability to predict future hidden state classification, and Mean Absolute Percentage Error), Posterior predictive checks (PPCs) along with computational efficiency. The Deviance Information Criterion (DIC) was employed to determine the optimal number of hidden states when the proposed algorithm for CT-mMixHMMs was initialized with a large set of hidden states using simulated data. The results from both simulation studies and real-world data application were compared with MCMC to assess robustness and relative efficiency.

**Results:** The SVB method outperformed MCMC in identifying absorbing states and estimating transition probabilities, offering faster computation and stable convergence when sample size and measurement occasions are large. The influence of age demonstrated a distinct progression across hypertension stages, resulting in greater treatment requirements and elevated mortality risk. These findings underscore the importance of early intervention among MFUS participants. The CT-mMixHMM model-fitted state sequences from SVB closely matched observed data for individual MFUS participants, capturing general trends effectively and supporting the CT-mMixHMM's validity for longitudinal blood pressure analysis.

**Conclusion:** This research introduces fast approximate Bayesian estimation algorithms for CT-mMixHMM models. The SVB algorithm efficiently handles complex and large datasets, thereby facilitating accurate analysis of intensive longitudinal data. By solving computational issues in such models, the study provides valuable tools for interdisciplinary applications, such as in healthcare. These advancements will support data-driven decision-making and tailored interventions.

## 4.3 Background

Continuous-time (CT) models account for irregular observation intervals by treating time as a continuum, enabling flexible modeling of complex temporal dynamics in ILD. They allow us to explore how expected change in underlying phenomena depends on the time interval between measurements. For example, CT models such as continues-time hidden Markov models (CT-HMMs) have been used to explore the state transition dynamics among blood pressure states and to investigate the risk factors that are affecting the progression of hypertension in older adults, providing insights into how a patient's blood pressure fluctuates and responds to treatment [15]. The following paragraphs focus on the review of continuous time HMMs and mixed effects HMMs and their applications in diverse fields of ILD datasets.

Hidden Markov models (HMMs) - also known as dependent mixture models - are commonly used for ILD due to their ability to relate an observed outcome process to an underlying, unobservable finite-state transition process. In these models, the observed data are governed by an unseen random process that adheres to Markov property. The popularity of HMMs stems from their versatility and mathematical tractability; they offer linear-time likelihood computation, straightforward handling of ignorable missing observations, and the ability to easily derive the marginal, conditional and forecast distributions [16, 17]. Consequently, HMMs have garnered considerable attention in various disciplines involving complex ILD [5, 12, 17, 18].

While traditional HMMs require fixed discrete time (DT) steps, often necessitating biased imputations for missing intervals -CT-HMMs allow transitions to occur at any point in CT [19, 20]. This makes CT-HMMs suitable for handling irregularly sampled temporal data, such as clinical measurements collected during sporadic hospital visits [21]. To date, research on CT-HMMs has predominantly relied on frequentist frameworks, often relying on the expectation-maximization (EM) algorithm [5, 22].

Early research in DT HMMs integrated generalized linear mixed effects model (GLMM) framework to incorporate fixed and random effects for individual covariates. This approach was applied

to multiple sclerosis patients' data to identify the heterogeneity in reducing mean lesion count among such patients [8], a framework later extended to multivariate setting in smoking cessation trial dataset [23]. Other developments include multilevel HMMs with Poisson-Lognormal emission distribution to identify the temporal dynamics of neural activity [24]. Despite these advancements and comprehensive reviews of mixed effects HMMs applied to longitudinal data [16], practical applications remain hindered by the DT framework and the computational complexities involved in estimating large and unknown state spaces using frequentist methods.

***Bayesian Approaches and CT-HMMs***

To address these limitations, researchers have pivoted toward Bayesian frameworks. For example, [23] applied a fully Bayesian framework to multivariate longitudinal data in the context of smoking cessation trials, while [25] used MCMC for CT-HMMs to study dementia progression. However, they did not address the estimation process with a large unknown number of states due to high computational demands. More complex models have since emerged, such as the reversible jump MCMC algorithm used to model COPD progression [7] and Bayesian CT-HMMs incorporating covariate selection and measurement error to identify risk factors for cigarette smoking abstinence [22].

***Scalability via Variational Bayes***

Despite these strides, literature remains scarce regarding optimization-based approximate Bayesian estimation studies on CT models that include both fixed effects and random effects, specifically focusing on the state transition dynamics.

Unlike traditional Bayesian sampling methods, optimization-based approaches such as variational Bayes (VB) avoid Markov chain sampling by replacing the posterior with a tractable approximating distribution and optimizing it directly. VB constructs an objective (typically an evidence lower bound) and uses it to fit the variational parameters, which often makes the procedure substantially more scalable than methods such as MCMC or exact Bayesian inference [26]. Here, "scalability" refers to how

computational demands change as data size and model complexity increase, and is typically assessed using metrics such as computational complexity, convergence speed, memory usage, and runtime. In this sense, VB has demonstrated strong computational efficiency for large, structured Bayesian models, including variational Bayesian analyses for sparse item response models and two-part latent variable models [49-51]. Specifically, [50] developed variational Bayesian inference for sparse item response theory models and reported computational benefits by reducing the need for costly sampling. Similarly, [49] used VB for a two-part latent variable model and demonstrated that variational optimization yields accurate inference while improving runtime compared with more simulation-heavy alternatives. [51] further showed that efficient and accurate variational inference can be achieved for multilevel threshold autoregressive models, highlighting the practical speed-accuracy trade-off of VB in complex hierarchical settings. Common VB schemes include mean-field variational Bayes (MFVB) and sequential variational Bayes (SVB), which differ in how they factorize and update the latent-variable approximation.

An approximate Bayesian approach was developed to facilitate the computation for CT-HMMs and demonstrated its performance in the disease progression dynamics in real-world data application like coronary allograft vasculopathy [28]. A key limitation of these algorithmic studies is their reliance on batch-based data processing and fixed effects in CT models for state transition dynamics in ILD datasets. For example, large-scale inference challenges for unknown hidden states have led to the development of batch-based methods like the MFVB algorithm for CT-HMMs, as demonstrated in active learning data [27]. To the best of our knowledge, no studies in literature have introduced CT models incorporating both fixed and random effects to account for unit-specific heterogeneity in the state transition dynamics of multivariate study variables in ILD datasets. Additionally, the application of Sequential Variational Bayes (SVB) approaches remains unexplored to tackle computational issues in large scale inference within longitudinal data processing techniques.

This study evaluated the performance of the SVB algorithm for CT-mMixHMM, comparing it to

the standard MCMC in both synthetic and real data. The objectives are: (1) to assess how varying sample sizes affect the performance of the SVB algorithm in terms of parameter accuracy and computational time compared to the MCMC approach in simulated data; and (2) to examine how CT-mMixHMM can account for attrition and capture dynamic hypertension progression in MFUS data, by modeling state transition and covariate effects.

## 4.4 Methods

### 4.4.1 *Data Sources*

After validating the proposed algorithm (detailed in Section 4.4.2.6.) with simulated data, we applied it to the Manitoba Follow-Up Study (MFUS), one of the world's longest-running studies of health and aging. Established in 1948 at the University of Manitoba, MFUS has prospectively documented health and disease in a cohort of 3,983 male Royal Canadian Air Force aircrew recruits [29]. For this analysis, we selected a subset of $N = 1007$ participants who met the ILD requirements, specifically those with 20 to 48 repeated observations recorded between 1948 and 2019.

The cohort's baseline measurement of systolic and diastolic blood pressure (mean ± standard deviation) was 121 ± 10mmHg, and 76 ± 8 mmHg. This dataset is suitable to explore the modeling disease transition dynamics among participants due to high temporal accuracy. Disease transition dynamics track how individuals move between health states (e.g., healthy, subclinical, diseased) over time, such as in the progression of hypertension (Normal, Elevated, Hypertension stage 1, Hypertension stage 2). In this study, systolic blood pressure (SBP) and diastolic blood pressure (DBP) serve as multivariate longitudinal responses. We treated these measures as indicator of an underlying, latent hypertension progression process that is not directly observable.

### 4.4.2 *Continuous time multivariate mixed hidden Markov model (CT-mMixHMM) development*

#### 4.4.2.1 *Notation*

Let $\boldsymbol{y_{it}}$ denote a *R*-dimensional multivariate longitudinal response for subject $i$ $(i = 1, 2, \cdots, N)$ at time $t(t = 1,2, \cdots, n_i)$ $n_i$ represents the total number of observations or time points for subject $i$ and $z_{it}$ denote the hidden process for subject $i$ at some common time points $t$. Let us denote $\boldsymbol{y_i}$ as the vector comprising all observations of $\boldsymbol{y_{it}} = [y_{i1t}, y_{i2t}, \cdots, y_{iRt}]'$ for subject $i$ , and $\boldsymbol{y}$ as the collection of all $\boldsymbol{y_i}$ for every subject $i$ . Furthermore, we denoted $\boldsymbol{z_i}$ to comprise all value of $z_{it}$ for subject $i$ , and $\boldsymbol{z}$ to be all values of $\boldsymbol{z_i}$ . We assumed $z_{it}$ arises from an K-state homogeneous continuous time Markov process taking values on the finite integer set $\{1,2, \cdots, K\}$ where $K$ is (un)known. This study primarily focused on HMMs arising from the first-order Markov assumption, where the transition matrix is denoted as $\boldsymbol{P}$. The initial probability (row) vector $\boldsymbol{\pi}$ represents the initial state distribution of the hidden process for each individual. We assumed that initial state probabilities, indicated by $\boldsymbol{\pi}$, remains constant over time. In other words, only the starting distribution is fixed, while the evolution over time is governed by the transition model. The elements of $\boldsymbol{P}$ are $p_{jk} = P\big(z_{i,t} = k|\, z_{i,t-1} = j\big)$ for all $t > 1$, and $\pi_k = P\big(z_{i,1} = k\big)$; where $j, k = 1,2, \cdots, K$. Suppose that a HMM generates the joint longitudinal response such that $\boldsymbol{y_{it}}$ and $\boldsymbol{y_{it'}}$ were independent given the current hidden state $z_{it}$ $and$ $z_{it'}$ , respectively, for two different time points $t \neq t'$. Further, we incorporate subject-specific random effects, denoted as $\boldsymbol{u_i}$, to account for inter-individual variation. Let $\boldsymbol{u}$ to be all values of $\boldsymbol{u_i}$ for all $i$. These random effects are assumed to be independent and identically distributed (i.i.d.) according to a common distribution with zero mean and covariance matrix $\Sigma_u$.

Conventionally, using standard arguments based on the Chapman-Kolmogorov and Kolmogorov equations, for individuals $i$ the transition probability from state $j$ to state $k$ in the time interval of length $\Delta_t$ between observation times $t$ and $t - 1$ takes the form

$$p_{jk}(\Delta_t) = P\big(z_{i,t} = k|\, z_{i,t-1} = j, \Delta_t\big) = \exp(\Delta_t \boldsymbol{Q}_i)_{jk} \quad , \tag{1}$$

where $\boldsymbol{Q_i} = (q_{i,jk})$ for $1 \leq k, j \leq K$ is the infinitesimal generator for the continuous time Markov process $\{z_{it}\}$ which represents the instantaneous risk of moving from state $j$ to state $k$ at time-interval $\Delta_t$ and $\sum_{j \neq k} q_{jk} = -q_{jj} > 0$ for $1 \leq j \leq K$, and exp(**A**) is the matrix exponential of matrix **A.** To incorporate baseline subject-level covariate $\boldsymbol{V} \in \mathbb{R}^P$ into this generator $Q$, we defined the log-rate parameters via a linear predictor for $1 \leq k \neq j \leq K$, $\log q_{jk} = \boldsymbol{V}^T \xi_{jk}$ , where $\xi_{jk}$ is the coefficient vector for $q_{jk}$, assuming that the rate in any interval is determined by the values of covariates at the start of the interval. The initial state distribution was $\pi_k = P(z_1 = k)$ for $k = 1,2, \cdots, K$.

#### 4.4.2.2 *Linear Mixed Effects Observation Model*

The observation process assumes that the observed data are drawn from an exponential family conditional on the hidden state. Here, we focused only on adding random effects to the conditional observation model and assume that random effects do not appear for the hidden process transition matrix. We also assume that the hidden process is homogenous with transition probabilities defined in equation 1 and initial probabilities $\pi_k$ that are common to all subjects.

Using the theory of generalized linear mixed models (GLMMs) [32] and letting $\boldsymbol{\theta}$ be the vector of all model parameters, we assume that, conditional on the random effects, $\boldsymbol{u}$, and the hidden states $z_{it} = k$, $y_{irt}$ are independent and follow a distribution in the exponential family for each $R$

$$p(y_{irt}|z_{it} = k, \boldsymbol{u}, \boldsymbol{\theta}) = \exp\{(y_{irt}\, \eta_{irtk} - c(\eta_{irtk}))/a(\phi) + d(y_{irt}, \phi)\}, \quad (2)$$

where the link function is $\eta_{irtk} = \boldsymbol{e'_{irtk}\alpha_r} + \boldsymbol{x'_{irt}\beta_{rk}} + \boldsymbol{w'_{irt}u_i}$ . Here, $\boldsymbol{\alpha_r}$ correspond to the fixed effect intercept for each hidden state, as stated through a $K$-dimensional hidden vector, $\boldsymbol{e_{irtk}}$ **.** We follow [23] such that for hidden state $k$ , $\boldsymbol{e'_{irtk}\alpha_r}$ defines the cumulative sum of the first $k$ elements of $\boldsymbol{\alpha_r}$. Furthermore, $\boldsymbol{\beta_{rk}}$ is the fixed effect when $z_{it} = k$, $\boldsymbol{x'_{irt}}$ are vector of covariates for individual $i$ at time $t$ , and $\boldsymbol{w_{irt}}$ is the row vector of model matrix for the random effects for $r$th longitudinal response for individual $i$ at time $t$ in state $k$ . We denote the distribution of random effects by $f_{\boldsymbol{u_i}}(\boldsymbol{u_i}|\boldsymbol{\Sigma})$ and assume that random

effects are independent of the hidden states (see Figure 4.1).

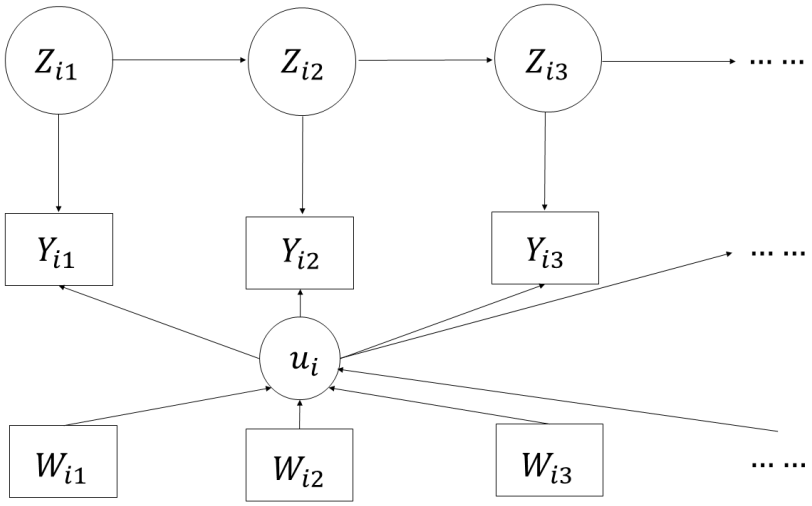



**Figure 4.1** A typical representation of mixed effects hidden Markov model structure for subject *i*.

We denoted $\boldsymbol{S_i}$ as an indicator variable that comprises all values of $\boldsymbol{S_{it}}$ for subject $i$ and $\boldsymbol{S}$ to all values of $\boldsymbol{S_i}$ for all *i.* Here, $\boldsymbol{S_{it}} = \left(S_{it,1}, \cdots, S_{it,K}\right)'$ is an indicator random vector with $S_{it,K} = 1$ if $z_{it} = k$ and 0 otherwise.

When the response may have missing values, we considered a state-dependent missingness mechanism where missingness depends on the hidden state but not on the value of the missing data. We defined $\boldsymbol{y_{obs}}$ be the set of observed values and $\boldsymbol{y_{miss}}$ be the set of missing values. We defined the multivariate missingness indicator vector $\boldsymbol{M_{it}} = \left[M_{i1t}, M_{i2t,} \cdots, M_{iRt}\right]'$, Missingness was encoded using component-specific indicators $M_{iRt}$, where $M_{iRt} = 1$ if the $R$ th response component was missing and belongs to $\boldsymbol{y_{miss}}$ and 0 if response component belongs to $\boldsymbol{y_{obs}}$. The resulting vector $\boldsymbol{M}_{it}$ allows for partially observed response vectors and facilitates state-dependent missingness modeling. The model is thus parameterized by $\boldsymbol{\Theta} = \left\{\xi_{jk}, 1 \leq j, k \leq K, \boldsymbol{\pi}, \boldsymbol{B}, \boldsymbol{\Sigma_u}, \boldsymbol{\varphi}\right\}$, where $\boldsymbol{B}$ represents parameters associated with the canonical link of the observed process in Equation (2), $\boldsymbol{\Sigma}$ related to random effects and $\boldsymbol{\varphi}$ related to missingness parameter**.**

#### 4.4.2.3 *The Likelihood*

If $\{z_s\}$ is observed continuously over the time interval $[0, t]$, the likelihood contribution from an individual subject $i$ is defined [39] as

$$\prod_{l=1}^{K}\prod_{m\neq l} q_{l,m}^{N_{l,m}(\Delta_t)} \exp\{-q_{l,m}D_l(\Delta_t)\},$$

where, $N_{l,m}(\Delta_t)$ is the number of transitions from state $l$ to state $m$ in the time interval $\Delta_t$and $D_l(\Delta_t) = \int_0^{\Delta_t} \mathbb{I}\,(z_s = l)ds$ is the total time that the process has spent in state $l$ in $\Delta_t$. The quantities $N_{l,m}(\Delta_t)$ and $D_l(\Delta_t)$ can be computed given the realization of the latent process on $\Delta_t$. Therefore, the complete (observed) data likelihood for proposed CT-mMixHMM can be defined as follows:

$$\mathcal{L}\,(\boldsymbol{\Theta}; \mathbf{y}) = p(\boldsymbol{y}|\boldsymbol{\Theta}) = \int_{\boldsymbol{u}} \sum_{\boldsymbol{z}} p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\Theta})\, p(\boldsymbol{z}|\boldsymbol{u}; \boldsymbol{\Theta})\, p\,(\boldsymbol{u}; \boldsymbol{\Theta})\; d\boldsymbol{u}$$

$$= \int_{\boldsymbol{u}} \sum_{\boldsymbol{z}} \left\{ \prod_{r=1}^{R}\prod_{i=1}^{N}\prod_{t=1}^{n_i} p(y_{irt}|z_{it}, \boldsymbol{u}, \boldsymbol{\Theta}) \right\}$$

$$\times \left\{ \prod_{i=1}^{N} \pi_{z_{i1}} \left[ \prod_{t=2}^{n_i}\prod_{l=1}^{K}\prod_{m\neq l} q_{i,l,m}^{N_{i,l,m}(\Delta_{i,t})} \exp\{-q_{i,l,m}D_{i,l}(\Delta_{i,t})\} \right] \right\} p(\boldsymbol{u}; \boldsymbol{\Theta})\, d\boldsymbol{u}$$

$$= \int_{\boldsymbol{u}} \prod_{r=1}^{R}\prod_{i=1}^{N} \left\{ \sum_{\boldsymbol{z_i}} \pi_{z_{i1}}\, p(y_{ir1}|z_{i1}, \boldsymbol{u}, \boldsymbol{\Theta}) \right.$$

$$\left. \times \prod_{t=2}^{n_i}\prod_{l=1}^{K}\prod_{m\neq l} q_{i,l,m}^{N_{i,l,m}(\Delta_{i,t})} \exp\{-q_{i,l,m}D_{i,l}(\Delta_{i,t})\}\; p(y_{irt}|z_{it}, \boldsymbol{u}, \boldsymbol{\Theta}) \right\} p(\boldsymbol{u_i}; \boldsymbol{\Theta})\, d\boldsymbol{u_i}. \qquad (3)$$

The complete (observed) data log-likelihood is then given by

$$\ell(\boldsymbol{\Theta}) = \log p(\boldsymbol{y}|\boldsymbol{\Theta}) = \log p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\Theta}) + \log p(\boldsymbol{z}|\boldsymbol{u}; \boldsymbol{\Theta}) + \log p\,(\boldsymbol{u}; \boldsymbol{\Theta})$$

$$= \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{t=1}^{n_i} \log p(y_{irt}|z_{it}, \boldsymbol{u}, \boldsymbol{\Theta}) + \sum_{i=1}^{N} \log \pi_{z_{it}} + \sum_{i=1}^{N}\sum_{t=2}^{n_i} d_{i,t} + \log p(\boldsymbol{u}; \boldsymbol{\Theta}).$$

Now, the log-likelihood written in terms of the latent state indicator random vectors $\{\boldsymbol{S_k}\}_{k=1}^{K}$ is,

$$\ell(\boldsymbol{\Theta}) = \log p(\boldsymbol{y}|\boldsymbol{\Theta})$$

$$= \sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{t=1}^{n_i}\sum_{k=1}^{K} S_{i,t,k} \log p(y_{irt}|z_{it}, \boldsymbol{u}, \boldsymbol{\Theta}) + \sum_{i=1}^{N}\sum_{k=1}^{K} S_{i,1,k} \log \pi_{z_{i1}}$$

$$+ \sum_{i=1}^{N}\sum_{t=2}^{n_i}\sum_{j=1}^{K}\sum_{k=1}^{K} S_{i,t-1,j}\, S_{i,t,k}\, d_{i,t}^{j,k} + \log p(\boldsymbol{u}; \boldsymbol{\Theta}), \qquad (4)$$

where,

$$d_{i,t}{}^{j,k} = \sum_{l=1}^{K}\sum_{m\neq l} \{N_{i,l,m}^{j,k}(\Delta_{i,t})\boldsymbol{V_i}^T \xi_{lm} - \exp(\boldsymbol{V_i}^T \xi_{lm})\, D_{i,l}^{j,k}(\Delta_{i,t})\},$$

records the probability of transition from state $j$ to state $k$ in the interval $\Delta_{i,t}$ and $N_{i,l,m}^{j,k}(\Delta_{i,t})$ and $D_{i,l}^{j,k}(\Delta_{i,t})$ are the number of jumps from state $l$ to state $m$, and time spent in state $l$ for a Markov chain initiated at state $j$ at time $t-1$ and ending in state $k$ at time $t$ for each individual $i$.

***Likelihood with state-dependent missingness***

By incorporating the state-dependent missingness, the likelihood defined in equation (3) can be written as

$$\mathcal{L}(\boldsymbol{\Theta}; \mathbf{y}, \boldsymbol{M}) = p(\boldsymbol{y}, \boldsymbol{M}|\boldsymbol{\Theta}) = \left[\sum_{z} p(\boldsymbol{M}|\boldsymbol{z}; \boldsymbol{\varphi}) \times \int_{\boldsymbol{u}} p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\Theta})\, p(\boldsymbol{z}|\boldsymbol{u}; \boldsymbol{\Theta})\, p(\boldsymbol{u}; \boldsymbol{\Theta})\, d\boldsymbol{u}\right]$$

And the log-likelihood defined in equation (4) now can be written as

$$\ell(\boldsymbol{\Theta}) = \log p(\boldsymbol{y}, \boldsymbol{M}|\boldsymbol{\Theta}) = \log\left[\sum_{z} p(\boldsymbol{M}|\boldsymbol{z}; \boldsymbol{\varphi}) \times \int_{\boldsymbol{u}} p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\Theta})\, p(\boldsymbol{z}|\boldsymbol{u}; \boldsymbol{\Theta})\, p(\boldsymbol{u}; \boldsymbol{\Theta})\, d\boldsymbol{u}\right].$$

Let $\boldsymbol{z}(\boldsymbol{M})$ denote the set of latent state trajectories that are compatible with the observed missingness

pattern, in the sense that they assign positive probability to the observed missingness under the assumed state-dependent missingness mechanism. In this research, under deterministic death-induced missingness, we restricted admissible state paths to those for which $z_{it} = 6$ if and only if $M_{irt} = 1$.

Since, missingness was deterministic in our case given state $z_{it} = 6$ for *i*th individual at time *t* with *r* longitudinal response:

$$p(\boldsymbol{M} \mid \boldsymbol{z}; \boldsymbol{\phi}) = \prod_{t} \prod \mathbb{I}(M_{irt} = 1 \iff z_{it} = 6),$$

then the final observed data log-likelihood becomes,

$$\ell(\boldsymbol{\Theta}; \boldsymbol{y}, \boldsymbol{M}) = \log p(\boldsymbol{y} \mid \boldsymbol{\Theta}) = \log \left[ \sum_{\boldsymbol{z} \in \boldsymbol{Z}(\boldsymbol{M})} \int p(\, \boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\Theta}\,)\, p(\, \boldsymbol{z} \mid \boldsymbol{u}; \boldsymbol{\Theta}\,)\, p(\boldsymbol{u}; \boldsymbol{\Theta})\, d\boldsymbol{u} \right] (4.a)$$

where $\boldsymbol{Z}(\boldsymbol{M})$is the set of state paths compatible with the observed missingness pattern.

The observed data log-likelihood is the log of the marginal likelihood, obtained by summing over latent state trajectories and integrating over subject-specific random effects. Because missingness enters through a state-dependent mechanism, the likelihood takes a log–sum–integral form that is analytically intractable in most hidden-state models. Consequently, estimation relies on optimization-based approaches, such as variational inference methods for continuous-time mixed hidden Markov models.

#### 4.4.2.4 *Proposed Sequential Variational Algorithm for CT-mMixHMMs*

In a sequential VB framework, such as that proposed by [33], the observations $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ are incorporated sequentially so that at iteration $i$ , one targets an approximation $q_i(\boldsymbol{\Theta}) = q(\boldsymbol{\Theta}; \boldsymbol{\lambda_i}) = N\,(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}\,)$ , where $\boldsymbol{\Sigma_i}$ is a full covariance matrix, and seeks closed-form updates for $\boldsymbol{\lambda_i} = \{\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}\}$ that is closest in a KL sense to the "pseudo-posterior" $p(\boldsymbol{y_i}|\boldsymbol{\Theta}\,)\, q_{i-1}(\boldsymbol{\Theta})/\mathcal{H}_i$ , where

$$\mathcal{H}_i = \int p(\boldsymbol{y_i}|\boldsymbol{\Theta}\,)\, q_{i-1}(\boldsymbol{\Theta})\, d\boldsymbol{\Theta}.$$

Let us denote, $\boldsymbol{y}_{1:i} = (\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_i^T)^T$ a collection of observations from individual 1 to N , $i = 1, \cdots, N$. By assumption of conditional independence between observations $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i$ given the parameters $\boldsymbol{\Theta}$, the KL divergence between the variational distribution $q_i(\boldsymbol{\Theta})$ and the posterior distribution $p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i})$ can be expressed as

$$KL\left(q_i(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i})\right) = \int q_i(\boldsymbol{\Theta}) \log \frac{q_i(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i})}\, d\boldsymbol{\Theta}$$

$$= \mathrm{E}_{q_i}(\log q_i(\boldsymbol{\Theta}) - \log p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i-1}) - \log p(\boldsymbol{y_i}|\boldsymbol{\Theta})) + \log p(\boldsymbol{y_{1:i}}) - \log p(\boldsymbol{y_{1:i-1}}).$$

The posterior distribution after incorporating the first $i-1$ groups of observations, $p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i-1})$ is approximated by the variational distribution $q_{i-1}(\boldsymbol{\Theta})$ to give

$$KL\left(q_i(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\boldsymbol{y}_{1:i})\right)$$

$$\approx \mathrm{E}_{q_i}(\log q_i(\boldsymbol{\Theta}) - \log q_{i-1}(\boldsymbol{\Theta}) - \log p(\boldsymbol{y_i}|\boldsymbol{\Theta}))$$

$$+ \log p(\boldsymbol{y_{1:i}}) - \log p(\boldsymbol{y_{1:i-1}})\,. \tag{5}$$

As the last two terms in the right-hand side of the equation (5) do not depend on $\boldsymbol{\Theta}$, the optimization problem is equivalent to finding

$$\arg min_{\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}}\ \mathrm{E}_{q_i}(\log q_i(\boldsymbol{\Theta}) - \log q_{i-1}(\boldsymbol{\Theta}) - \log p(\boldsymbol{y_i}|\boldsymbol{\Theta}))\,. \tag{6}$$

Differentiating the expectations of (6) with respect to $\boldsymbol{\mu_i}$ and $\boldsymbol{\Sigma_i}$ , setting the derivatives to zero, and rearranging the resulting equations, yield the following recursive updates for the variational mean $\boldsymbol{\mu_i}$ and precision matrix $\boldsymbol{\Sigma_i}^{-1}$,

$$\boldsymbol{\mu_i} = \boldsymbol{\mu_{i-1}} + \boldsymbol{\Sigma_{i-1}}\, E_{q_i}\left(\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}|\boldsymbol{\Theta})\right), \text{and} \tag{7}$$

$$\boldsymbol{\Sigma_i}^{-1} = \boldsymbol{\Sigma_{i-1}^{-1}} - E_{q_i}\left(\nabla^2_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}|\boldsymbol{\Theta})\right)\,. \tag{8}$$

These updates are implicit as they require the evaluation of expectations with respect to $q_i(\boldsymbol{\Theta})$. Under the assumption that $q_i(\boldsymbol{\Theta})$ is close to $q_{i-1}(\boldsymbol{\Theta})$, [34] proposed replacing $q_i(\boldsymbol{\Theta})$ with $q_{i-1}(\boldsymbol{\Theta})$ in

equations (7) and (8) and replacing $\boldsymbol{\Sigma_{i-1}}$ with $\boldsymbol{\Sigma_i}$ on right hand side of (6) to yield an explicit scheme called recursive variational Gaussian algorithm (R-VGA).

**Algorithm 2 Damped R-VGADM**

**Input:** Observations $\{\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_N}\ ;\ \boldsymbol{M_1}, \boldsymbol{M_2}, \cdots, \boldsymbol{M_N}\ ;\ \boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_{,N}};\ \boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w_N}\ ; \boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_N}\}$, initial variational parameter $\boldsymbol{\mu_0}$ and $\boldsymbol{\Sigma_0}$ for $\boldsymbol{\Theta}$ **,** number of damping parameters $n_{\text{damp}}$ , number of inner damping steps $L$ .

**Output:** Variational parameters $\boldsymbol{\mu_i}$ and $\boldsymbol{\Sigma_i}$ for $i = 1,2, \cdots, N$.

**Initialization:** Set $q_0(\boldsymbol{\Theta}) = N\ (\boldsymbol{\mu_0}\ , \boldsymbol{\Sigma_0}\ )$

**for** $i = 1,2, \cdots, N$ **do**

**if** $i\ \leq\ n_{\text{damp}}$ **then**

set $a = \ {}^{1}/_{L}\ ,\ \boldsymbol{\mu_{i,0}} = \ \boldsymbol{\mu_{i-1}}\ ;\ \boldsymbol{\Sigma_{i,0}} = \ \boldsymbol{\Sigma_{i-1}}$

**for** $l = 1,2, \cdots, L$ **do**

set $q_{i,l-1}(\boldsymbol{\Theta}) = N\ (\boldsymbol{\mu_{i,l-1}}\ , \boldsymbol{\Sigma_{i,l-1}}\ )$

$\boldsymbol{\mu_{i,l}} = \ \boldsymbol{\mu_{i,l-1}} + a\ \boldsymbol{\Sigma_{i,l}}\ \mathbf{E}_{q_{i,l-1}}\left(\nabla_{\boldsymbol{\Theta}} \log \widehat{p(\boldsymbol{y_\iota}}\ |\boldsymbol{\Theta})\ \right)$

$\boldsymbol{\Sigma_{i,l}}^{-1} = \ \boldsymbol{\Sigma_{i,l-1}}^{-1} - a\ \mathbf{E}_{q_{i,l-1}}\left(\nabla^2_{\boldsymbol{\Theta}} \log \widehat{p(\boldsymbol{y_\iota}}\ |\boldsymbol{\Theta})\ \right)$

**end for**

set $\boldsymbol{\mu_i} = \ \boldsymbol{\mu_{i,l}}\ ; \boldsymbol{\Sigma_i} = \ \boldsymbol{\Sigma_{i,l}}\ ;\ q_i(\boldsymbol{\Theta}) = N\ (\boldsymbol{\mu_i}\ , \boldsymbol{\Sigma_i}\ )$

else

$\boldsymbol{\mu_i} = \ \boldsymbol{\mu_{i-1}} + \ \boldsymbol{\Sigma_i}\ \mathbf{E}_{q_{i-1}}\left(\nabla_{\boldsymbol{\Theta}} \log \widehat{p(\boldsymbol{y_\iota}}\ |\boldsymbol{\Theta})\ \right)$

$\boldsymbol{\Sigma_i}^{-1} = \ \boldsymbol{\Sigma_{i-1}}^{-1} - \ \mathbf{E}_{q_{i-1}}\left(\nabla^2_{\boldsymbol{\Theta}} \log \widehat{p(\boldsymbol{y_\iota}}\ |\boldsymbol{\Theta})\ \right)$

**end if**

**end for**

Now, for CT-MixHMMs, the likelihood function defined in equations (3), (4) and (4.a), the "pseudo-posterior" for SVB takes the form, $p(\boldsymbol{y_i}|\boldsymbol{z_i}, \boldsymbol{u_i}, \boldsymbol{\Theta})\ q_{i-1}(\boldsymbol{z_i}, \boldsymbol{\Theta})/\mathcal{H}_i$ , where,

$$\mathcal{H}_i = \int p(\boldsymbol{y_i}|\boldsymbol{z_i}, \boldsymbol{u_i}, \boldsymbol{\Theta})\ q_{i-1}(\boldsymbol{z_i}, \boldsymbol{\Theta})\ d\boldsymbol{\Theta}\ . \tag{9}$$

The above Recursive Variational Gaussian Approximation algorithm for Dependent Mixtures (R-VGADM) has been specifically developed for parameter estimation in dependent mixture models, such as for continuous-time multivariate mixed hidden Markov models using the algorithmic framework presented in [33]. To tackle the instability of the first few observations, we considered the damping approach of [33] to stabilize our algorithm during the initial few steps.

The partial-log-likelihood function, $\log p(\boldsymbol{y_i}\ |\boldsymbol{\Theta})$ used in equation (6) has the following components.

*Components of the log-likelihood:*

- $\log p(\boldsymbol{z_1}/\boldsymbol{\pi})$**:** The log-probability of the initial state, using the initial probability vector $\boldsymbol{\pi}$.
- $\sum_{\boldsymbol{t=2}}^{\boldsymbol{T}} \log\ p(\boldsymbol{z_t}/\boldsymbol{z_{t-1}}, \boldsymbol{P})$: The sum of the log-transition probabilities, where $\boldsymbol{P}$ depends on the transition intensities **Q** for continuous time hidden Markov model.
- $\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \log\ p(\boldsymbol{y_t}/\boldsymbol{z_t}, \boldsymbol{u}, \boldsymbol{\phi})$: The sum of the log-emission probabilities, where the distribution of the observation $\boldsymbol{y_t}$ depends on both the current hidden state $\boldsymbol{z_t}$ and the random effect $\boldsymbol{u}$**.** The parameters $\boldsymbol{\phi}$ define the base emission characteristics such as mean and variances.
- $\log p(\boldsymbol{u}/\boldsymbol{\Sigma})$: The log-prior distribution of the random effects $\boldsymbol{u}$ **.** In our model, $\boldsymbol{u}$ considered as a Gaussian random variable with mean 0 and variance $\boldsymbol{\Sigma_u}$ **.**

The R-VGADM updates in (7) and (8) require the gradient and Hessian of the "partial" log-likelihood for the *i*th observation. The detail derivation of the approximation of the gradient with Fisher's identity and Hessian with Loui's identify for $\log p(\boldsymbol{y_i}\ |\boldsymbol{\Theta})$ are given in supplementary materials (S1: Appendix A).

#### 4.4.2.5 *Prior Specification*

Specifying a prior for $\boldsymbol{\Theta}$ depends on the specific circumstances and the target of inference. It is often convenient for sampling to put independent conjugate priors on each component of $\boldsymbol{\Theta}$**.** In general, it is convenient to factor $[\boldsymbol{\Theta}]$ as

$$[\boldsymbol{\Theta}, \mathbf{u}] = [\boldsymbol{\beta}, \boldsymbol{\alpha}] \times [\boldsymbol{\xi}] \times [\boldsymbol{\pi}] \times [\boldsymbol{u}|\boldsymbol{\Sigma}_{\boldsymbol{u}}] \times [\boldsymbol{\Sigma}_{\boldsymbol{u}}] \times [\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}].$$

In most cases, we choose conjugate priors for each component of $\boldsymbol{\Theta}$**,** as this makes sampling from the required full conditional distribution generally straightforward. Here, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is precision parameter associated with the observation process. In the case of hyperparameters, we will use standard "default" choices or ones that are relatively uninformative because of having tractable density functions. Default hyperparameters with tractable density functions are chosen because they simplify mathematical and computational tasks. This can simplify the process of model fitting and make it easier to understand the behavior of the model [35]. However, when sample size per subject is small, an informative prior may be adopted. In such cases, we will use normal priors with small variances on the original or log scale for the regression parameters. Priors are discussed in detail in supplementary materials S1: Appendix A.

#### 4.4.2.6 *Simulation Setup*

We simulated data that reflects the characteristics of ILD with a multilevel, nested structure and temporal dependencies, using a continuous time multivariate hidden Markov model (CT-mHMM). In this context, specifying a data-generating process that matches the analytic model is a standard and well-established first step in simulation research [47, 48]. This approach is widely used to evaluate internal validity - specifically, whether an estimation procedure can recover known parameters under correct model specification - before examining performance under misspecification. This logic is well documented in the broader simulation and verification literature [46, 47]. Furthermore, the simulated parameter values were chosen to reflect the motivating MFUS datasets, in which the modes of the hidden states are clearly

differentiated, and the outcome variables demonstrate considerable heterogeneity. This methodology aligns with established practices documented in prior research on the multivariate mixed effects hidden Markov model [23].

***1.*** *Data Generation*

*Step 1: Define Simulation Parameters*

- Number of individuals $N$: Sample sizes (e.g., $N$=200,500,1000) were selected to reflect realistic population sizes typically found in empirical intensive longitudinal and longitudinal cohort studies.
- Time points per individual or number of observations per individual, $n_i$: The number of observations per individual (e.g., $n_i \in \{20,30,40,50\}$).
- Number of latent states $K$: The initial number of hidden states in the CT-mMixHMM (e.g., $K$=5, 10, 12, 15) was used as a tuning and robustness parameter for estimation, rather than as a substantive epidemiological assumption.
- Number of dependent variables: two dimensional ($R$=2) longitudinal responses.
- State dependent missingness variable $\boldsymbol{M}$: The missing proportion (e.g., $p$=0.2,0.3,0.4). The missingness proportion is a marginal consequence of entering the Death state.
- State transition rates $\boldsymbol{Q}$: The transition between states is governed by a rate matrix $\boldsymbol{Q}$ defined in section 4.4.2.1 and given in section 4.5.1.
- Initial state distribution $\boldsymbol{\pi}$: A vector defining the initial probability distribution over the hidden states defined in section 4.4.2.1 and given in section 4.5.1.
- Emission model with single covariate: For each latent state, the observations are drawn from a likelihood distribution with observed data defined in equation (4).
- Modified emission model: For each latent state, the observations are drawn from a likelihood distribution modified by the dropout rate defined in equation (4.a).

*Step 2: Generate Latent State Sequences*

We simulated the hidden states using the rate matrix **Q** and $\boldsymbol{\pi}$ defined in section 4.4.2.1 for each $i$ and $t$. At the same time, for each $i$ and $t$ , we generated the random vector $\boldsymbol{S_{i,t}}$ from the multinomial distribution with parameters $a_{i,t} = (a_{i,t,1}, \cdots, a_{i,t,K})$. Moreover, we generated $\boldsymbol{\pi}$ from a Dirichlet distribution with random vector $\boldsymbol{S_{i,t}}$ as defined in prior specification of supplementary file (see supplementary material (S1): Appendix A)**.**

*Step 3: Generate Observed Data*

For each latent state, we simulated observations using a state-dependent observation model. For simplicity, we assumed multivariate Gaussian observation model for our simulation study. For each individual, the observation times were randomly spaced to reflect the continuous nature of the data.

*Step 4: Simulation Scenarios*

We investigated multiple simulation scenarios to assess the robustness of the algorithm:

- *Scenario 1*: Varied the number of individuals, time points and proportion of missingness to evaluate the scalability of the proposed algorithm.
- *Scenario 2*: Changed the complexity of the hierarchical structure to assess how well the proposed algorithm of the model captures individual level variation [33].

***2.*** *Model Setup: Continuous time multivariate mixed hidden Markov model*

Set up a multivariate mixed hidden Markov model where:

- Each individual had their own set of latent states but shares a common structure for the transition and multivariate emission distribution with baseline covariate.
- Random effects were included to represent individual differences in the transition probabilities between hidden states or in the emission distributions (random intercept).

#### 4.4.2.7 *Model Performance Measures*

We evaluated the performance of our proposed damped R-VGADM algorithm compared to standard MCMC in terms of speed gains and accuracy. Specifically, we examined how accurately the R-VGADM estimates reproduced the true parameters (the values used to generate the data) and compared these accuracy metrics to those obtained with MCMC. To assess accuracy, we considered both the bias of the posterior means for the initial state distribution parameters, the transition rates and the associated fixed and random effect parameters, as well as the coverage of their 95% central credible intervals (the central portion of the posterior distribution that contains 95% of the values). This evaluation was based on 250 fitted models per simulation condition. The actual coverage of their 95% credible intervals is defined as follows:

$$\text{Coverage Rate} = \frac{1}{b}\sum_{b=1}^{B} I(\theta_{true} \in CCI_b) ,$$

where B is the number of simulations and $I(\theta_{true} \in CCI_b)$ is an indicator function, where $I = 1$ if the true parameter $\theta_{true}$ is within the central credible interval and $I = 0$ otherwise.

We also considered deviance information criteria (DIC) proposed in [38] for HMM to select the number of hidden states when our algorithm was initialized with large number of hidden states. DIC was selected primarily for its computational efficiency and practical feasibility in the proposed high-dimensional continuous-time mixed-effects hidden Markov framework. Unlike alternatives such as Bayesian cross-validation or marginal likelihood estimation, DIC can be computed directly from posterior approximations without repeated model fitting or explicit specification of the effective number of model parameters.

$$DIC = -2\log p\left(\boldsymbol{y} \mid \widetilde{\boldsymbol{\Theta}}\right) + 2p_D ,$$

where, $\widetilde{\boldsymbol{\Theta}}$ is the posterior estimates of the parameters of interest and $p_D$ is the model complexity measured

in equation 4 (page 230) of [38]. To assess the model fit, we used posterior predictive checks (PPC) [36, 37]. For MCMC, we considered trace plots and auto-correlation plots to evaluate stationarity and mixing. These evaluations were also applied to the CT-mMixHMM on the empirical dataset.

## 4.5 Results

### 4.5.1 *Simulation study results*

In this section, we present the results of simulation study of the CT-mMixHMM. The simulated dataset was designed to reflect the blood pressure transition patterns among older population, whereas the empirical analysis applies the proposed model to real MFUS data to examine how age, as both fixed and random effects covariates, influences progression between hypertensive states. Both the simulation study and real data application were implemented in the R software version 4.5.2. In the simulation study, the MCMC used a modified version of the function *mHMM_cont* from the *mHMMbayes* package [40].

For simulation scenario 1, the sequential variational Bayes algorithm, R-VGADM, was initialized with a large number of hidden states to correctly identify the true number of hidden states in the simulated datasets. We used MFUS-derived empirical estimates for the emission means and variances and for the transition intensities matrix $\boldsymbol{Q}$ with one baseline covariate as starting values to mitigate the label switching. We varied the number of individuals $N$= 200,500,1000, the number of time points $T$ =20,30,40,50, and the marginal proportion of missingness about 2%, 3%, and 4%. Missingness was generated as a deterministic, state-dependent process. More specifically, observations were fully observed in all non-terminal states and became missing with probability of one upon entry into the full absorbing state. The results obtained from these models were sufficient and did not exhibit label switching across different combinations of $N$, $T$ and missing proportions.

Here, we presented results of two simulated datasets, comprising of 15,000 ($M$=500, $T$= 30), 30,000

($M$=1000, $T$= 30) observations, respectively, from a six-state CT-mMixHMM where we considered one baseline covariate $V_1 \sim N(0,1)$ and the following associated coefficients matrices: $\boldsymbol{\xi_0}$ was used for intercept and $\boldsymbol{\xi_1}$was used for intercept for generating constrained transition intensity matrix Q , which are given below:

$$\boldsymbol{\xi_0} = \begin{bmatrix} 0.000 & -1.10 & -2.00 & -0.12 & -0.13 & -0.02 \\ -0.10 & 0.00 & -1.19 & -1.32 & 0.100 & -0.47 \\ -0.16 & 0.110 & 0.00 & -1.01 & 0.080 & -0.12 \\ -3.10 & 0.020 & 0.120 & 0.00 & -0.84 & 0.125 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -3.11 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix},$$

$$\boldsymbol{\xi_1} = \begin{bmatrix} 0.000 & 1.000 & 2.000 & 0.500 & 1.250 & 1.225 \\ 1.000 & 0.000 & 1.000 & 0.200 & 1.000 & 1.250 \\ 1.111 & 1.111 & 0.000 & 0.500 & 2.000 & 1.000 \\ 1.000 & 2.000 & 1.200 & 0.000 & 2.000 & 1.250 \\ 0.000 & 0.000 & 0.000 & 0.000 & 2.500 & 2.500 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix},$$

and with the initial state probability,

$$\boldsymbol{\pi} = [0.5, 0.3, 0.1, 0.05, 0.05, 0].$$

We also considered one-time varying covariate $X_1 \sim N(0, 0.5)$ with coefficient vector $\boldsymbol{B} = [\boldsymbol{B_1}\ \boldsymbol{B_2}]'$, where $\boldsymbol{B_1} = \begin{pmatrix} 111.2 & 122.4 & 131.7 & 141.2 & 139.2 & NA \\ 0.12 & 0.24 & 0.32 & 0.43 & 0.60 & NA \end{pmatrix}$, and

$$\boldsymbol{B_2} = \begin{pmatrix} 71.3 & 71.7 & 82.3 & 86.2 & 81.3 & NA \\ 0.02 & 0.12 & 0.25 & 0.24 & 0.35 & NA \end{pmatrix}.$$

The random effects were i.i.d and had a covariance matrix $\boldsymbol{\Sigma_u} = \begin{bmatrix} 15.21 & 0.5 \\ 0.5 & 10.36 \end{bmatrix}$ assuming conditional independence and the inverse of residual error variance $\boldsymbol{\Sigma_{\varepsilon,R}} = 1/\log \sigma^2_{\varepsilon R} = 10$,

Here, state 5 was considered as partially absorbing state (to align with clinical meaningfulness condition of blood pressure medication of MFUS participants) which can only move the next state (state 6) and state 6 was considered as full absorbing state means once an individual visits this state will not return anymore from that state. Note that we modeled each of the two outcome variables as independent

normal distributions with mean $\eta_{irtk}$.

**Table 4.1** Population values of the group level means for each outcome variable for simulated data

| Variables | States | Mean (SD) |
|---|---|---|
| **Dependent Variable 1** | State 1 | 111.2 (3.11) |
| | State 2 | 122.4 (2.10) |
| | State 3 | 131.7 (5.12) |
| | State 4 | 141.2 (11.70) |
| | State 5 | 139.2 (15.2) |
| | State 6 | NA (NA) |
| **Dependent Variable 2** | State 1 | 71.3 (3.57) |
| | State 2 | 71.7 (2.56) |
| | State 3 | 82.3 (2.89) |
| | State 4 | 86.2 (8.06) |
| | State 5 | 81.2 (6.24) |
| | State 6 | NA (NA) |

That is, the covariance between any two outcome variables is assumed to be zero after conditioning on the hidden state. However, marginal dependence between outcomes was captured through random effects covariance matrix $\boldsymbol{\Sigma_u}$. For the simulation study, we chose $\boldsymbol{w_{i1t}} = [1 \quad 0]$ and $\boldsymbol{w_{i2t}} = [0 \quad 1]$. Empirical estimates of the key parameters (means and SD) for these states are presented in Table 4.2.

The state-dependent missingness with $P(D_t = 1|Z_t = 6) = 1$, which indicated missingness only occur due to state 6 (a non-ignorable missingness which could happen due to an event such as death). We

explored the results of a sequential variational Bayes analysis with several different numbers of initial states to examine how results varied by individual count and time points.

#### 4.5.1.1 *Convergence and model selection*

Table 4.2 ($N$=500, $T$ = 30) and table 4.3 ($N$=100, $T$ = 30) report: (i) the initial number of states used to state the CT-mMixHMM; (ii)the final number of states selected automatically by the sequential variational Bayes analysis; (iii) the estimated sequential variational posterior means and standard errors for the normal emission densities;(iv) the variational estimate of the deviance information criterion (DIC), used as a model selection criteria for VB-based HMM [38]. The results show that the algorithm correctly identifies six-state solutions for each time for each dependent variable.

For the $N$=500, $T$=30, in total 15, 000 observation datasets (with 2% missingness), we were successfully able to recover a six-state solution, with good posterior estimates for the parameters of both variables across all different number of initial states. For the large dataset from $N$=1000, $T$=30, (30,000 observations with 2% missingness), the six-state solution was recovered in all but one of the initial number of states chosen.

**Table 4.2** Results from the Sequential Variational Bayes (SVB) fit of a six-state CT-mMixHMM to the 15,000-observation simulated dataset.

| | | **No. of initial states** | **No. of states found** | **Estimated posterior means** | **Estimated posterior st. error.** | **DIC** |
|---|---|---|---|---|---|---|
| **Dependent Variable 1** | N=500, T=30,2% missingness | 15 | 6 | 111.5,121.9,131.5, 141.7, 139.5, NA | 3.1,2.0, 5.2, 12.0, 14.3, NA | −205.3 |
| | | 12 | 6 | 111.1,121.6,131.0, 142.2, 138.9, NA | 3.1,2.1, 5.1, 10.0, 15.3, NA | −204.9 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 10 | 6 | 112.1,122.3,132.0, 141.2, 139.9, NA | 4.1,2.0, 5.1, 10.0, 14.3, NA | −203.7 |
| | | 5 | 6 | 111.1,121.3,131.0, 141.2, 139.9, NA | 3.1,2.0, 5.2, 11.0, 15.3, NA | −205.5 |
| Dependent Variable 2 | N=500, T=30, 2% missingness | 15 | 6 | 70.0,78.0,86.0,92.0, 82.0, NA | 3.9, 3.8, 4.8, 4.9, 5.7, NA | −653.2 |
| | | 12 | 6 | 70.1,71.1,86.1,91.9, 82.0, NA | 3.8, 3.9, 4.9, 4.8, 5.6, NA | −652.5 |
| | | 10 | 6 | 71.0,71.3,82.0,86.0, 81.1, NA | 3.4, 2.7, 2.5, 7.1, 6.1, NA | −653.0 |
| | | 5 | 6 | 71.3,71.7,82.3,86.1, 81.1, NA | 3.5, 2.8, 2.7, 8.1, 6.1, NA | −655.0 |

Note: DIC-Deviance Information Criteria

**Table 4.3** Results from the Sequential Variational Bayes (SVB) fit of a six-state CT-mMixHMM to the 30,000-observation (N=1000, T= 30) simulated dataset.

| | | **No. of initial states** | **No. of states found** | **Estimated posterior means** | **Estimated posterior st. error.** | **DIC** |
|---|---|---|---|---|---|---|
| **Dependent Variable 1** | N=1000, T=30,2% missingness | 15 | 8 | 111.0,122.0,132.0, 141.3,139.0,138.2, 140.2, NA | 3.6, 2.5, 5.6, 9.6,14.5,3.5,4.5, NA | −1450.8 |
| | | 12 | 6 | 111.1,122.1,132.0, 141.0, 139.0, NA | 3.5,2.5, 5.6, 9.5, 14.5, NA | −1451.5 |
| | | 10 | 6 | 111.2,121.9,131.8, 141.2, 139.9, NA | 3.1,2.1,5.2,11.0, 15.0, NA | −1451.2 |
| | | 5 | 6 | 111.1,121.4,131.6, 141.2, 139.1, NA | 3.1,2.1,5.2,11.6, 15.1, NA | −1452.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dependent Variable 2** | N=1000, T=30,2% missingness | 15 | 8 | 70.0,72.0,82.0,88.0, 82.0,89.2, 84.6, NA | 3.6,2.5, 4.5, 7.6, 6.4,6.2, 7.1, NA | −1,350.6 |
| | | 12 | 6 | 70.0,72.0,82.0,88.0, 82.0, NA | 3.6,2.5, 4.5, 7.6, 6.4, NA | −1,351.3 |
| | | 10 | 6 | 71.2,71.6,82.4,86.1, 81.3, NA | 3.4,2.7, 2.6, 8.1, 6.1, NA | −1,351.0 |
| | | 5 | 6 | 71.3,71.7,82.3,86.1, 81.1, NA | 3.5,2.8, 2.7, 8.1, 6.1, NA | −1,352.8 |

Note: DIC-Deviance Information Criteria

Tables 4.2 and 4.3 show that, in general, increasing the number of observations in the sample leads to posterior estimates that are closer to the population values of the model parameters (given in table 4.1) that were used to generate the data from the six-state CT-mMixHMM. Moreover, the solution obtained by starting with only five initial states also lead to smaller DIC values. These results suggest to us that the initial number of states can affect how observations are classified as the algorithm converges. Interestingly, in our results this effect was more pronounced when the number of observations was higher. Initializing the analysis for $N$=1000, $T$=30 observation dataset with 15 states leads to 8-state solution. However, two of these 8-states only had five observations assigned to it. This suggests that, if the initial number of states is too large, some of the observations can become "stuck" in their initial allocation. Intuitively, we would expect that, the larger the dataset, the more opportunity there is for that to happen, as there are more observations available to lend support to a superfluous state in the model. Therefore, using too many initial states may require caution.

For this thesis, the selection of the number of initial states was guided by a combination of prior knowledge and empirical model assessment. Existing research, expert insights, and established process stages were used to define a reasonable starting range for the number of possible states. This helped to ensure that the model remains interpretable and substantively meaningful. In addition, an empirical

approach was typically adopted by fitting models with different numbers of hidden states and comparing their performance using model comparison criteria, as well as examining interpretability, parameter stability, and the plausibility of inferred state dynamics. Therefore, the number of states was selected by balancing model fit, parsimony, and interpretability, rather than relying on a single criterion.

Because there is no observed data for state 6, its emission parameters cannot be updated using the likelihood. As a result, their posterior remains equal to the fixed informative prior. Missingness arising solely from the Death state has a localized and interpretable impact: it primarily affects inference on death-related transition intensities while leaving the estimation of non-terminal dynamics largely intact (see Figure 4.6). When appropriately modeled within a CT-MixHMM framework, such missingness did not degrade inference quality and may even strengthen identification of terminal states. More specifically, in terms of parameter estimation, missingness concentrated in the Death state has minimal impact on the estimation of emission parameters for the non-death states, since observations prior to death remain fully informative (see on supplementary materials S2: Appendix B, Table B1). Finally, Table 4.3 and Appendix B Table B1 illustrate the effect of increasing the proportion of missing observations from 2% to 3% for $N$=1000, $T$=30 on the performance of the Sequential Variational Bayes (SVB) estimator for CT-MixHMM. Overall, the results indicate that a modest increase in missingness leads to minimal impact on degradation in estimation precision and model fit when state-dependent missingness is explicitly modeled. This robustness highlights the importance of explicitly incorporating state-dependent missingness mechanisms when modeling ILD with absorbing events.

*Model convergence*

*ELBO Convergence for SVB algorithm of CT-mMixHMM*

Figure 4.2 presents the Evidence Lower Bound (ELBO) convergence trajectories from 250 independent runs of the Sequential Variational Bayes (SVB) algorithm applied to the six-state CT-mMixHMM. Each gray line represents the ELBO progression for a single run, while the solid blue curve

denotes the mean ELBO across all runs. The shaded orange region indicated the ±1 standard deviation (SD) band, reflecting the variability in convergence across runs.

Overall, the ELBO increased monotonically with iteration, demonstrating the stability and consistency of the SVB optimization process. After approximately 120–150 iterations, the ELBO values plateaued, suggesting that the algorithm reached a stable variational optimum. The relatively narrow SD band indicated low variability across runs, implying that the SVB algorithm yields reproducible solutions even when initialized with different random seeds. ELBO trajectories confirmed that SVB rapidly converged within approximately 150 iterations, demonstrating computational efficiency.

A few early fluctuations observed in the initial 50 iterations likely correspond to adaptive updates of the variational parameters as the posterior approximation adjusts to the latent structure. As iterations progressed, both the individual and mean ELBO curves stabilized, confirming that the SVB algorithm effectively optimized the variational evidence bound without signs of divergence or overfitting.

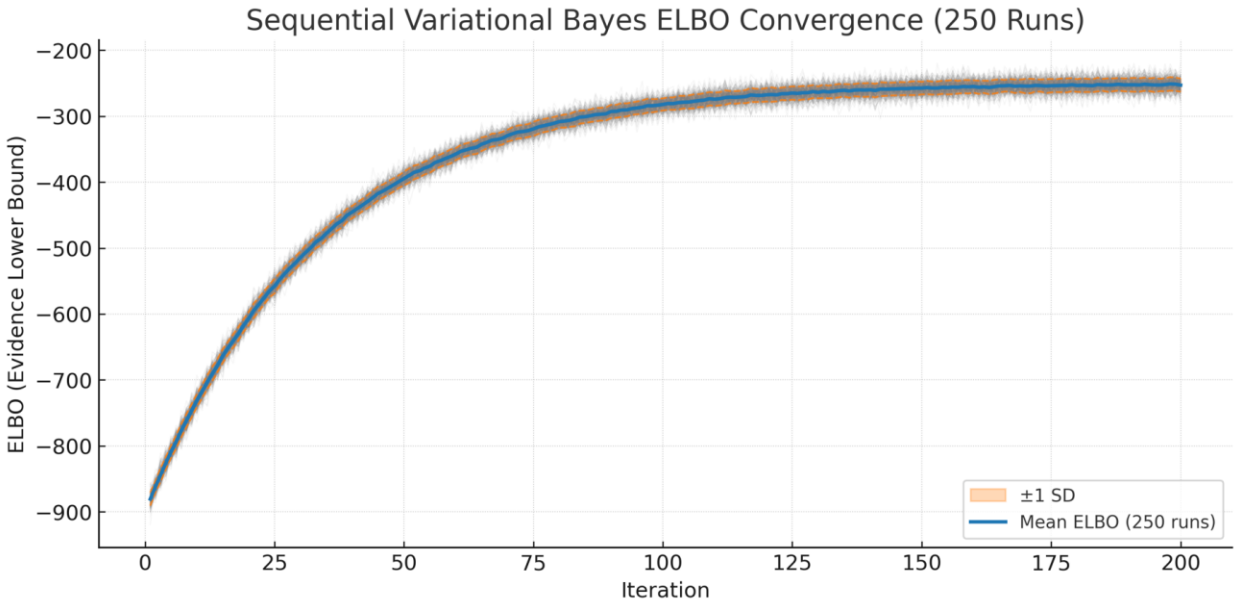



**Figure 4.2** Sequential Variational Bayes ELBO convergence plot for 250 iterations

In summary, the convergence profile supports the computational efficiency and robustness of the

SVB framework for parameter estimation in CT-mMixHMMs, even under complex hierarchical and continuous-time dependencies.

#### 4.5.1.2 *Emission Distribution*

Table 4.4 presents the mean (SD) values of the two dependent variables across the six identified states. For dependent variable 1, the mean rises from 111.2 mmHg (SD = 3.10) to 141.2 mmHg (SD = 11.72) in state 4, with state 3 at 131.7 (5.13) and state 5 at 139.2 (15.1). The slight drop in state 5 relative to state 4 may reflect partial self-control or an intervention effect. For dependent variable 2, the mean increases from 71.3 (3.58) mmHg in state 1 to 86.2 (8.04) in state 4, then decreases to 81.2 (6.23) in state 5. Moreover, the estimated standard deviations vary considerably across states, with lower states exhibiting relatively small variability and higher states showing substantially greater dispersion. This pattern suggests that while early states are well-defined and homogeneous, later states may capture more heterogeneous or transitional dynamics and may also reflect potential over-specification in the number of latent states.

**Table 4.4** Summary statistics of the estimated group-level means for each of the 6 states (component) distributions on the two dependent variables

| Parameter/Variables | States | Mean (SD) |
|---|---|---|
| **Dependent Variable 1** | State 1 | 111.2 (3.10) |
| | State 2 | 121.4 (2.12) |
| | State 3 | 131.7 (5.13) |
| | State 4 | 141.2 (11.72) |
| | State 5 | 139.2 (15.1) |
| | State 6 | NA (NA) |
| **Dependent Variable 2** | State 1 | 71.3 (3.58) |
| | State 2 | 71.7 (2.58) |
| | State 3 | 82.3 (2.86) |

| | | |
|---|---|---|
| | State 4 | 86.2 (8.04) |
| | State 5 | 81.2 (6.23) |
| | State 6 | NA (NA) |

Overall, the state classifications reflect the expected clinical progression of a condition - likely hypertension in case of MFUS dataset - and that the observed reductions may be associated with interventions such as antihypertensive treatment.

Figure 4.3 illustrates the distribution of outcome variables across the states and highlights the degree of separation or overlap among states.

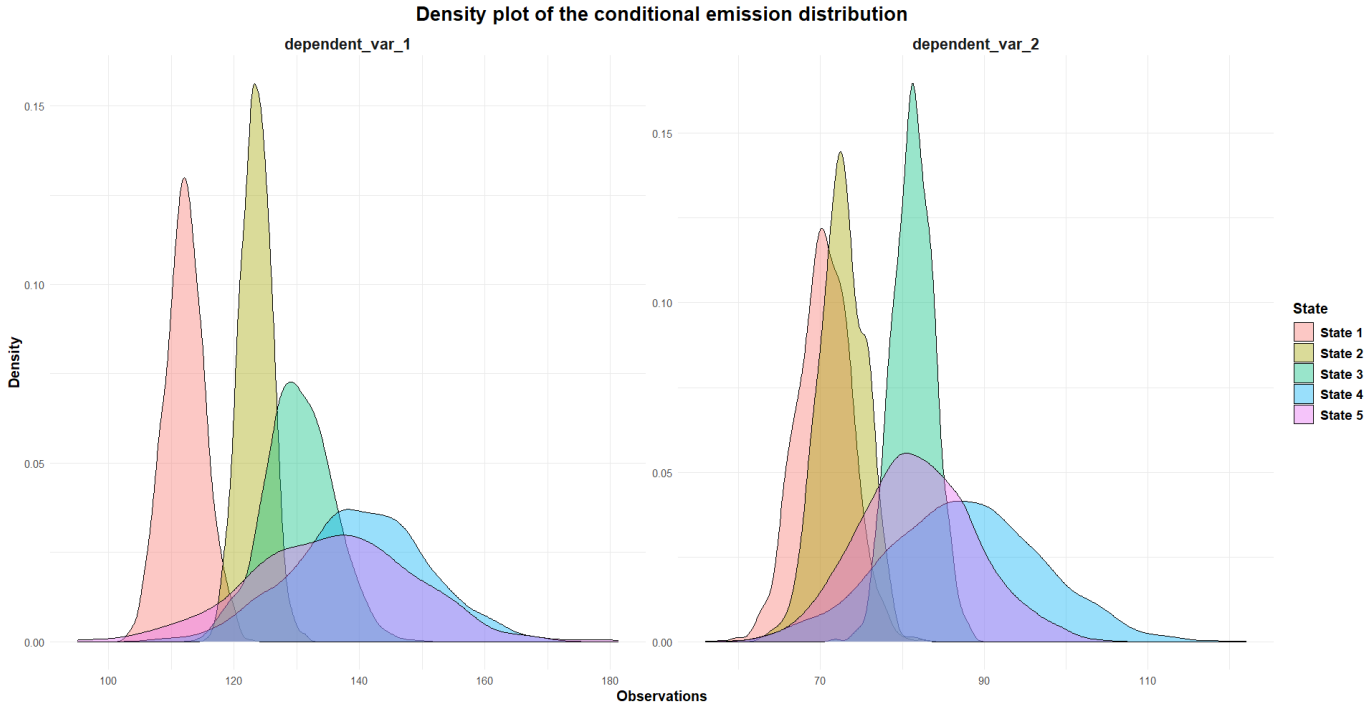



**Figure 4.3** Estimated density plot of the conditional emission distribution

#### 4.5.1.3 *Computational performance*

All experiments were conducted on the High-Performance Computer system of the Digital Research Alliance of Canada. The platform enabled parallel computation during the importance sampling phase, allowing for the concurrent calculation of the gradient, the Hessian of the joint log-likelihood, and the associated weights. Additionally, the system supported parallelization across Monte Carlo samples for

estimating expectations in the R-VGADM algorithm.

**Table 4.5** Average computing time (in seconds) for SVB and MCMC approaches in 250 simulated data sets from CT-mMixHMM model.

| | | **T=20** | **T=30** | **T=40** | **T=50** |
|---|---|---|---|---|---|
| **N=200** | MCMC | 410.85 | 1985.60 | 3025.90 | 4120.45 |
| | SVB | 178.40 | 480.70 | 465.30 | 390.50 |
| | **Ratio (MCMC/SVB)** | 2.30 | 4.13 | 6.50 | 10.55 |
| **N=500** | MCMC | 460.95 | 2880.75 | 5890.40 | 9225.60 |
| | SVB | 165.80 | 545.90 | 735.60 | 925.20 |
| | **Ratio (MCMC/SVB)** | 2.78 | 5.28 | 8.01 | 10.59 |
| **N=1000** | MCMC | 465.28 | 2955.27 | 3385.27 | 16695.92 |
| | SVB | 156.30 | 642.20 | 525.20 | 1404.20 |
| | **Ratio (MCMC/SVB)** | **2.96** | **5.50** | **6.40** | **11.89** |

Note: *N* represents number of individuals and *T* represents number of time points

Table 4.5 presents the average computing times (in seconds) for the Sequential Variational Bayes (SVB) and MCMC approaches, averaging over 250 simulated datasets generated from the CT-mMixHMM model. The results are reported across different sample sizes *N*=200, 500, 1000 and observation lengths *T*=20, 30, 40, 50. For each combination, MCMC times exceed SVB times, with the relative gain increasing as both *N* and *T* grow. Overall, SVB was 2 to 11 times faster than MCMC across all scenarios, with larger gains for bigger *N* and longer *T*. These findings highlight SVB's scalability for fitting CT-mMixHMM.

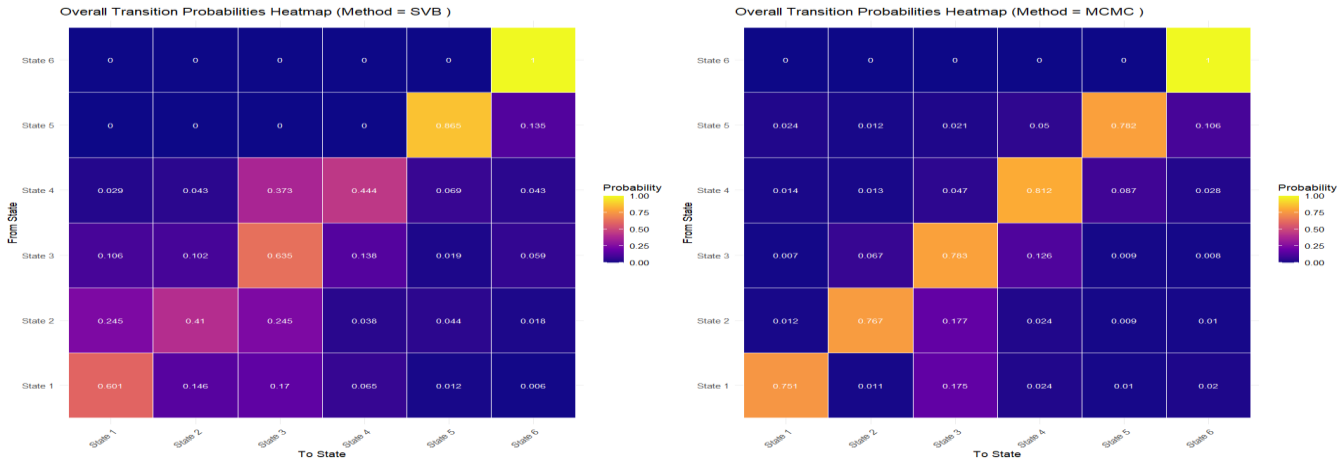
#### 4.5.1.4 *Transition Probability Matrix (TPM) accuracy and biases measures*

Figures 4.4 (a) and (b) compare the estimated average transition probabilities between the six generated states for $N$=1000, $T$=30, with 3% missingness: (a) SVB estimates and (b) MCMC estimates. Both heatmaps displayed a strong diagonal pattern, indicating high persistence in state over time. The full absorbing state (state 6) showed a probability of 1.0 in both estimates, and the partial absorbing state (state 5) exhibited high self-transition probabilities (0.865 for SVB vs. 0.846 for True), suggesting good accuracy of SVB in this category. Moreover, transitions from state 3 to state 4 are both close to 0.07 across methods.

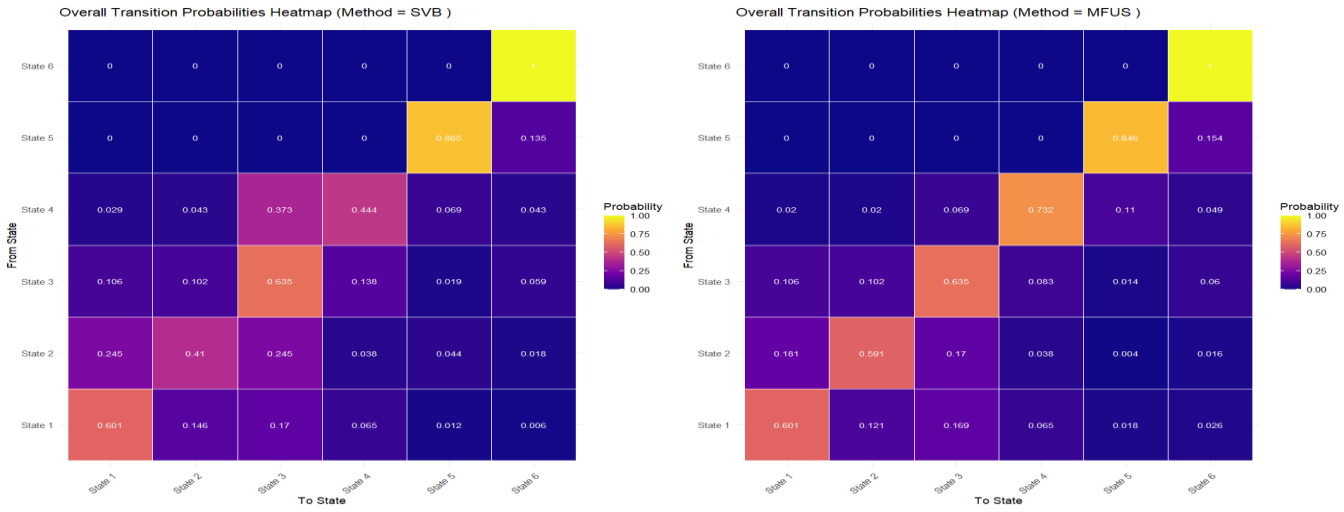
Overall, high diagonal entries indicate substantial state persistence over time. SVB achieves good accuracy of TPM relative to the true values and performed better than MCMC. High differences between MCMC and True TPM highlight method-specific misalignment that could influence clinical interpretation and policy decisions. Moreover, the SVB approach more accurately identifies state 6 as a full absorbing state and state 5 as a partial absorbing state than MCMC, making SVB the superior method for modeling transition probabilities.

The reason behind the efficiency of SVB is that in case of MCMC, it must simulate or marginalize over many latent state paths (especially in continuous time), (partial) absorbing events make the state path updates very sparse. This causes inefficient use of samples: most iterations repeat identical sequences and rarely sample partial or full absorbing state such as death events. Therefore, the posterior variance of transition intensities to (partial/full) absorbing states remains inflated, obscuring the absorbing nature in the posterior mean. Whereas SVB enforces (partial/full) absorbing constraints and penalizes unobserved transitions deterministically. Overall, SVB's bias–variance trade-off is better suited for models with (partial/full) absorbing or constrained transition structures-giving cleaner, interpretable transition matrices, especially when absorbing events are infrequent. Therefore, we report no further MCMC results beyond a computational time comparison for the simulation study.

Supplementary Table B2 (see Supplementary material S2, Appendix B) reports the bias of SVB and MCMC estimates relative to the MFUS benchmark across transition categories when $N$=1000, $T$=30 and missingness proportion was 3%.




a)




b)

**Figure 4.4** a) Population-level transition probability matrices (TPM): estimated via SVB and MCMC, b) Comparison of the SVB-estimated population-level TPM with the true underlying TPM used in the simulations.

Overall, SVB demonstrated smaller bias and lower mean absolute bias than MCMC, indicating closer agreement with the reference transition structure. In particular, MCMC exhibits larger deviations for dominant self-transition probabilities, whereas SVB provides more stable estimates across persistent and

deterioration transitions. Although MCMC performs comparably for some low-probability transitions, SVB shows more consistent alignment with the benchmark overall.

**Table 4.6** Estimated values of population-level TPM from SVB for simulation study

| | | **To (occasion t+1)** | | | | | |
|---|---|---|---|---|---|---|---|
| **From (occasion t)** | | State 1 | State 2 | State 3 | State 4 | State 5 | State 6 |
| | State 1 | 0.601 | 0.146 | 0.170 | 0.065 | 0.012 | 0.006 |
| | State 2 | 0.245 | 0.410 | 0.245 | 0.036 | 0.044 | 0.018 |
| | State 3 | 0.106 | 0.102 | 0.635 | 0.138 | 0.019 | 0.059 |
| | State 4 | 0.029 | 0.043 | 0.373 | 0.444 | 0.069 | 0.043 |
| | State 5 | 0.000 | 0.000 | 0.000 | 0.000 | 0.865 | 0.135 |
| | State 6 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |

Table 4.6 shows a well-ordered and interpretable transition structure among hidden health states such as blood pressure evolution, characterized by (1) strong persistence within states, (2) predominant forward progression toward more severe or treated stages, (3) limited transitions to earlier states, and (4) clear absorbing behavior in the terminal state. This pattern indicated the model's ability to capture both the chronic and terminal dynamics of health states such as hypertension in a longitudinal cohort.

*Model evaluation for estimated TPM*

*Bias Patterns*

Figure 4.5 presents the percentage bias in estimated transition probabilities under SVB only across under combinations of measurement occasions (20–50) and subject sample sizes ($N$ = 200, 500, 800, 1000). Each panel corresponds to a specific transition between hypertension-related states, with different line types representing sample sizes and shaded regions indicating variability across replicates.

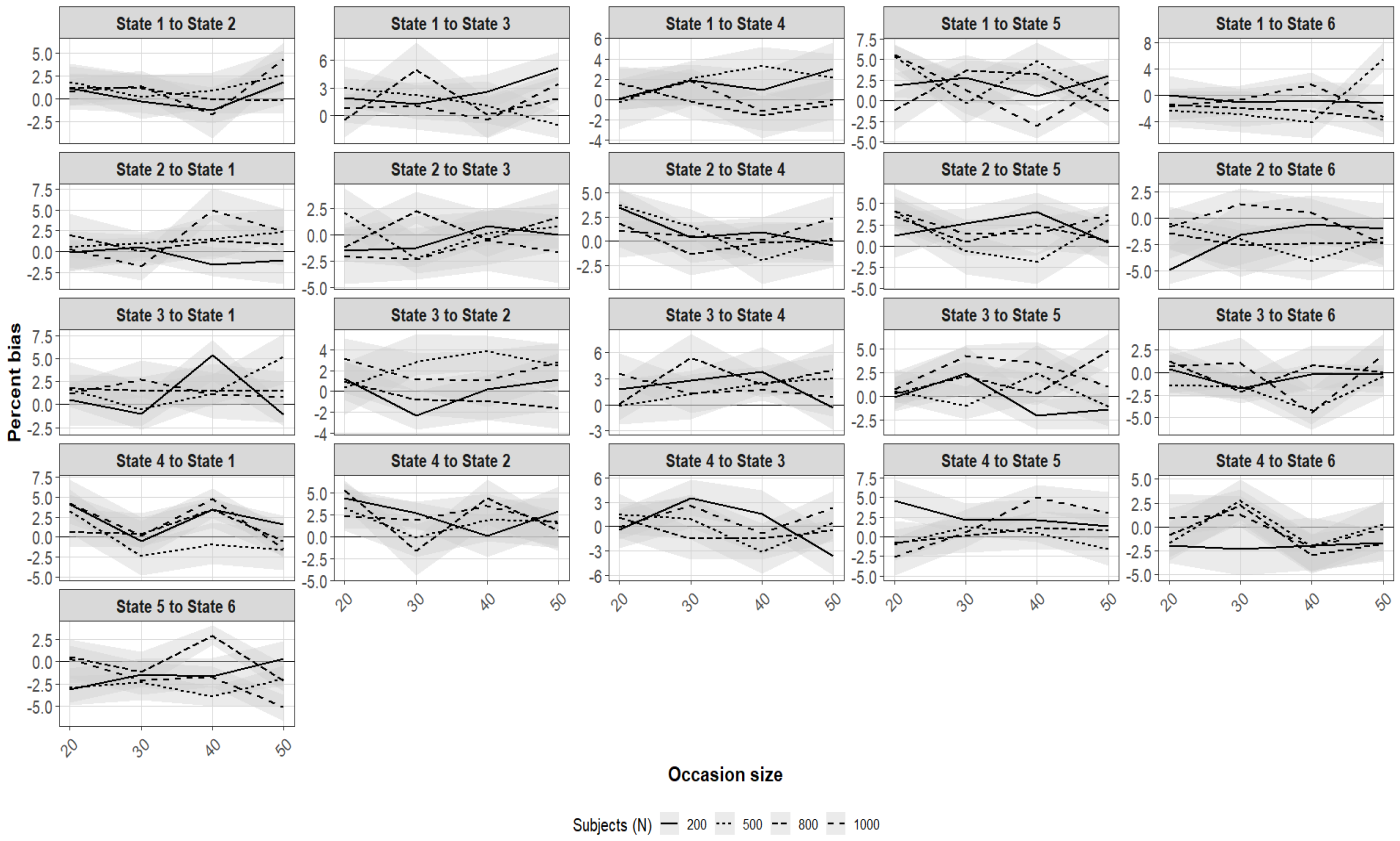



**Figure 4.5** Bias of estimated transition probabilities across various sample sizes and occasions

In general, the estimation process produces almost unbiased estimates of transition probabilities of most transitions, and these percent biases are usually clustered around the value of zero and fall within ±5%. Unsurprisingly, increasing the sample size reduces both bias and variability, indicating a higher precision of the estimates in bigger datasets. Expanding the number of measurement occasions from 20 to 50 yields only modest gains but contributes to greater stability.

Transitions involving less frequently visited states, such as those to or from State 5, show greater variability and are not always zero-biased, likely due to the lower number of observations for these transitions. Collectively, these results suggest that the SVB estimation method is robust across realistic simulation settings, with greater accuracy and reliability with larger samples.

*Coverage Patterns*

Figure 4.6 represents the fraction of the coverage that is achieved by the 95% confidence intervals associated with the estimated transition probabilities across combinations of sample sizes ($N$ = 200, 500, 800, 1000) and numbers of observation occasions ($T$=20, 30, 40, 50). Each panel represents a specific state-to-state transition; line types represent the varying sample size conditions.

Overall, empirical coverage is close to 95% for most transitions, thus suggesting that the proposed estimation procedure yields well-calibrated uncertainty quantification across a range of simulated scenarios. Transitions such as State 1 →State 2 and State 1 →State 3 show stable and near nominal coverage for all conditions examined, especially for larger sample sizes ($N$=800), thus showing promising potential for robust estimation of interval estimates for early-stage progression transitions. Conversely, transitions involving State 5 or State 6, e.g. State 5 → State 6 and State 2 → State 6, exhibit a slightly wider variability and cases of under-coverage, particularly at smaller sample sizes ($N$ = 200) and missingness solely from state 6, indicating increased uncertainty in calculating rare or terminal transitions.

Forward progression transitions (i.e., State 2 → State 3 and State 3→ State 4) resulted in coverage rates approaching 95% for most conditions, while reverse transitions (e.g., State 4 → State 1 and State 4 → State 2) exhibit slightly lower and more variable coverage, even in the large sample size.

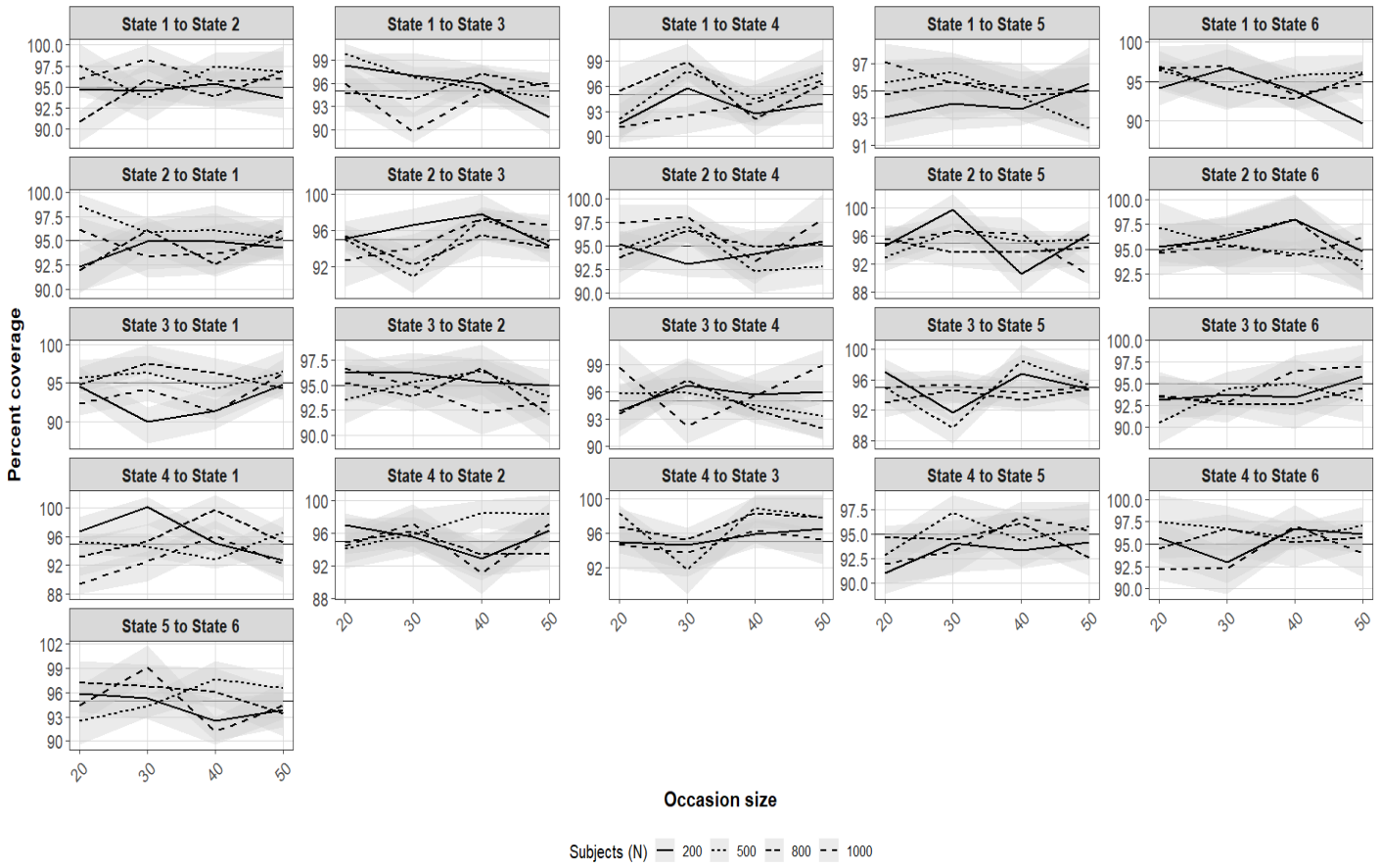



**Figure 4.6** Percent coverage of estimated transition probabilities across various sample sizes and occasions.

Increasing the number of occasions to $T = 40$ yields modest improvements in coverage stability. Taken together, these results show that the proposed SVB methodology provides reliable confidence interval estimates for common transitions, but transitions involving rare or less frequent events may result in sample sizes that are too large to maintain nominal levels of coverage. In other words, rare or terminal transitions may require larger sample sizes to maintain nominal coverage.

#### 4.5.1.5 *Maximum a posterior (MAP) estimates of model parameters with single covariate*

Supplementary Table B3 (Supplementary S2: Appendix B) presents the posterior median (MAP) estimates, posterior standard deviations, and 95% central credible intervals (CCI) for group-level and subject-level parameters obtained from 250 simulation replications of a six-state continuous-time

multivariate mixed-effects hidden Markov model (CT-mMixHMM) fitted using SVB. The simulated data consisted of $N = 1000$ subjects observed over $T = 30$ time points, with chosen ground-truth parameter values.

*Emission model parameters*

For both dependent variables (DV 1 and DV 2), the state-specific intercepts ($\alpha_{0k}$) exhibited a monotonic increase across latent states, consistent with the data-generating ordering. Across all states, the MAP estimates closely track the true intercept values, and 95% CCIs generally contain the true values. A modest positive bias was observed for higher-numbered states, particularly for DV 2 in States 4 and 5. This pattern is consistent with partial state overlap and reduced effective sample sizes in less frequently occupied states, which can lead to conservative upward shifts in posterior location estimates.

The state-specific slope parameters ($\beta_{1k}$) were well recovered for both outcomes. MAP estimates were close to the true values, and the associated CCIs consistently included the data-generating parameters. These results indicated that SVB accurately captures the temporal trends within states, even in the presence of latent state uncertainty and random effects.

Within-state residual variances ($\sigma^2_{\text{within}}$) are slightly overestimated across most states for both dependent variables. The degree of overestimation increases with state severity, a phenomenon commonly observed in mixture and hidden Markov models, where uncertainty in state allocation contributes to inflated residual variability. Importantly, the true variance values remained within the estimated CCI, suggesting appropriate uncertainty quantification.

*Subject-level random effects*

The variances of the subject-specific random intercepts for DV 1 and DV 2 ($\tau_1^2$ and $\tau_2^2$) were

moderately overestimated, although the true values were well covered by the 95% CCIs. This finding reflects the known tendency of variational inference methods to attribute a greater proportion of unexplained variability to random effects when jointly modeling latent states and hierarchical structures. Nevertheless, the results demonstrate that SVB successfully identifies substantial between-subject heterogeneity in longitudinal trajectories.

*Transition probability model*

The transition probability model (TPM) parameters, estimated on the logit scale, were recovered with high accuracy. Both the transition intercepts and covariate slopes exhibited minimal bias, tight posterior dispersion, and credible intervals that consistently include the true values. Transitions corresponding to clinically plausible state progressions (e.g., $1 \rightarrow 2$, $3 \rightarrow 4$, $4 \rightarrow 5$) were particularly well identified, while less frequent transitions (e.g., $4 \rightarrow 1$) displayed slightly wider credible intervals but no systematic bias.

The subject-level random-effect variance in the TPM was also accurately estimated, with MAP values closely matching the true parameter and narrow credible intervals. This indicated that SVB effectively captured individual-level heterogeneity in transition dynamics in a continuous-time setting.

*Initial state distribution*

The initial state probabilities were estimated with minimal bias, and the MAP estimates closely approximate the true proportions. As expected, credible intervals were wider for states with lower initial prevalence, reflecting greater uncertainty due to limited information at baseline. Nonetheless, the overall structure of the initial distribution was correctly recovered.

*Missingness mechanism*

The state-dependent missingness parameters were accurately estimated: the probability of missingness conditional on the absorbing state was recovered almost exactly, and the marginal missingness proportion closely matched the true value of 3%. These results demonstrate the SVB framework can jointly model the longitudinal process and informative missingness without introducing appreciable bias.

Overall, Supplementary Table B3 demonstrated that the proposed SVB approach provides reliable parameter recovery for a complex six-state CT-mMixHMM under realistic sample sizes and moderate observation lengths. The method performs particularly well in estimating transition dynamics and temporal effects, while exhibiting mild conservatism in variance and intercept estimation for higher latent states. Importantly, credible intervals consistently achieve appropriate coverage, supporting the validity of SVB for large-scale longitudinal modeling with latent state dynamics, subject-specific heterogeneity, and informative missingness.

### 4.5.2 Empirical Application Results

For the empirical application of our proposed model, we considered MFUS dataset. This section describes the posterior mean and standard errors of outcome variables for each hidden state, the estimated conditional emission distribution density plot and compares estimation methods for modeling state transitions and analyzes the age-related effects on transition rates. It highlights differences in model accuracy and presents individual case studies illustrating disease progression and management.

#### 4.5.2.1 *Model comparisons and state labelling*

Initially, we compared five- and six-states CT-mMixHMM via our proposed estimation algorithm and obtained the posterior mean and standard errors of two outcome variables - SBP and DBP - within

each hidden state. The results are shown in Table 4.7. The DIC indicated that the six-state model provided a better fit than the five-state model. We then labeled the six states based on the clinical relevance of BP measures.

**Table 4.7** Modelling results for the CT-mMixHMM with 5-7 latent states for MFUS datasets

| Variables | No. of states | Estimated posterior means (SBP, mmHg) | Estimated posterior std. error. (mmHg) | DIC |
|---|---|---|---|---|
| **SBP** | 5 | 111.0, 122.0, 132.0, 141.3, NA | 3.6, 2.5, 5.6, 9.6, NA | −1351.8 |
| | 6 | 112.1, 123.3, 130.4, 140.2, 136.2, NA | 3.2, 2.4, 5.5, 10.9, 13.1, NA | **−1355.5** |
| | 7 | 111.5, 124.5, 130.6, 143.5, 136.2, 142.6, NA | 3.6, 3.4, 6.5, 11.9, 12.1, 13.6, NA | **-1352.3** |
| **DBP** | 5 | 70.0, 72.0, 82.0, 88.0, NA | 3.6, 2.5, 4.5, 7.6, NA | −1,240.6 |
| | 6 | 70.3, 72.7, 81.3, 87.3, 82.2, NA | 3.2, 2.8, 2.4, 9.1, 7.2, NA | **−1,260.3** |
| | 7 | 71.3, 73.7, 82.5, 85.3, 87.2, 82.3, NA | 4.2, 4.8, 3.4, 8.1, 8.2, 8.3, NA | **−1,252.1** |

Table 4.8 shows the posterior means and standard errors of SBP and DBP for the six SVB-derived hidden states. From State 1 through State 4, both SBP and DBP increase monotonically. In State 5, SBP remains in the State 4 range, while DBP is slightly lower but still elevated, with greater between-subject variability and larger SEs, reflecting heterogeneity within this state.

**Table 4.8** Posterior means (standard errors) of SBP and DBP variables for each hidden state from six-states CT-mMixHMM via SVB

| State | SBP Mean | SBP SE | DBP Mean | DBP SE | State label |
|---|---|---|---|---|---|
| 1 | 112.1 | 3.26 | 70.3 | 3.28 | Normal |
| 2 | 123.3 | 2.41 | 72.7 | 2.88 | Elevated |
| 3 | 130.4 | 5.53 | 81.3 | 2.46 | Hypertension stage 1 |
| 4 | 140.2 | 10.98 | 87.3 | 9.04 | Hypertension stage 2 |
| 5 | 136.2 | 13.16 | 82.2 | 7.23 | On Hypertension Medication |
| 6 | NA | NA | NA | NA | Death |

State 6 is absorbing, with no SBP/DBP measurements. We label the states as Normal, Elevated, Hypertension Stage 1 (HTN 1), Hypertension Stage 2, On Hypertension Medication, and Death to aid interpretation of progression and treatment effects and to align with standard hypertension care.

#### 4.5.2.2 *Emission distribution by states*

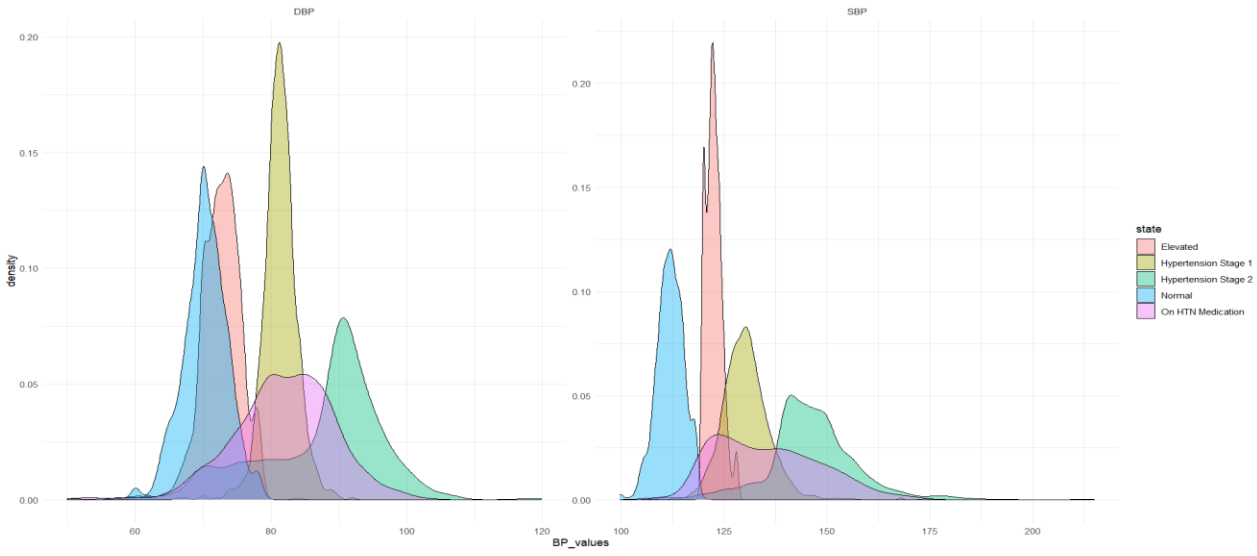



**Figure 4.7** Estimated emission density plot for SBP and DBP conditional on each hidden state

Figure 4.7 shows the estimated emission densities for SBP and DBP conditional on each hidden state. Normal and Elevated states cluster at lower end of SBP/DBP, while HTN1 and HTN2 progressively shift toward higher values. Patients on hypertension medication span both normal and hypertensive ranges, reflecting treatment effects. Overall, Figure 4.7 demonstrated graded differences across BP categories, but also notable overlap between states and substantial within-state variability, consistent with heterogeneity and medication effects.

We additionally performed a comprehensive performance analysis comparing SVB and MCMC for (a) estimating both individual transition probabilities and (b) aggregate age effects on transition rates across all participants (refer to Supplementary Figure 4-0-2 for details). Furthermore, aging effects on transition rates were evaluated for two specific case studies using our proposed algorithm (refer to Supplementary Figure4-0-3 with details).

#### 4.5.2.3 Maximum a posterior (MAP) estimates of model parameters with covariate

The MAP parameter estimates obtained from six state CT-mMixHMM are shown in Table 4.9a. We use medians of the posterior distributions as MAPs and use the standard deviations and 95% credible intervals (CI) as measures of uncertainty. In this analysis, age was group-mean centered at everyone's sample mean. Group-mean centering ensures emission model to individual specific baselines, prevents confounding between state occupancy and age distribution (means it is allowing latent states to be interpreted independently of the age distribution of the individuals who occupy them) and improves numerical stability in SVB.

For SBP, the estimated state-specific intercepts ($\beta_0$) representing the expected SBP evaluated at each individual's average age over follow-up, increased progressively from the Normal (=112.3 mmHg) state to the On HTN Medication state (=141.4 mmHg), reflecting clear graduations by hypertension severity. The age slope captures the within person effect of aging on SBP levels in different latent states.

Specifically, as an individual ages by one year relative to their own average age, their SBP increased steadily from 0.10mmHg conditional on remaining in the Normal state to 0.62mmHg conditional on remaining in On HTN Medication state, suggesting stronger longitudinal aging effects in increased SBP at higher hypertension states. Within-state residual variances ($\sigma^2_{within}$) of SBP rose from 8.6 $(mmHg)^2$ in the Normal state to 18.2 $(mmHg)^2$ in the On HTN Medication state, indicating greater within-person variability in systolic blood-pressure conditional on latent state membership, particularly among individuals in more advanced hypertensive states. The relatively large variance suggests that SBP trajectories in the On HTN Medication state are comparatively unstable after accounting for individual aging effects. Moreover, by the 95% CCIs of the parameter estimates, we can see that the uncertainty is larger around the fixed and random effect estimates of the "On HTN Medication" state. This is not unexpected given that it is the hardest state to estimate because it is always "stuck" across all other state distributions (see figure 4.7). Hence, it is associated with a large measure of uncertainty.

The random effects for SBP revealed substantial between-subject variability in baseline SBP after accounting for latent state membership and fixed covariate effects. The estimated random intercept variance ($\tau^2_{int}$) of 20.4 mmHg2 (95% CCI: [15.6, 26.4]) which reflects individuals differ markedly in their underlying SBP levels, even occupying the same latent state, and the random slope variance ($\tau^2_{age\_c}$)of 0.052 mmHg2 ([0.025, 0.084]) which indicated heterogeneity in age related SBP trajectories across individuals. More specifically, the effect of aging on SBP is not constant across subjects; rather, individuals exhibit varying rates of SBP increase with age within latent states. This supports the inclusion of random slopes to capture personalized disease progression patterns. The positive covariance between intercept and age ($\tau_{int,age\_c}$ = 0.91, 95% CCI: [0.37, 1.53]) which is credibly different from zero, indicating that individuals with higher baseline SBP also tend to experience steeper age-related increases in SBP over time.

**Table 4.9a** Point estimates MAP (medians) of emission parameters from six-states CT-mMixHMM via SVB method

| Variable | Parameter | MAP (SD) | 95% CCI |
|---|---|---|---|
| **SBP - Normal** | $\beta_0$ (Intercept) | 112.3 (1.0) | [110.1, 114.6] |
| | $\beta_{age\_c}$ (Age slope) | 0.1 (0.04) | [0.04, 0.16] |
| | $\sigma^2_{within}$(Residual variance) | 8.6 (0.8) | [7.3, 10.1] |
| **SBP - Elevated** | $\beta_0$ (Intercept) | 125.7 (1.0) | [123.4, 127.9] |
| | $\beta_{age\_c}$ (Age slope) | 0.22 (0.04) | [0.13, 0.31] |
| | $\sigma^2_{within}$(Residual variance) | 10.9 (0.8) | [9.2, 12.8] |
| **SBP - HTN Stage 1** | $\beta_0$ (Intercept) | 136.8 (1.0) | [134.1, 139.6] |
| | $\beta_{age\_c}$ (Age slope) | 0.33 (0.04) | [0.21, 0.45] |
| | $\sigma^2_{within}$(Residual variance) | 12.5 (0.8) | [10.5, 14.9] |
| **SBP - HTN Stage 2** | $\beta_0$ (Intercept) | 148.9 (1.0) | [146.0, 151.9] |
| | $\beta_{age\_c}$ (Age slope) | 0.47 (0.04) | [0.33, 0.61] |
| | $\sigma^2_{within}$(Residual variance) | 15.8 (0.8) | [13.2, 18.6] |
| **SBP - On HTN Med** | $\beta_0$ (Intercept) | 141.4 (1.0) | [137.8, 165.2] |
| | $\beta_{age\_c}$ (Age slope) | 0.62 (0.04) | [0.45, 0.78] |
| | $\sigma^2_{within}$(Residual variance) | 18.2 (0.8) | [15.1, 21.9] |
| **Random Effects - SBP** | $\tau^2_{int}$(Intercept) | 20.4 (2.9) | [15.6, 26.4] |
| | $\tau^2_{age\_c}$(Age) | 0.052 (0.015) | [0.025, 0.084] |
| | $\tau_{int,age\_c}$ (Intercept, Age) | 0.91 (0.30) | [0.37, 1.53] |

A similar pattern was observed for DBP in table 4.9b. The intercepts which indicated the expected DBP evaluated at each individual's average age over follow-up, rose from 70.5 mmHg in the Normal state to 98.6 mmHg in the On HTN Medication state, while the age effects on the level of DBP, when an individual ages by one year relative to their own average age, their DBP increased from 0.04 mmHg conditional on remaining in the Normal state to 0.37mmHg on being in the On HTN Medication state

suggesting a stronger age related increase in DBP. Residual variance of DBP also increased from 5.9 (mmHg)2 in the Normal state to 10.7 (mmHg)2 in the On HTN Medication state, mirroring the heterogeneity pattern seen in SBP. The random effects for DBP showed notable between-person variation ($\tau^2_{int}$= 11.6; $\tau^2_{age\_c}$= 0.026), with a moderate positive intercept-age covariance ($\tau_{int,age\_c}$= 0.39), again implying that participants with higher baseline DBP exhibit stronger age-related increases.

**Table 4.9b** Point estimates MAP (medians) of emission parameters from six-states CT-mMixHMM via SVB method

| Variable | Parameter | MAP (SD) | 95% CCI |
|---|---|---|---|
| **DBP - Normal** | $\beta_0$ (Intercept) | 70.5 (0.9) | [69.1, 72.0] |
| | $\beta_{age\_c}$ (Age slope) | 0.04 (0.03) | [0.01, 0.08] |
| | $\sigma^2_{within}$(Residual variance) | 5.9 (0.6) | [4.8, 7.1] |
| **DBP - Elevated** | $\beta_0$ (Intercept) | 78.2 (0.9) | [76.5, 80.0] |
| | $\beta_{age\_c}$ (Age slope) | 0.1 (0.03) | [0.04, 0.17] |
| | $\sigma^2_{within}$(Residual variance) | 6.7 (0.6) | [5.5, 8.1] |
| **DBP - HTN Stage 1** | $\beta_0$ (Intercept) | 85.6 (0.9) | [83.7, 87.5] |
| | $\beta_{age\_c}$ (Age slope) | 0.2 (0.03) | [0.11, 0.29] |
| | $\sigma^2_{within}$(Residual variance) | 7.8 (0.6) | [6.3, 9.5] |
| **DBP - HTN Stage 2** | $\beta_0$ (Intercept) | 91.8 (0.9) | [89.6, 94.0] |
| | $\beta_{age\_c}$ (Age slope) | 0.28 (0.03) | [0.17, 0.39] |
| | $\sigma^2_{within}$(Residual variance) | 9.4 (0.6) | [7.5, 11.4] |
| **DBP - On HTN Med** | $\beta_0$ (Intercept) | 98.6 (0.9) | [96.2, 101.1] |
| | $\beta_{age\_c}$ (Age slope) | 0.37 (0.03) | [0.25, 0.50] |
| | $\sigma^2_{within}$(Residual variance) | 10.7 (0.6) | [8.7, 12.9] |
| **Random Effects - DBP** | $\tau^2_{int}$(Intercept) | 11.6 (1.9) | [8.2, 15.4] |
| | $\tau^2_{age\_c}$ (Age) | 0.026 (0.008) | [0.011, 0.045] |
| | $\tau_{int,age\_c}$ (Intercept, Age) | 0.39 (0.15) | [0.08, 0.65] |

Overall, the component distribution random effects are large on the "On HTN Med" state for both SBP and DBP outcome variables (see table 4.9a-b), which suggests that it may be useful to incorporate some additional covariates such as BMI to explain differences between subjects.

*MAP estimates of state transition probability*

In table 4.10 for the transition probability model (TPM), the baseline logit-scale intercepts ($\alpha$) representing baseline measures of the transition when the person is at their own average age, were negative and consistent with relatively low transition probabilities between hypertension states in any single time step. Transition intensities generally decreased as the gap between states widened-for example, $\alpha_{12} = -1.15$ and $\alpha_{13} = -2.38$ - indicating that adjacent-state transitions (e.g., Normal → Elevated) are more likely than larger jumps (e.g., Normal → HTN1). The reverse transition from HTN2 → Normal ($\alpha_{41} = -3.15$) had the smallest estimated probability, consistent with the low likelihood of returning to normotensive levels.

**Table 4.10** Point estimates MAP of TPM parameters from the six-states CT-mMixHMM via SVB

| Variable | Parameter | MAP (SD) | 95% CCI |
|---|---|---|---|
| **TPM (Intercepts )** | $\xi_{0,12}$ (1→2) | -1.15 (0.17) | [-1.47, -0.85] |
| | $\xi_{0,13}$ (1→3) | -2.38 (0.26) | [-2.90, -1.93] |
| | $\xi_{0,23}$ (2→3) | -1.22 (0.19) | [-1.64, -0.87] |
| | $\xi_{0,34}$ (3→4) | -1.05 (0.21) | [-1.48, -0.65] |
| | $\xi_{0,45}$ (4→5) | -0.85 (0.18) | [-1.21, -0.52] |
| | $\xi_{0,41}$ (4→1) | -3.15 (0.33) | [-3.79, -2.53] |
| **TPM Random Effects** | $\tau^2_{int,(TPM)}$(Intercept) | 0.68 (0.11) | [0.48, 0.94] |
| | $\tau^2_{age\_c\ (TPM)}$(Age) | 0.052 (0.020) | [0.021, 0.096] |
| | $\tau_{int,age\_c\ (TPM)}$ (Intercept, Age) | -0.015 (0.007) | [-0.029, -0.003] |

Since, age was group-mean centered at the subject's baseline age, the random intercept variance reflects between-subject heterogeneity in baseline transition propensities at the baseline age, while the random slope variance reflects heterogeneity in sensitivity to age deviations from baseline age. The small negative intercept-slope covariance indicated that individuals with higher baseline propensities tend to have slightly weaker age effects on transitions.

The random effects for the TPM showed moderate between-subject variability in both intercepts ($\tau^2_{int,(TPM)}$= 0.68) and age slopes ($\tau^2_{age\ (TPM)}$= 0.052). Here, $\tau^2_{age\ (TPM)}$= 0.052 indicated a moderate level of between person variability in age effects on transition probabilities across individuals. In other words, not everyone ages the same way in terms of how their transition likelihoods change -some individuals show stronger age effects relative to their average age, others weaker. For example, in hypertensive disease progression, some patients may show rapid transitions with age relative to their baseline age, while others remain stable. Overall, some people's transition probabilities are more sensitive to age than others (see supplementary figure 4-0-2 and 4-0-3).

The small but significant negative covariance ($\tau_{int,age\ (TPM)}$ = –0.015) indicated that individuals with higher baseline transition propensities tend to have slightly weaker age effects on transition probabilities. The predicted probability of missingness conditional on the "Death" state was recovered almost exactly.

Overall, these findings underscore clear differences in blood pressure patterns across latent hypertension states, with age exerting a stronger influence and within-person variability increasing at more severe stages. Importantly, the impact of aging on blood pressure is not uniform; individuals show different rates of increase with age within these states. In addition, the estimated transition parameters indicate that progression typically occurs through neighboring states, while abrupt shifts or complete

reversals are unlikely.

***Density of posterior transitions probabilities***

The observed differences in the transition probabilities across subjects is not unexpected given the substantial heterogeneity among MFUS participants in medication initiation. Figure 4.8 displays the density estimates of the 26 posterior means transitions probabilities in the six-states CT-mMixHMM fitted to MFUS data.

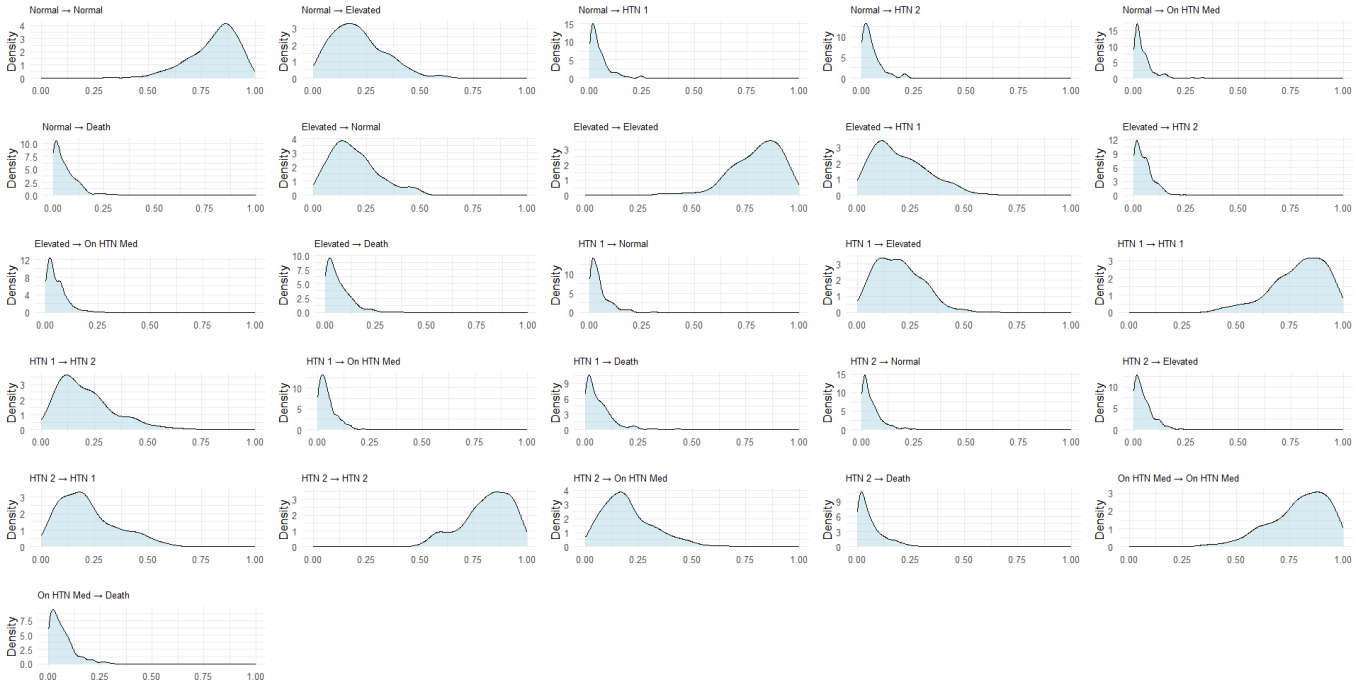



**Figure 4.8** Density estimate of the 26 possible transition sets of maximum a posterior (MAP) mean transition probabilities for the six-states CT-mMixHMM

In Figure 4.8, the diagonal panels (e.g., Normal → Normal, Elevated → Elevated, HTN Stage 1 → HTN Stage 1, HTN Stage 2 → HTN Stage 2, On HTN Medication → On HTN Medication) peak near one, indicating strong self-transitions and persistence in the current blood pressure state over consecutive observations. Such persistence aligns with the chronic and relatively stable nature of hypertension once established.

Forward transitions (Normal → Elevated, Elevated → HTN Stage 1, HTN Stage 1 → HTN Stage 2, and HTN Stage 2 → On HTN Medication) show moderately high-density peaks between 0.6 and 0.9. These transitions reflect gradual worsening of hypertension conditions over time, consistent with the expected physiological progression from normal blood pressure to more severe stages and eventual treatment initiation. The smooth, unimodal shapes of these densities indicate stable and well-estimated progression probabilities across individuals.

Backward transitions (e.g., Elevated → Normal, HTN Stage 1 → Elevated, HTN Stage 2 → HTN Stage 1) display densities skewed toward lower probabilities (typically below 0.3), implying that recovery or improvement to milder states is relatively uncommon.

The transition from On HTN Medication → Death has a slightly higher and broader density, indicating greater variability in mortality risk among treated hypertensive individuals. This is consistent with clinical observations that those requiring medication typically represent more advanced disease stages.

Overall, the transition probability densities reveal a coherent and clinically plausible dynamic structure for MFUS participants. The model captures high within-state persistence, dominant forward progression, limited regression, and a fully absorbing death state. The results indicate that blood pressure dynamics follow a gradual, largely irreversible trajectory from normal to elevated and hypertensive states, with increasing mortality risk as the disease advances.

#### 4.5.2.4 *Individual state sequences estimate*

Figure 4.9 presents state-sequence comparisons for four individuals, showing how closely the model-predicted states track the observed states over time. The SVB fitted state sequences align well with the observed data across individuals, capturing general trends effectively, demonstrating the practicality

of CT-mMixHMM for longitudinal blood pressure analysis. Overall, the fitted state sequences largely match the observed states, supporting the model's validity.

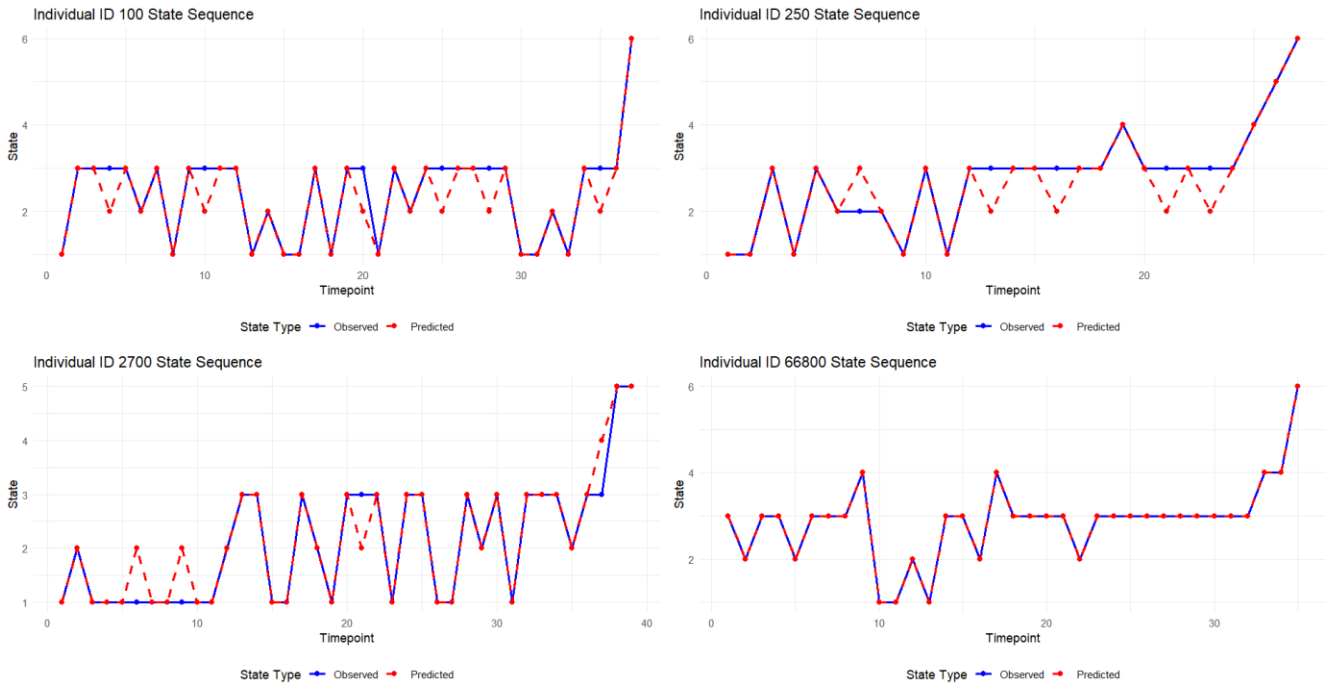



**Figure 4.9** Fitted states sequences from SVB method for 4 randomly selected individuals

### 4.5.2.5 *Model fit through posterior prediction checks (PPC)*

We checked whether the model produces extreme results on the component (state) distribution means and variances. For the state means, the posterior predictive p-values indicated that the simulated values are not extreme compared to the observed datasets for Normal, Elevated, HTN1 and HTN2 across both outcome variables ($0.25 \leq P_{psterior} \leq 0.57$ ). However, for the On HTN Med state, the p-values suggest some extremeness in the simulated data compared with the observed dataset (see Figure 4.10).

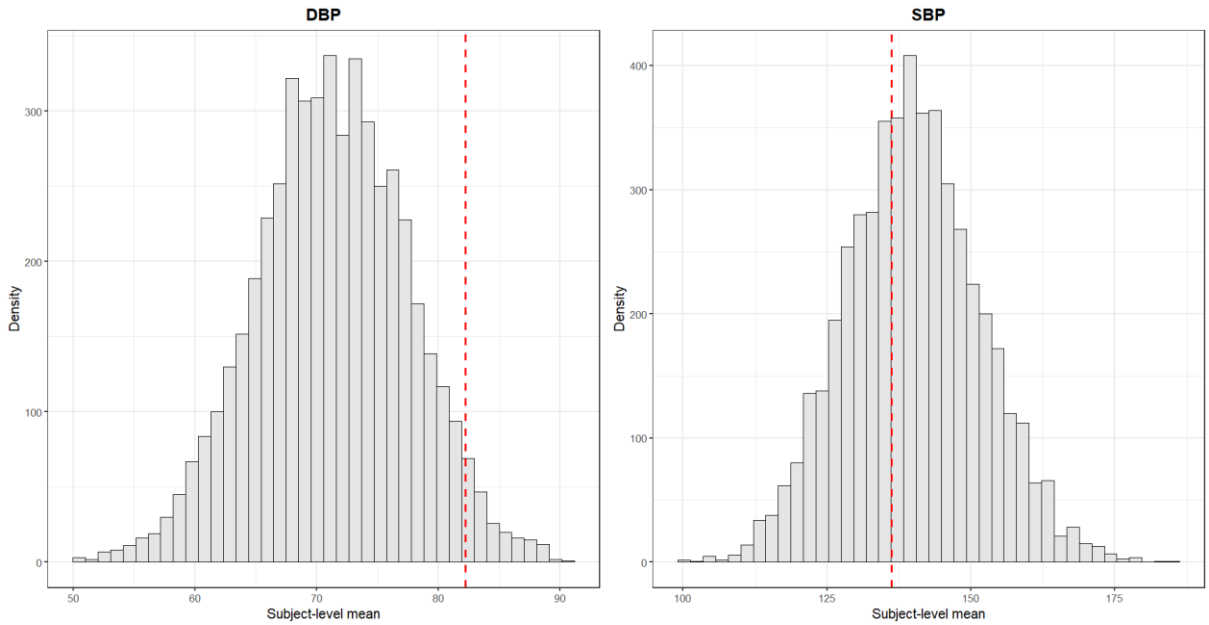



**Figure 4.10** Results of PPC for the "On HTN Med" state of the two emission variables. Vertical dashed line (in red) represents the observed mean; DBP: Diastolic Blood Pressure and SBP: Systolic Blood Pressure

Figure 4.10 plots SVB-based predictions: the red vertical line marks the observed mean, and the grey histogram shows the distribution of means from the generated samples. The generated samples overshoot the observed mean for the second outcome variable (SBP) and undershoot the observed mean for the first outcome variable (DBP) for the On HTN Medication state. This is not surprising given that the component distribution of the "On HTN Med" state overlaps significantly with the other states on all two outcome variables (see Figure 4.7). PPC indicated no extreme means or variances for Normal, Elevated, HTN1, HTN2; the On HTN Med state shows some extremeness, as noted above. Overall, the model fits MFUS BP emissions adequately.

#### 4.5.2.6 *Clinical interpretations and implications*

The empirical application of the CT-mMixHMM to the MFUS dataset demonstrated that blood pressure dynamics in the cohort follow a gradual and largely irreversible progression from normal to

elevated and hypertensive states, with mortality risk increasing as the disease advances. The model's transition probability densities and fitted state sequences provide strong evidence of high within-state persistence and dominant forward movement, with regression to lower-risk states being infrequent. These findings align with the chronic and progressive nature of hypertension and support the clinical plausibility of the modelling approach.

The posterior predictive checks (PPC) indicate that the latent states produced by the model adequately reproduce the observed distributional features for most states; the On HTN Med state shows some overlap with other states, reflecting the heterogeneity of treatment effects, but the overall fit remains robust. Clinically, the model can help identify individuals at higher risk of progression or death, inform risk stratification and targeted interventions, and illuminate how treatments influence disease trajectories. Overall, this continuous-time, multivariate mixed effects HMM provides a valuable framework for dissecting complex, latent disease pathways in longitudinal BP data and supports more precise and responsive patient care.

**4.6 Discussion**

The present study systematically evaluated the performance and interpretability of continuous-time multivariate mixed effects Hidden Markov Models (CT-mMixHMM) for ILD, using both simulation scenarios and empirical analysis of the MFUS blood pressure data. The simulation study established the model’s robustness under increasing complexity, demonstrating consistent recovery of hidden state sequences and parameter estimates across varying state structures, random effects, and covariate influences. These controlled scenarios provided a benchmark for model validity and sensitivity, allowing for a nuanced appraisal of CT-mMixHMM’s flexibility and reliability. Moreover, the state sequences inferred via the SVB approach in the empirical analysis closely aligned with observed transition patterns at the individual level, providing further evidence of the model’s ability to accurately reconstruct

underlying latent disease processes.

The simulation results demonstrate that the SVB approach provides a scalable, numerically stable, and fully integrated inference strategy for complex CT-mMixHMMs. The framework accommodates multiple longitudinal outcomes, state-dependent trajectories, hierarchical random effects, covariate-dependent transition dynamics, and informative missingness within a single variational objective. Although variational inference tends to be mildly conservative in variance estimation a known property- it consistently preserves structural interpretability and avoids the computational burden of fully Bayesian MCMC approaches. These findings for the variance estimation are consistent with previous studies [38, 42]. Moreover, these findings support SVB as a practical and theoretically sound alternative for large-scale continuous-time latent state modeling in longitudinal health data [33]. Empirical application to the MFUS data furnished real-world corroboration, with fitted transition dynamics and emission distributions aligning well with clinical expectations. Together, the simulation and empirical findings reinforce the CT-mMixHMM framework as a powerful approach for dissecting latent disease progression pathways in longitudinal BP studies. These findings related to empirical applications are also consistent with previous studies [7, 8, 10].

The model's transition probability structure revealed a clinically coherent progression from normotensive to hypertensive states, characterized by high within-state persistence and dominant forward movement. Recession to lower-risk states was rare, mirroring the largely irreversible nature of advancement of hypertension. The presence of a fully absorbing death state further underscores the model's capacity to capture terminal events in chronic disease contexts. These findings carry significant clinical implications: the observed state dynamics not only reflect the natural history of BP elevation but also enable stratification of patients by risk and trajectory, providing a basis for targeted intervention and monitoring which are consistent with previous studies [15, 42]. Furthermore, the influence of aging on transition rates

observed in empirical data applications demonstrated a stepwise progression from Normal to Elevated, HTN1, and HTN2 stages, accompanied by an increased probability of treatment initiation (Medication) and mortality (Death) at subsequent stages. These findings emphasize the critical value of early identification and management of elevated blood pressure to delay or prevent advancement to later stages, which are linked to greater morbidity and mortality risks, consistent with prior research [42, 43].

Model adequacy was assessed with the Deviance Information Criterion (DIC) [38] and the posterior predictive checks (PPC) [44]. PPC results indicated no evidence of extreme means or variances for most latent states, affirming the model's ability to reproduce the distributional features of the observed data. The notable exception was the "On HTN Med" state, where discrepancies in simulated versus observed means suggest some overlap and potential misclassification, likely due to the heterogeneity of treatment effects and patient responses which is also consistent with the existing study [44]. Nevertheless, overall DIC values and graphical diagnostics support the appropriateness of the CT-mMixHMM for the MFUS BP data, with residual deviations offering insight into areas for model refinement. SVB-estimated states closely matched observed individual transitions, validating accurate recovery of latent disease dynamics.

While general trends were well captured, some individual-level mismatches- particularly in the "On HTN Med" state - highlight the challenge of modelling states with substantial overlap or clinical ambiguity. These limitations are consistent with the simulation study, which showed that increased state complexity and emission overlap modestly reduced sequence recovery fidelity that supports the findings of the existing study [44]. Despite these caveats, the high correspondence between predicted and empirical sequences demonstrated the practical utility of CT-mMixHMM in real-world longitudinal data contexts.

### 4.6.1 Strengths and Limitations

The primary strength of the CT-mMixHMM approach lies in its capacity to model continuous-

time, multivariate, and mixed effects processes, thereby accommodating the rich heterogeneity and irregular sampling inherent in longitudinal BP studies. Its latent state framework enables nuanced characterization of disease dynamics, facilitating both individual and population-level inference. Furthermore, the proposed sequential variational Bayes algorithm (R-VGADM) addresses the intractability of computing the partial log-likelihoods' gradient and Hessian at each observation in dependent mixture models by leveraging Fisher's and Louis' identities to obtain substantially faster unbiased estimates.

Despite these advantages, several limitations should be considered. First, the model relies on parametric emission distributions, which may not adequately capture non-standard or highly skewed outcome variables. Second, the overlap of emission distributions, particularly in states reflecting treatment or transitional phases, can impair states' distinguishability and posterior inference [44]. Third, computational demands increase with model complexity, underscoring the need for scalable inference algorithms in large-scale applications [45].

The selection of six states in this research was motivated by two considerations. (1) It represents a moderately complex latent structure-more demanding than simple two-or three-state models, yet still feasible for stable estimation in a continuous time multilevel setting [38]. (2) Empirical applications of hidden Markov and regime-switching models to intensive longitudinal data frequently involve a small to moderate number of latent states, making six a reasonable and representative choice for a baseline justification. Importantly, our findings are not inherently tied to the specific number of states. That said, the quantitative performance metrics (e.g., convergence rates, estimator variance, and computation time) may vary as the number of states changes [44, 38].

Finally, the simulation study conducted in this thesis was a matched data-generating and analytic model to establish baseline properties such as bias, efficiency, coverage and convergence of our proposed method, and to demonstrate correct parameter recovery prior to comparing alternative modeling

frameworks [46, 47]. Although conducting simulation studies is considered good practice in latent variable and longitudinal modeling research, it often neglects the evaluation of robustness [48].

In case of real data application, the "On HTN Medication" state was intended to represent a clinically observable treatment status rather than a detailed pharmacological treatment-response model. Due to limitations in the available MFUS dataset and the scope of this thesis, medication use was incorporated as a simplified binary treatment-related latent state indicating whether antihypertensive medication had been initiated. The current CT-mMixHMM framework therefore assumed that entering the medication state reflects an important clinical transition in disease management, but it did not explicitly differentiate medication types, dosage intensity, treatment adherence, combination therapies, or changing treatment effectiveness over time. These factors may contribute to heterogeneity within the medication state and may influence both transition dynamics and outcome trajectories. This is consistent with the larger variability observed within the "On HTN Medication" state, where patients may respond differently to treatment depending on individual clinical characteristics and medication effectiveness.

From a modeling perspective, incorporating detailed time-varying medication information would substantially increase model complexity - for example, by modeling medication type and effectiveness as time-varying covariates in treatment-specific transition intensities, or by allowing dynamic intervention effects within the hidden-state framework. Such extensions would require more detailed pharmacological information, larger effective sample sizes within treatment subgroups, and substantially more complex estimation procedures. Therefore, the current thesis viewed it as an initial latent-state disease progression framework rather than a full causal treatment-effect model.

### 4.6.2 Future Directions

Building on the strengths and limitations identified in the present study, several avenues for future research emerge. First, the development and evaluation of non-parametric or semi-parametric emission

models could help address the challenges posed by non-standard or skewed outcome distributions, potentially enhancing model flexibility and fit in heterogeneous clinical datasets [11, 14]. Second, strategies to better resolve overlapping latent states-such as incorporating auxiliary biomarkers or leveraging advanced clustering techniques-should be explored to improve state distinguishability, particularly in treatment and transitional phases [44].

Additionally, future work might investigate the integration of time-varying covariate effects and the explicit modelling of external interventions, such as medication changes or lifestyle modifications, to further refine the predictive and explanatory power of the CT-mMixHMM framework. In the current model, we included main covariate effects, age, because age is strongly associated with hypertension progression and transition dynamics. However, higher-order interaction effects, such as age and BMI, age and medication use, or BMI and comorbidity status, were not explicitly incorporated in the current implementation. We acknowledge that such interaction effects may be clinically meaningful because the impact of one risk factor may depend on another. For example, increasing BMI may accelerate hypertension progression differently across age groups, or medication effectiveness may vary according to age and obesity status. Therefore, incorporating interaction effects could represent an important future extension of the CT-mMixHMM framework and could further improve individualized disease progression modeling and clinical interpretability.

The development of scalable inference algorithms will be critical for applying these models to large population-based studies. Additionally, we openly recognize that evaluating robustness against misspecification-such as continuous-time generative processes, different dependence structures, or simplified latent dynamics-may be a valuable next phase and an obvious expansion of this study. Finally, research focused on enhancing model interpretability and embedding CT-mMixHMM outputs within clinical decision support tools could accelerate the translation of these statistical advances into routine patient care and population health management.

### 4.6.3 Study Implications

The findings have immediate relevance for clinical monitoring and risk stratification in hypertension management. By elucidating the latent trajectories underlying observed BP patterns, CT-mMixHMM provides a principled framework for anticipating disease progression and tailoring intervention strategies. For researchers, the integration of simulation and empirical validation offers a template for robust model evaluation, highlighting the importance of both controlled experimentation and real-world testing. Future methodological work should explore flexible, possibly nonparametric, emission models and strategies for mitigating state overlap, as well as the incorporation of dynamic covariate effects and external interventions. Enhanced model interpretability and integration with clinical decision support systems represent promising avenues for translating statistical advances into improved patient outcomes.

## 4.7 Conclusions

This study demonstrated the utility and robustness of continuous-time multivariate mixed effects Hidden Markov Models (CT-mMixHMM) for analyzing ILD in both simulated and real-world clinical settings. The approach effectively captures latent disease progression, accommodates complex data structures, and provides clinically meaningful insights into risk stratification and patient trajectories. While the model exhibits strong performance in reconstructing hidden states and transition dynamics, challenges remain regarding emission distribution overlap and computational scalability. Addressing these limitations through methodological innovations - such as nonparametric emission models, enhanced state resolution strategies, and scalable inference algorithms-will further strengthen its application in large-scale population health studies. Eventually, the integration of CT-mMixHMM outputs into clinical decision support systems holds promise for advancing personalized hypertension management and improving patient outcomes in diverse healthcare contexts.

## 4.8 References


1. S. De Haan-Rietdijk, J. M. Gottman, C. S. Bergeman, and E. L. Hamaker, “Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation,” *Psychometrika*, vol. 81, no. 1, pp. 217–241, 2016, doi: 10.1007/s11336-014-9417-x.
2. J. H. L. Oud and M. J. M. H. Delsing, "Continuous time modeling of panel data by means of SEM," in *Longitudinal research with latent variables*, K. van Montfort, J. H. L. Oud, and A. Satorra, Eds., Springer, 2010, pp. 201–244.
3. M. C. Voelkle, J. H. L. Oud, E. Davidov, and P. Schmidt, "An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia," *Psychological Methods*, vol. 17, no. 2, pp. 176-192, Apr. 2012, doi: 10.1037/a0027543.
4. P. R. Deboeck and K. J. Preacher, "No need to be discrete: A method for continuous time mediation analysis," *Struct. Equ. Modeling: A Multidisciplinary Journal*, vol. 23, pp. 61–75, 2016, doi: 10.1080/10705511.2014.973960.
5. C. H. Jackson and L. D. Sharples, “Hidden Markov model for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients,” *Statistics in Medicine*, vol. 21, pp. 113–128, 2002.
6. F. Bortolucci and A. Farcomeni, “A shared-parameter continuous-time hidden Markov and survival model for longitudinal data with informative dropout,” *Statistical Methods in Medical Research*, vol. 38, pp. 1056–1073, 2019.
7. Y. Luo and D. A. Stephens, "Bayesian inference for continuous-time hidden Markov models with an unknown number of states," *Statist. Comput.*, vol. 31, 2021, doi: 10.1007/s11222-021-10032-8. Available: https://doi.org/10.48550/arXiv.2106.10660

8. R. M. Altman, “Mixed hidden Markov models-An extension of the hidden Markov model to the longitudinal data setting,” *Journal of the American Statistical Association*, vol. 102, pp. 201–210, 2007.
9. F. Bartolucci, A. Farcomeni, and F. Pennoni, *Latent Markov Models for Longitudinal Data*, 1st ed. New York: Chapman and Hall/CRC, 2012. doi: 10.1201/b13246
10. A. Bureau, S. Shiboski, and J. P. Hughes, “Application of continuous-time hidden Markov models to the study of misclassified disease outcomes,” *Statistical Methods in Medical Research*, vol. 22, pp. 441–462, 2003.
11. E. H. Ip, W. J. Rejeski, T. B. Harris, and S. B. Kritchevsky, “Partially ordered hidden mixed Markov model for the disablement process of older adults,” *Journal of the American Statistical Association*, vol. 108, pp. 370–384, 2013.
12. K. Kang, J. Cai, X. Song, and H. Zhu, “Bayesian hidden Markov model for delineating the pathology of Alzheimer disease,” *Statistical Methods in Medical Research*, vol. 28, pp. 2112–2124, 2019.
13. Y. Y. Liu, S. Li, F. Li, and J. M. Song, “Efficient learning of continuous hidden Markov models for disease progression,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 28, pp. 3599–3607, 2015.
14. X. Zhou, K. Kang, and X. Song, “Two-part hidden Markov models for semi-continuous longitudinal data with nonignorable missing covariates,” *Statistical Methods in Medical Research*, vol. 39, pp. 1801–1816, 2020.
15. Z. Huang, G. Wang, J. B. Jonas, C. Ji, S. Chen, Y. Yuan, C. Shen, Y. Wu, and S. Wu, "Blood pressure control and progression of arteriosclerosis in hypertension," *J. Hypertens.*, vol. 39, no. 6, pp. 1221-1229, Jun. 2021, doi: 10.1097/hjh.0000000000002758. PMID: 33470733

16. A. Maruotti, “Mixed hidden Markov models for longitudinal data: An overview,” *International Statistical Review*, vol. 79, no. 3, pp. 427–454, 2011, doi: 10.1111/j.1751-5823.2011.00160.x.
17. S. Pandolfi, F. Bartolucci, and F. Pennoni, "A hidden Markov model for continuous longitudinal data with missing responses and dropout," *Biom. J.*, vol. 65, no. 5, p. e2200016, Jun. 2023, doi: 10.1002/bimj.202200016. PMID: 37035989.
18. Cappe, O., Moulines, E., and Ryden, T. (2005) Inference in Hidden Markov Models, 2nd Edition, Spinger, NY.
19. D. R. Cox and H. D. Müller, *The Theory of Stochastic Processes*, 1st ed. London, U.K.: Chapman and Hall, 1965.
20. C. H. Jackson, “Multi-state models for panel data: The msm package,” *Journal of Statistical Software*, vol. 38, no. 8, pp. 1–29, 2011, doi: 10.18637/jss.v038.i08.
21. N. Bartolomeo, P. Trerotoli, and G. Serio, “Progression of liver cirrhosis to HCC: An application of hidden Markov model,” *BMC Medical Research Methodology*, vol. 11, Art. no. 38, pp. 1–8, 2011.
22. J. Lange, R. Hubbard, L. Inoue, and V. Minin, “A joint model for multi-state disease process and random information observation times, with application to electronic health records data,” *Biometrics*, vol. 71, pp. 821–831, 2015.
23. D. J. Raffa and J. A. Dubin, “Multivariate longitudinal data analysis with mixed effect hidden Markov models,” *Biometrics*, 2015, doi: 10.1111/biom.12096.
24. S. Mildiner-Moraga and E. Aarts, "A Bayesian multilevel hidden Markov model with Poisson-lognormal emissions for intense longitudinal count data," *arXiv:2403.12561v1 [stat.ME]*, Mar. 19, 2024. [Online]. Available: https://arxiv.org/abs/2403.12561

25. J. P. Williams, C. B. Storlie, T. M. Therneau, C. R. Jack Jr., and J. Hannig, “A Bayesian approach to multi-state hidden Markov models: Application to dementia progression,” *Journal of the American Statistical Association*, pp. 1–21, 2019.

26. T. P. Minka, *A Family of Algorithms for Approximate Bayesian Inference*, Ph.D. dissertation, MIT, 2001.

27. S. Ji, B. Krishnapuram, and L. Carin, “Variational Bayes for continuous hidden Markov models and its application to active learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 522–532, 2006.

28. A. Tancredi, “Approximate Bayesian inference for discretely observed continuous-time multi-state models,” *Biometrics*, vol. 75, no. 3, pp. 966–977, Sep. 2019, doi: 10.1111/biom.13019.

29. R. B. Tate, T. E. Cuddy, and F. A. L. Mathewson, "Cohort Profile: The Manitoba Follow-up Study (MFUS)," *Int. J. Epidemiol.*, vol. 44, no. 5, pp. 1528-1536, Oct. 2015, doi: 10.1093/ije/dyu141. PMID: 25064641.

30. F. A. Mathewson and G. S. Varnam, "Abnormal electrocardiograms in apparently healthy people. I. Long term follow-up study," *Circulation*, vol. 21, pp. 196-203, Feb. 1960, doi: 10.1161/01.cir.21.2.196. PMID: 14422280.

31. E. M. Hoque, E. F. Acar, and M. Torabi, “A time-heterogeneous D-vine copula model for unbalanced and unequally spaced longitudinal data,” *Biometrics*, vol. 79, no. 2, pp. 734–746, 2023.

32. McCulloch, R. E., & Tsay, R. S. (1994). Bayesian Analysis of Autoregressive time series via the Gibbs Sampler. *Journal of Time Series Analysis*, *15*(2), 235–250. https://doi.org/10.1111/j.1467-9892.1994.tb00188.x

33. A. Bao, D. Gunawan, and A. Z. Mangion, “R-VGAL: A sequential variational Bayes algorithm for generalized linear mixed models,” *Statistics and Computing*, vol. 34, Art. no. 110, pp. 1–15, 2024.

34. M. Lambert, S. Bonnabel, and F. Bach, “The recursive variational Gaussian approximation (R-VGA),” *Statistics and Computing*, vol. 32, no. 11, pp. 1–14, 2022.

35. C. Arnold, L. Biedebach, A. Küpfer, and M. Neunhoeffer, "The role of hyperparameters in machine learning models and how to tune them," *Political Sci. Res. Methods*, vol. 12, no. 4, pp. 841-848, Oct. 2024, doi: 10.1017/psrm.2023.61.

36. S. M. Lynch and B. Western, "Bayesian posterior predictive checks for complex models," *Sociol. Methods Res.*, vol. 32, no. 3, pp. 301–335, 2004, doi: 10.1177/0049124103257303.

37. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. New York: Chapman and Hall/CRC, 2013. doi: 10.1201/b16018.

38. C. A. McGrory and D. M. Titterington, “Variational Bayesian Analysis for Hidden Markov Models”, *Aust. N.Z. J. Stat.,* 51(2), 227-244, 2009, doi: 10.1111/j.1467-842X.2009.00543.x.

39. M. Bladt and M. Sorensen, “Statistical Inference for discretely observed Markov jump process”, *J. Royal. Stat. Soc. Series B (Statistical Methodology),* 67, 395-410, 2005.

40. Aarts, E and Mildiner, S. “ mHMMbayes - An R package for multilevel Hidden Markov Models using Bayesian Estimation, version 1.1.1.

41. D. M. . Hughes, M. García-Fiñana, and M.P. Wand. “Fast approximate inference for multivariate longitudinal data,” *Biostatistics.* Dec 12;24(1):177-192, 2022. doi: 10.1093/biostatistics/kxab021. PMID: 33991420; PMCID: PMC9748580.

42. V. A. Shaffer, P. Wegier, K. D. Valentine, J. L. Belden, S. M. Canfield, M. Popescu, L. M. Steege, A. Jain, and R. J. Koopman, “Use of Enhanced Data Visualization to Improve Patient Judgments about Hypertension Control”, *Med Decis Making*. 40(6):785-796, 2020. doi: 10.1177/0272989X20940999.

43. T. A. Gowan, M. D. Tringali, J. A. Hostetler, J. Martin, L. I. Ward-Geiger, and J. M. Johnson, "A hidden Markov model for estimating age-specific survival when age and size are uncertain". *Ecology*. 102(8):e03426, 2021, doi: 10.1002/ecy.3426.

44. G. Jasper, S.M. Morga, and E. Arts, "Sample size considerations for Bayesian Multilevel Hidden Markov Models: A Simulation Study on Multivariate continuous time data with highly overlapping component distributions based on sleep data", arxiv.org.2001.09033: 1-45, 2022.

45. Wang, P., and P. Blunsom, "Collapsed Variational Bayesian Inference for Hidden Markov Models", *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS),* 2013, Scottsdale, AZ, USA.

46. S. Robinson, "Tutorial on simulation model verification and validation", *Proceedings of the Operational Research Society Simulation Workshop,* 2021, DOI: https://doi.org/10.36819/SW21.006.

47. R. G. Sargent, "Verification and validation of simulation models", *Journal of Simulation,* 7(1), 12-24, 2013, DOI: https://doi.org/10.1057/jos.2012.20.

48. Y. Chen, and S. Zhang, "A Latent Gaussian process model for analysing intensive longitudinal data". *The British journal of mathematical and statistical psychology*, 73(2), 237-260, 2019.

49. Y. Xia, J. Chen, and D. Jiang, "Variational Bayesian analysis for two-part latent variable model". *Computational Statistics,* 2024, 39(4), 2259-2290.

50. Y. Xia, Y. Xue, and D. Jiang, "Variational Bayesian inference for sparse item response theory models". *British Journal of Mathematics and Statistical Psychology*. 2026.

51. A. Rahman, D. Jiang, and L. Lix, "Efficient and accurate variational inference for multilevel threshold autoregressive models in intensive longitudinal data. *British Journal of Mathematics and Statistical Psychology*, 2025, https://doi.org/10.1111/bmsp.12381.

## 4.9 Supplementary Material

### (S1) Appendix A: Derivation of gradient and Hessian expressions for R-VGADM algorithm using Fisher's and Louis' identities

The gradient of the observed-data log-likelihood defined in equation (4.a) and used in the optimization function for equation (6), is given by the variational posterior expectation of the complete-data score function, while the Hessian decomposes into the expected complete-data curvature term reflecting uncertainty in the latent variables. This decomposition explains both the intractability of direct maximization and the effectiveness of R-VGADM algorithm for continuous-time hidden Markov models considered in section 2.

***Gradient approximation with Fisher's identity***

Consider the $i$th iteration. Fisher's identity (Cappe et al., 2005) for the gradient of $\log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ is

$$\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \left[\sum_{\boldsymbol{z}\in\mathcal{Z}(\boldsymbol{M})} \int p(\,\boldsymbol{z_i}\mid \boldsymbol{u_i};\boldsymbol{\Theta}\,)\, p(\boldsymbol{u_i};\boldsymbol{\Theta})\, \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{y_i}\mid \boldsymbol{z_i},\boldsymbol{u_i},\boldsymbol{\Theta}\,)\; d\boldsymbol{u_i}\right]. \qquad (A.1)$$

We now give the expression for $\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ for the continuous time multivariate mixed hidden Markov model considered in section 2.

For $i = 1, \cdots, N$, the gradient $\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ can be written as

$$\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{y_i}\mid \boldsymbol{z_i},\boldsymbol{u_i},\boldsymbol{\Theta}\,) + \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{z_i}\mid \boldsymbol{u_i},\boldsymbol{\Theta}\,) + \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{u_i}\mid \boldsymbol{\Theta}\,) \qquad (A.2)$$

$$= \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{y_i}\mid \boldsymbol{z_i},\boldsymbol{u_i},\mathbf{B},\boldsymbol{\Sigma}_{\varepsilon}\,) + \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{z_i}\mid \boldsymbol{\pi},\boldsymbol{\xi},\boldsymbol{V_i}\,) + \nabla_{\boldsymbol{\Theta}} \log p(\,\boldsymbol{u_i}\mid \boldsymbol{\Sigma}\,), \qquad (A.3)$$

where,

$$\log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon)$$

$$= -\frac{n_i R}{2}\log(2\pi) - \frac{n_i}{2}\log|\boldsymbol{\Sigma}_\varepsilon| - \frac{1}{2}\,(\boldsymbol{y_i} - \boldsymbol{\eta_i})^{\mathrm{T}}\left(\mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}\right)(\boldsymbol{y_i} - \boldsymbol{\eta_i}), \qquad (A.4)$$

$$\log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) = -\frac{q}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma_u}| - \frac{1}{2}\,\boldsymbol{u_i}^T\,\boldsymbol{\Sigma_u}^{-1}\,\boldsymbol{u_i}\ , \qquad (A.5)$$

and from equation (4) for $i$ th individual on notational simplicity, we avoid indicator variables and express the latent-state likelihood directly in terms of the realized state sequence.

$$\log p(\boldsymbol{z_i} \mid \boldsymbol{\pi}, \boldsymbol{\xi}, \boldsymbol{V_i}) = \log \pi_k + \sum_{t=2}^{n_i} {d_{i,t}}^{j,k}\ , \qquad (A.6)$$

where,

$${d_{i,t}}^{j,k} = \sum_{l=1}^{K}\sum_{m\neq l}\left\{N_{i,l,m}^{j,k}(\Delta_{i,t})\,{\boldsymbol{V_i}}^T\xi_{lm} - \exp\left({\boldsymbol{V_i}}^T\xi_{lm}\right) D_{i,l}^{j,k}(\Delta_{i,t})\right\},$$

records the probability of transition from state $j$ to state $k$ in the interval $\Delta_{i,t}$ and $N_{i,l,m}^{j,k}(\Delta_{i,t})$ and $D_{i,l}^{j,k}(\Delta_{i,t})$ are the number of jumps from state $l$ to state $m$, and time spent in state $l$ for a Markov chain initiated at state $j$ at time $\tau_{i,t-1}$ and ending in state $k$ at time $\tau_{i,t}$.

Now the gradient of the joint $\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i} \mid \boldsymbol{\Theta})$ is

$$\nabla_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i} \mid \boldsymbol{\Theta}) = \nabla_{\boldsymbol{\Theta}}\left[\log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon) + \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) + \log p(\boldsymbol{z_i} \mid \boldsymbol{\pi}, \boldsymbol{\xi}, \boldsymbol{V_i})\right] \qquad (A.7)$$

$$= \begin{pmatrix} \nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) \\ \nabla_{\boldsymbol{\beta}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) \\ \nabla_{\boldsymbol{\Sigma}_{\varepsilon}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) \\ \nabla_{\boldsymbol{\Sigma}_u} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) \\ \nabla_{\boldsymbol{\pi}} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) \\ \nabla_{\boldsymbol{\xi}} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) \end{pmatrix}, \qquad (A.8)$$

where each component is given by

$$\nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) = \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{\Sigma}_{\varepsilon}^{-1} (\boldsymbol{y_i} - \boldsymbol{\eta_i}), \qquad (A.9)$$

$$\nabla_{\boldsymbol{\beta}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\Sigma}_{\varepsilon}^{-1} (\boldsymbol{y_i} - \boldsymbol{\eta_i}) \qquad (A.10)$$

$$\nabla_{\boldsymbol{\Sigma}_{\varepsilon}} \log p(\boldsymbol{y_i} \mid \boldsymbol{z_i}, \boldsymbol{u_i}, \mathbf{B}, \boldsymbol{\Sigma}_{\varepsilon}) = -\frac{n_i}{2} [\boldsymbol{\Sigma}_{\varepsilon}^{-1} - \boldsymbol{\Sigma}_{\varepsilon}^{-1} (\boldsymbol{y_i} - \boldsymbol{\eta_i}) (\boldsymbol{y_i} - \boldsymbol{\eta_i})^{\mathrm{T}} \boldsymbol{\Sigma}_{\varepsilon}^{-1}] \qquad (A.11)$$

$$\nabla_{\boldsymbol{\Sigma}_u} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) = -\frac{1}{2} [\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1} \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_u^{-1}] \qquad (A.12)$$

$$\nabla_{\boldsymbol{\pi}} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) = \frac{\mathbb{I}(\boldsymbol{z}_{i1} = k) + a_k - 1}{\pi_k} \quad (include\ the\ Dirichlet\ prior\ gradient) \qquad (A.13)$$

$$\nabla_{\boldsymbol{\xi}} \log p(\boldsymbol{u_i} \mid \boldsymbol{\Sigma_u}) = \sum_{t=2}^{n_i} \sum_{l=1}^{K} \sum_{m \neq l} \left\{ N_{i,l,m}^{j,k}(\Delta_{i,t}) - \exp(\boldsymbol{V_i}^{\mathrm{T}} \xi_{lm}) D_{i,l}^{j,k}(\Delta_{i,t}) \right\} \boldsymbol{V_i}, \qquad (A.14)$$

The gradient of the complete-data log likelihood decomposes additively across emission, transition, and random-effect components, yielding block-separable score functions with respect to the model parameters.

***Hessian approximation with Louis' identity***

Similarly, the approximation of the Hessian using Louis' identify requires $\nabla^2_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ in particular,

$\nabla^2_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$

$$= \begin{bmatrix} \nabla^2_{\boldsymbol{\alpha}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\xi}} \\ & & \cdots & & & \\ & & \cdots & & & \\ & & \cdots & & & \\ & & \cdots & & & \\ \nabla^2_{\boldsymbol{\xi},\boldsymbol{\alpha}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\xi},\boldsymbol{\beta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\xi},\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\xi},\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\xi},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) & \nabla^2_{\boldsymbol{\xi},} \end{bmatrix}$$

Now the components of $\nabla^2_{\boldsymbol{\Theta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ are given by

$$\nabla^2_{\boldsymbol{\alpha}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = -\boldsymbol{e}_i^{\mathrm{T}}\, \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{e}_i \qquad (A.15)$$

$$\nabla^2_{\boldsymbol{\beta}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = -\boldsymbol{x}_i^{\mathrm{T}}\, \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{x}_i \qquad (A.16)$$

$$\nabla^2_{\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \frac{1}{2}\, \boldsymbol{\Sigma}_\varepsilon^{-1} \otimes \boldsymbol{\Sigma}_\varepsilon^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{y_i} - \boldsymbol{\eta_i})\,(\boldsymbol{y_i} - \boldsymbol{\eta_i})^{\mathrm{T}}\, \boldsymbol{\Sigma}_\varepsilon^{-1} \otimes \boldsymbol{\Sigma}_\varepsilon^{-1} \qquad (A.17)$$

$$\nabla^2_{,\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \frac{1}{2}\, \boldsymbol{\Sigma}_u^{-1} \otimes \boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}(\boldsymbol{u_i})\,(\boldsymbol{u_i})^{\mathrm{T}}\, \boldsymbol{\Sigma}_u^{-1} \otimes \boldsymbol{\Sigma}_u^{-1} \qquad (A.18)$$

$$\nabla^2_{\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \begin{cases} -\dfrac{n_k + a_k - 1}{\pi_k^2}, \text{if } k = j. \\ 0 \end{cases} \qquad (A.19)$$

$$\nabla^2_{\boldsymbol{\xi},} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = -\sum_{i,t} \exp\left(\boldsymbol{V_i}^{\mathrm{T}} \xi_{lm}\right) D_{i,l}^{j,k}\left(\Delta_{i,t}\right)\, \boldsymbol{V_i}\, \boldsymbol{V_i}^{\mathrm{T}} \qquad (A.20)$$

$$\nabla^2_{\boldsymbol{\alpha},\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\beta},\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\Sigma_\varepsilon},\boldsymbol{\Sigma_u}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\Sigma_u},\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = 0 \qquad (A.21)$$

$$\nabla^2_{\boldsymbol{\alpha},\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = -\boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{e}^{\mathrm{T}}(\boldsymbol{y_i} - \boldsymbol{\eta_i}) \text{ and } \nabla^2_{\boldsymbol{\beta},\boldsymbol{\Sigma_\varepsilon}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = -\boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{y_i} - \boldsymbol{\eta_i}) \qquad (A.22)$$

$$\nabla^2_{\boldsymbol{\alpha},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\beta},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\alpha},\boldsymbol{\xi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\beta},\boldsymbol{\xi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = 0 \qquad (A.23)$$

$$\nabla^2_{\boldsymbol{\Sigma_\varepsilon},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\Sigma_\varepsilon},\boldsymbol{\pi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\Sigma_\varepsilon},\boldsymbol{\xi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = \nabla^2_{\boldsymbol{\Sigma_\varepsilon},\boldsymbol{\xi}} \log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta}) = 0 \qquad (A.24)$$

**Prior distributions for model parameters**

$$\boldsymbol{B} = (\boldsymbol{\alpha}, \boldsymbol{\beta}) \sim N\,(\mathbf{0}, \tau^2 \boldsymbol{I}),\ \ \tau^2 \ \rightarrow\ \infty$$

$$\boldsymbol{\xi_{0,lm}} \sim N\,(\mathbf{0}, \epsilon^2 \boldsymbol{I})\,, 1 \leq l \neq m \leq 6\,, \epsilon^2 \ \rightarrow\ \infty$$

$$\boldsymbol{\xi_{1,lm}} \sim N\,(\mathbf{0}, \rho^2 \boldsymbol{I}), 1 \leq l \neq m \leq 6\,, \rho^2 \ \rightarrow\ \infty$$

$$\boldsymbol{\Sigma_{\varepsilon,i}} = \ \text{blockdiag}\,[\boldsymbol{\Sigma_\varepsilon}\,, \boldsymbol{\Sigma_\varepsilon} \cdots, \boldsymbol{\Sigma_\varepsilon}\,],$$

where,

$$\boldsymbol{\Sigma_\varepsilon} = \text{diag}\,(\log \sigma^2_{\varepsilon 1}\,, \log \sigma^2_{\varepsilon 2}\,, \cdots, \log \sigma^2_{\varepsilon R}\,)$$

$$\log \sigma^2_{\varepsilon R}\ \sim N(\log(0.5^2), 1)$$

$$\boldsymbol{\Sigma_u} \sim IW\ (\vartheta_0\,, \boldsymbol{\Psi_0}), \boldsymbol{\Psi_0}\ \text{be the}\ 2 \times 2\ \text{identiy matrix}$$

An inverse-Wishart prior with identity scale matrix was placed on the random-effects covariance matrix, allowing flexible estimation of cross-outcome dependence while maintaining a weakly informative prior centered on independence. To enable Gaussian variational inference, simplex-constrained parameters were reparametrized using a logistic transform. Specifically, the initial state probabilities $\pi$and the latent state indicators $S_{i,t}$were represented through unconstrained logit variables, which were assigned multivariate normal distributions. This yields a logistic–normal variational approximation while preserving the original Dirichlet and multinomial model structure.

$$\boldsymbol{\pi} \sim Dirichlet\ (\boldsymbol{a_k})\ ,$$

where, $\boldsymbol{a_k} = \left(\frac{\boldsymbol{exp}\ (\lambda_k)}{\mathbf{1}+\sum_{l=1}^{K-1}\boldsymbol{exp}\ (\lambda_k)}\right);\ \boldsymbol{a_K} = \frac{\mathbf{1}}{\mathbf{1}+\sum_{l=1}^{K-1}\boldsymbol{exp}\ (\lambda_k)},\ \boldsymbol{\lambda} \sim N(\boldsymbol{\mu_\pi}, \boldsymbol{\Sigma_\pi})$

$\boldsymbol{S_{i,1,1}}$ is indicator random vector defined in methods section. Then

$$\boldsymbol{S}_{i,t} \sim \text{Multinomial}\left(1, \boldsymbol{c}_{i,t}\right), \quad \boldsymbol{c}_{i,t} = \left(c_{i,t,1} \cdots, c_{i,t,k}\right).$$

Here, $\boldsymbol{c}_{i,t} = \frac{\boldsymbol{exp}\ (\boldsymbol{\Gamma_{itk}})}{\sum_k \boldsymbol{exp}\ (\boldsymbol{\Gamma_{itk}})}\ ;\ \boldsymbol{\Gamma_{itk}} \sim N(\boldsymbol{\mu_S}, \boldsymbol{\Sigma_S})$

***The prior distribution as "initial variational distribution" for model parameters***

$$p(\boldsymbol{\Theta}) = q_0(\boldsymbol{\Theta}) = N\left(\begin{bmatrix}\boldsymbol{\mu_\alpha}\\ \boldsymbol{\mu_\beta}\\ \boldsymbol{\mu_\xi}\\ \boldsymbol{\mu_\pi}\\ \boldsymbol{\mu_u}\\ \boldsymbol{\mu_\varepsilon}\\ \boldsymbol{\mu_S}\end{bmatrix}, \text{blockdiag}\left(\boldsymbol{\Sigma_\alpha}, \boldsymbol{\Sigma_\beta}, \boldsymbol{\Sigma_\xi}, \boldsymbol{\Sigma_\pi}, \boldsymbol{\Sigma_u}, \boldsymbol{\Sigma_\varepsilon}, \boldsymbol{\Sigma_S}\right)\right).$$

At each iteration, we used $(i_u, i_S) = 100$ Monte Carlo samples to approximate the gradient and Hessian of $\log p(\boldsymbol{y_i}\,|\boldsymbol{\Theta})$ using Fisher's and Louis' identities defined above. Again, we used 200 Monte Carlo samples to approximate the expectations with respect to $q_{i-1}(\boldsymbol{\Theta})$ in the R-VGADM updates of the mean and precision matrix. Fisher blocks are computed only over admissible state paths $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}(\boldsymbol{M})$.

**(S2) Appendix B: Additional Simulation analysis results**

**Supplementary Table B1:** Results from the Sequential Variational Bayes (SVB) fit of a six-state CT-mMixHMM to the 30,000-observation (N=1000, T= 30) simulated dataset.

| | | No. of initial states | No. of states found | Estimated posterior means | Estimated posterior st. error. | DIC |
|---|---|---|---|---|---|---|
| **Dependent Variable 1** | N=1000, T=30,3% missingness | 15 | 8 | 111.1,121.0,132.1, 141.5,139.1,138.1, 140.5, NA | 3.7,2.5,5.4,9.5, 14.2,3.6,4.5, NA | −1460.9 |
| | | 12 | 6 | 111.0,122.0,132.1, 141.0, 139.0, NA | 3.4, 2.4, 5.6, 9.5, 14.1, NA | −1461.1 |
| | | 10 | 6 | 111.1,121.8,131.8, 141.2, 139.8, NA | 3.2,2.1,5.2,11.1, 15.0, NA | −1462.6 |
| | | 5 | 6 | 111.1,121.3,131.6, 141.2, 139.2, NA | 3.2,2.1,5.2, 11.5, 15.1, NA | −1461.8 |
| **Dependent Variable 2** | N=1000, T=30, 3% missingness | 15 | 8 | 70.1,72.0,82.1,88.0, 82.0,89.3, 84.6, NA | 3.6, 2.4, 4.5, 7.5, 6.4,6.2, 7.1, NA | −1362.7 |
| | | 12 | 6 | 70.1,72.0,82.0, 88.0, 82.1, NA | 3.5, 2.5, 4.5, 7.4, 6.4, NA | −1364.3 |
| | | 10 | 6 | 71.2,71.6,82.4, 86.2, 81.4, NA | 3.3, 2.7, 2.5, 8.1, 6.1, NA | −1365.8 |
| | | 5 | 6 | 71.3,71.6,82.2, 86.1, 81.1, NA | 3.4, 2.8, 2.7, 8.1, 6.1, NA | −1361.4 |

Note: DIC: Deviance Information Criteria6

**Supplementary Table B2:** Bias and Mean Absolute Bias of SVB and MCMC Relative to the MFUS

| *Transition Category* | *Transitions* | *True value* | *Bias (SVB)* | *Bias (MCMC)* | *Mean Absolute Bias (SVB)* | *Mean Absolute Bias (MCMC)* |
|---|---|---|---|---|---|---|
| *Persistent state* | state (1,1) | 0.601 | 0 | 0.15 | **0.081** | 0.103 |
| | state (2,2) | 0.591 | -0.181 | 0.176 | | |
| | state (3,3) | 0.635 | 0 | 0.148 | | |
| | state (4,4) | 0.732 | -0.288 | 0.08 | | |
| | state (5,5) | 0.846 | 0.019 | -0.064 | | |
| | state (6,6) | 1 | 0 | 0 | | |
| *move to poor state* | state (1,2) | 0.121 | 0.025 | -0.11 | **0.020** | 0.026 |
| | state(1,3) | 0.169 | 0.001 | 0.006 | | |
| | state(1,4) | 0.065 | 0 | -0.041 | | |
| | state(1,5) | 0.018 | -0.006 | -0.008 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | state(1,6) | 0.026 | -0.02 | -0.006 | | |
| | state(2,3) | 0.17 | 0.075 | 0.007 | | |
| | state(2,4) | 0.038 | 0 | -0.014 | | |
| | state(2,5) | 0.004 | 0.04 | 0.005 | | |
| | state(2,6) | 0.016 | 0.002 | -0.006 | | |
| | state(3,4) | 0.083 | 0.055 | 0.043 | | |
| | state(3,5) | 0.014 | 0.005 | -0.005 | | |
| | state(3,6) | 0.06 | -0.001 | -0.052 | | |
| | state(4,5) | 0.11 | -0.041 | -0.023 | | |
| | state(4,6) | 0.049 | -0.006 | -0.021 | | |
| | state(5,6) | 0.154 | -0.019 | -0.048 | | |
| ***move to better state*** | state (2,1) | 0.181 | 0.064 | -0.169 | **0.027** | 0.030 |
| | state(3,1) | 0.106 | 0 | -0.099 | | |
| | state(3,2) | 0.102 | 0 | -0.035 | | |
| | state(4,1) | 0.02 | 0.009 | -0.006 | | |
| | state(4,2) | 0.02 | 0.023 | -0.007 | | |
| | state(4,3) | 0.069 | 0.304 | -0.022 | | |
| | state(5,1) | 0 | 0 | 0.024 | | |
| | state(5,2) | 0 | 0 | 0.012 | | |
| | state(5,3) | 0 | 0 | 0.021 | | |
| | state(5,4) | 0 | 0 | 0.05 | | |

**Supplementary Table B3** Point estimates MAP (medians) for the group-level and subject level parameter estimates for N=1000, T=30 simulated data (with 250 runs) using SVB for six state CT-mMixHMM.

| Variable | Parameter | True value | MAP (SD) | 95% CCI |
|---|---|---|---|---|
| **DV 1 – State 1** | $\alpha_{01}$ (Intercept) | 111.2 | 112.3 (1.0) | [110.1, 114.6] |
| | $\beta_{11}$ (slope) | 0.12 | 0.1 (0.04) | [0.04, 0.16] |
| | $\sigma^2_{within}$(variance) | 8.10 | 8.6 (0.8) | [7.3, 10.1] |
| **DV 1 – State 2** | $\alpha_{02}$ (Intercept) | 121.9 | 125.7 (1.0) | [123.4, 127.9] |
| | $\beta_{12}$ (slope) | 0.24 | 0.22 (0.04) | [0.13, 0.31] |
| | $\sigma^2_{within}$(variance) | 10.0 | 10.9 (0.8) | [9.2, 12.8] |
| **DV 1 – Stage 3** | $\alpha_{03}$ (Intercept) | 131.8 | 136.8 (1.0) | [134.1, 139.6] |
| | $\beta_{13}$ (slope) | 0.32 | 0.33 (0.04) | [0.21, 0.45] |
| | $\sigma^2_{within}$(variance) | 10.2 | 12.5 (0.8) | [10.5, 14.9] |

| | | | | |
|---|---|---|---|---|
| **DV 1 – State 4** | $\alpha_{04}$ (Intercept) | 141.2 | 148.9 (1.0) | [146.0, 151.9] |
| | $\beta_{14}$ (slope) | 0.45 | 0.47 (0.04) | [0.33, 0.61] |
| | $\sigma^2_{within}$(variance) | 12.5 | 15.8 (0.8) | [13.2, 18.6] |
| **DV 1 – State 5** | $\alpha_{05}$ (Intercept) | 139.9 | 141.4 (1.0) | [137.8, 165.2] |
| | $\beta_{15}$ (slope) | 0.60 | 0.62 (0.04) | [0.45, 0.78] |
| | $\sigma^2_{within}$(variance) | 14.3 | 18.2 (0.8) | [15.1, 21.9] |
| **Random Effects – DV 1** | $\tau_1^2$(Intercept) | 15.21 | 20.4 (2.9) | [15.6, 26.4] |
| **DV 2 – State 1** | $\alpha_{01}$ (Intercept) | 71.2 | 70.5 (0.9) | [69.1, 72.0] |
| | $\beta_{11}$ (slope) | 0.02 | 0.04 (0.03) | [0.01, 0.08] |
| | $\sigma^2_{within}$(variance) | 5.0 | 5.9 (0.6) | [4.8, 7.1] |
| **DV 2 – State 2** | $\alpha_{02}$ (Intercept) | 71.6 | 78.2 (0.9) | [76.5, 80.0] |
| | $\beta_{12}$ (slope) | 0.12 | 0.1 (0.03) | [0.04, 0.17] |
| | $\sigma^2_{within}$(variance) | 6.0 | 6.7 (0.6) | [5.5, 8.1] |
| **DV 2 – State 3** | $\alpha_{03}$ (Intercept) | 82.4 | 85.6 (0.9) | [83.7, 87.5] |
| | $\beta_{13}$ (slope) | 0.25 | 0.2 (0.03) | [0.11, 0.29] |
| | $\sigma^2_{within}$(variance) | 7.0 | 7.8 (0.6) | [6.3, 9.5] |

## Continuation of Supplementary table B3

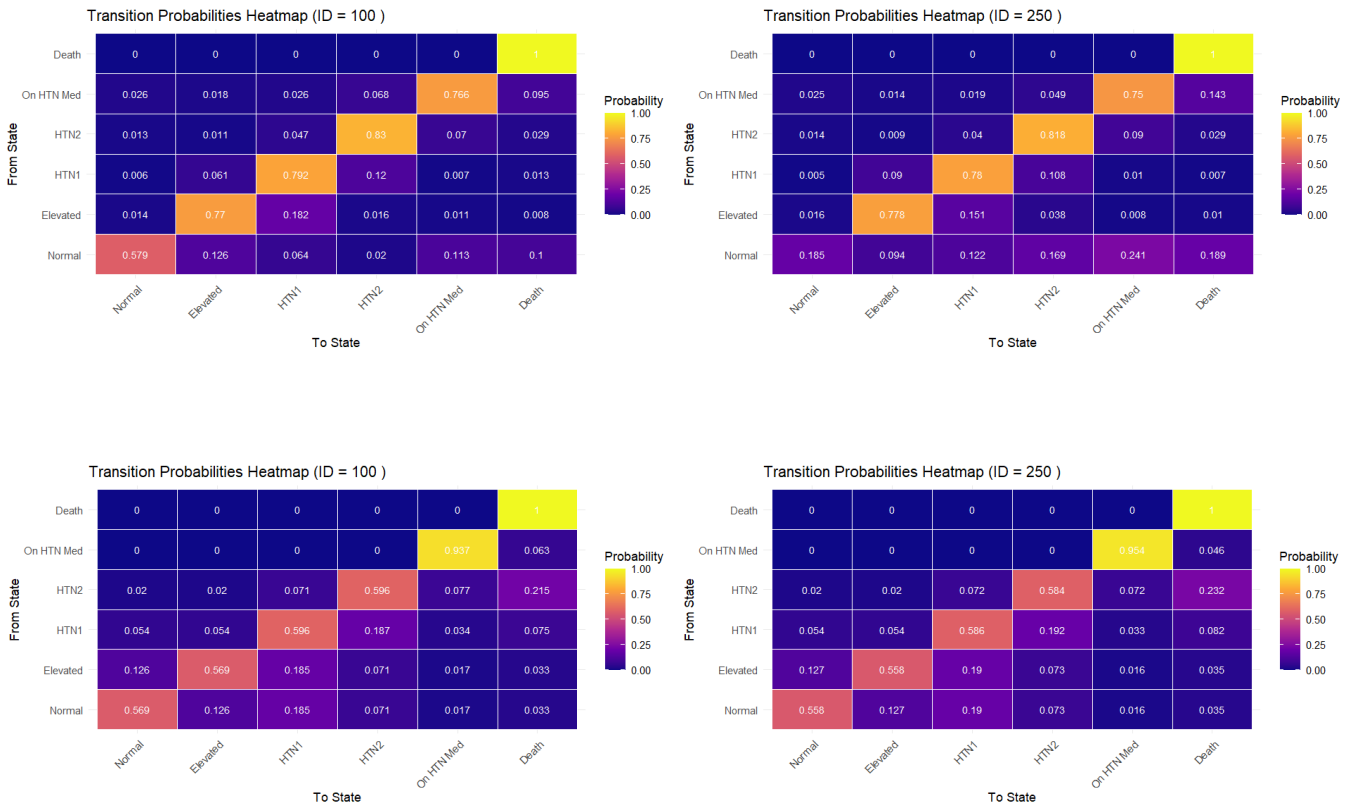
| Variable | Parameter | True value | MAP (SD) | 95% CCI |
|---|---|---|---|---|
| **DV 2 – State 4** | $\alpha_{04}$ (Intercept) | 86.1 | 91.8 (0.9) | [89.6, 94.0] |
| | $\beta_{14}$ (slope) | 0.24 | 0.28 (0.03) | [0.17, 0.39] |
| | $\sigma^2_{within}$(variance) | 7.0 | 9.4 (0.6) | [7.5, 11.4] |
| **DV 2 – State 5** | $\alpha_{05}$ (Intercept) | 81.3 | 98.6 (0.9) | [96.2, 101.1] |
| | $\beta_{15}$ (slope) | 0.35 | 0.37 (0.03) | [0.25, 0.50] |
| | $\sigma^2_{within}$(variance) | 10.0 | 10.7 (0.6) | [8.7, 12.9] |
| **Random Effects – DV 2** | $\tau_2^2$(Intercept) | 10.36 | 11.6 (1.9) | [8.2, 15.4] |
| **Random Effects-Covariance** | $\tau_{12}$ (Covariance) | 0.50 | 0.32 (0.13) | [0.20, 0.43] |
| **TPM Overall, Intercepts (Logit Scale)** | $\xi_{0,12}$ (1→2) | -1.10 | -1.15 (0.17) | [-1.47, -0.85] |

| | | | | |
|---|---|---|---|---|
| | $\xi_{0,13}$ (1→3) | -2.00 | -2.38 (0.26) | [-2.90, -1.93] |
| | $\xi_{0,23}$ (2→3) | -1.19 | -1.22 (0.19) | [-1.64, -0.87] |
| | $\xi_{0,34}$ (3→4) | -1.01 | -1.05 (0.21) | [-1.48, -0.65] |
| | $\xi_{0,45}$ (4→5) | -0.84 | -0.85 (0.18) | [-1.21, -0.52] |
| | $\xi_{0,41}$ (4→1) | -3.11 | -3.15 (0.33) | [-3.79, -2.53] |
| **TPM Overall, Slope (Logit Scale)** | $\xi_{1,12}$ (1→2) | 1.00 | 1.11 (0.15) | [1.08, 1.20] |
| | $\xi_{1,13}$ (1→3) | 2.00 | 2.35 (0.23) | [2.30, 2.40] |
| | $\xi_{1,23}$ (2→3) | 1.00 | 1.21 (0.17) | [1.20, 1.26] |
| | $\xi_{1,34}$ (3→4) | 0.50 | 1.05 (0.22) | [1.03, 1.09] |
| | $\xi_{1,45}$ (4→5) | 2.00 | 1.85 (0.16) | [1.80, 1.89] |
| | $\xi_{1,41}$ (4→1) | 1.00 | 1.15 (0.32) | [1.12, 1.16] |
| **TPM Random Effects (Subject level)** | $\tau^2_{int,(TPM)}$(Intercept) | 0.65 | 0.68 (0.11) | [0.48, 0.94] |
| **Initial Probabilities** | $\pi_1$ | 0.50 | 0.48 (0.33) | [0.42, 0.44] |
| | $\pi_2$ | 0.30 | 0.35 (0.31) | [0.32, 0.36] |
| | $\pi_3$ | 0.10 | 0.09 (0.18) | [0.08, 0.10] |
| | $\pi_4$ | 0.05 | 0.04 (0.16) | [0.01, 0.05] |
| | $\pi_5$ | 0.05 | 0.04 (0.34) | [0.03, 0.10] |
| **State Missingness** | $\boldsymbol{p}(M = \mathbf{1} \mid \boldsymbol{Z} = \mathbf{6})$ | 1.00 | 0.994 (0.08) | [0.975,0.999] |
| **Marginal miss prop** | $\boldsymbol{p}(M = \mathbf{1})$ | 0.03 | 0.029 (0.10) | [0.02,0.03] |

Note: CCI: Central Credible Interval

## (S3) Appendix C: Additional empirical analysis results

### Individual Transition probabilities and absorbing states




**Supplementary Figure 4-0-1** TPM from MCMC approach (upper panel) and SVB (lower panel)

From the upper panel of supplementary figure 4-0-1, it is observed that MCMC can capture death as a full absorbing state however it has failed to detect "On HTN Med" as a partial absorbing state as we can see that "On HTN Med" row contains some values to states Normal, Elevated, HTN1 and HTN2 that did not capture the constrained nature of the model, Whereas, from the lower panel, it is observed that SVB can capture accurately both death as a full absorbing state and "On HTN Med" as a partial absorbing state as we can see that "On HTN Med" row contains exactly zero values to states Normal, Elevated, HTN1 and HTN2 which supports the real-world scenario.

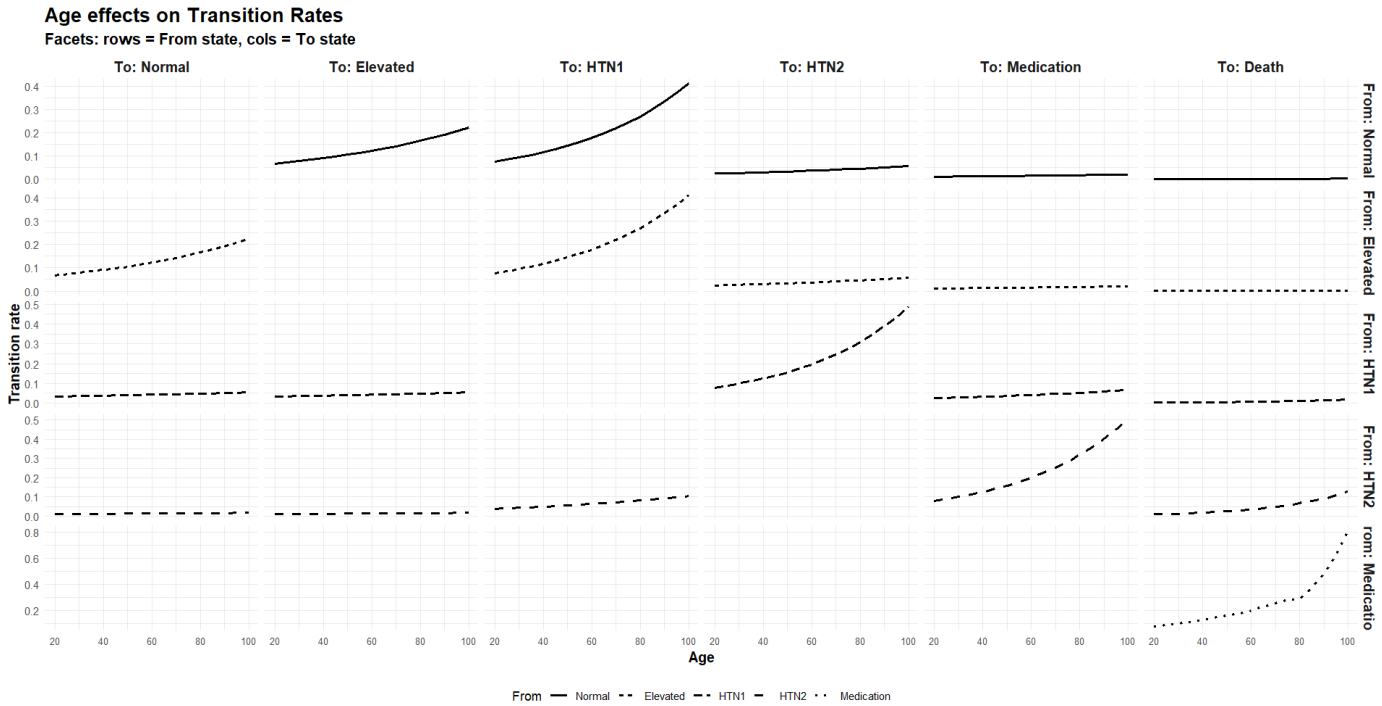
The SVB approach provides superior state identification because it more effectively constrains the latent transition structure through its optimization-based posterior estimation, reducing the sampling noise that can obscure absorbing or semi-absorbing dynamics under MCMC. The deterministic updates in SVB enhance numerical stability and better preserve low-probability transitions, enabling clearer distinction

between reversible and irreversible disease states.

Clinically, this improved state separation has meaningful implications: SVB more accurately reflects real-world hypertension progression, where treatment initiation (*On HTN Med*) typically represents a semi-permanent management phase preceding mortality (*Death*). This distinction helps clinicians interpret model outputs as evidence of treatment adherence, disease stabilization, and survival risk, thereby supporting more reliable longitudinal monitoring and policy evaluation.

Based on the accurate identification of the nature of the partial and full absorbing states, SVB is considered as superior method compared to MCMC. The next section provides the overall and individual age effect on the transition rates of the BP states for MFUS participants based on SVB method.

Participants in the Normal state showed increasing transition rates with age toward Elevated and HTN1, indicating age-related progression from healthy to higher-risk categories. However, transition rates from Normal directly to HTN2, Medication, or Death remained low and largely stable across the lifespan. Transition rates from normal to hypertensive states increase with age, emphasizing the importance of early detection.




**Supplementary Figure 4-0-2** Overall age effects on transition rates for all (N=1007) MFUS participants.

Supplementary Figure 4-0-2 presents the estimated effects of age on transition rates between blood pressure–related states, stratified by originating ("From") and destination ("To") states. Several key patterns were observed.

Individuals in the Elevated state demonstrated accelerated transitions to HTN1 and HTN2, with rates rising steadily as age increased. Elevated individuals also displayed moderate increases in transition to *Medication*, suggesting that older adults with elevated blood pressure were more likely to initiate treatment compared to younger individuals.

Transitions from HTN1 to HTN2 and Medication increased sharply with age, underscoring the higher risk of progression and treatment initiation once hypertension is established. Notably, transition to *Death* from HTN1 remained low, though some age-related increases were evident. For those in the HTN2 state, transitions to *Medication* and *Death* rose markedly with age. In particular, the risk of death from HTN2 increased sharply after midlife, suggesting that uncontrolled progression substantially contributes

to mortality risk in older adults. Moreover, transition of Medication → Death showed pronounced increase, especially after ~80. Even under treatment, mortality risk rises sharply with age, likely reflecting underlying severity and comorbidity rather than treatment failure per se.

Overall, the aging effects highlight a stepwise progression from Normal through Elevated, HTN1, and HTN2, with increasing likelihood of treatment initiation *(Medication)* and mortality *(Death)* at later stages. These results underline the importance of early detection and management of elevated blood pressure to delay or prevent progression to advanced stages associated with higher morbidity and mortality risks.

*Aging effects on transition rate for two specific case study*

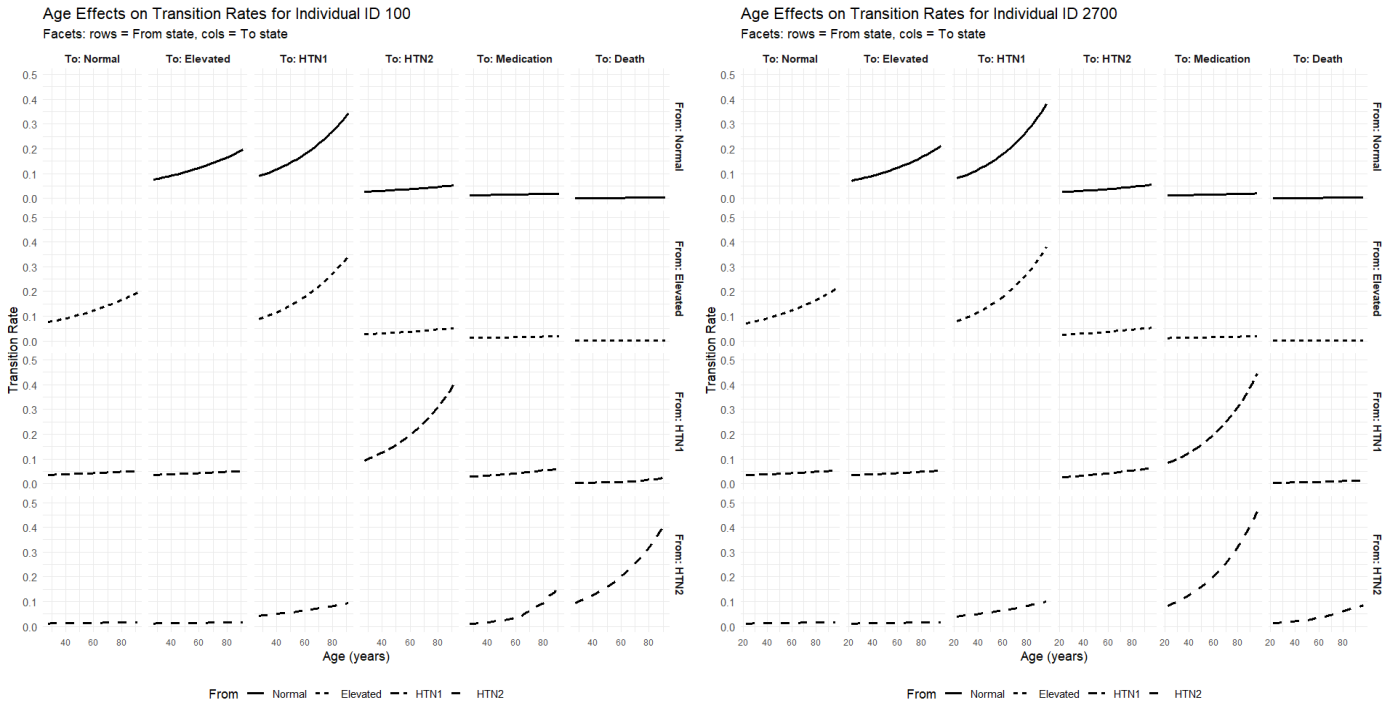
Supplementary Figure 4-0-3 demonstrated how the model dynamically reflects individualized blood pressure transition patterns across age, capturing both gradual and accelerated shifts toward medication use and mortality risk. These trajectories highlight the model's capacity to represent real-world heterogeneity in patient pathways, distinguishing between stable profiles (e.g., Individual 2) and high-risk progression (e.g., Individual 1), which is critical for personalized prediction and intervention planning.

*Interpretation of Aging Effects for Individual 1 (ID: 100) (Died)*

The curve for individual 1 (died) shows that Highest Rate occurs: HTN1 to HTN and Second Highest Rate occurs: HTN2 to Death. This is a classic and logical profile for someone who, unfortunately, passed away due to complications from hypertension. Let's break down why these two high rates tell the story of their health outcome. This combination of high rates creates a dangerous "one-two punch" or a vicious cycle:

High HTN1 to HTN2 Rate (the accelerator), this means that once Individual 1 developed Stage 1 Hypertension (HTN1), their condition was very aggressive and likely untreated. Instead of staying stable, getting on medication, or even improving, their blood pressure rapidly worsened to the more severe Stage 2 (HTN2). Taken together, these patterns suggest that aging not only increases the likelihood of disease

progression but also amplifies the pace of transition at more advanced stages of hypertension.




**Supplementary Figure 4-0-3** Effects of age on BP transition rates for individual 1(ID:100) (died) and individual 2 (ID:2700) (alive).

High HTN2 to Death Rate (the critical danger), this means that being in the HTN2 state was extremely dangerous for Individual 1. The risk of fatal events like heart attacks, strokes, or heart failure was very high. Because the first rate (HTN1 to HTN2) was also high, they were pushed into this high-danger state very quickly and spent a significant amount of time in it, greatly increasing their overall risk of death. The data curves tell a clear story, Individual 1's hypertension progressed very rapidly in a year from bad to worse (HTN1 → HTN2), and once it was severe (HTN2), their personal risk of death was exceptionally high. The combination of fast progression and high mortality in the severe state led to their death.

*Interpretation of Aging Effects for Individual 2 (ID: 2700) (Alive)*

In case of beneficial Transitions (Medication Initiation) such as HTN1 to Medication & HTN2 to

Medication, for Individual 2, these rates are likely meaningfully high and increase with age. This is crucial. It indicated that as their condition worsened or as they got older (and regular doctor visits became more common), they consistently started necessary treatment. Medication controls blood pressure, drastically reducing the risk of progression and death. Moreover, in case of disease progression transitions (the risky ones) such as Normal to Elevated, Elevated to HTN1, HTN1 to HTN2: These rates undoubtedly increase with age. However, for Individual 3, the *absolute value* of these rates-especially the progression to severe states (HTN1 to HTN2)-is likely moderate compared to Individual 1. The slopes might be less steep, indicating a slower, more manageable progression of the disease over their lifetime.

Finally, in case of the Most Critical Transition such as towards Death for instance HTN1 to Death & HTN2 to Death, while these risk rates will also increase with age (as with anyone), the key is that their absolute values are low across all ages, especially when compared to Individual 1.

Overall, Individual 2's observed trajectories of BP shows that while their risk of hypertension progression naturally increased with age, they also had a strong tendency to initiate medication. This treatment effectively managed their condition, keeping them from prolonged periods in the high-risk HTN2 state. Consequently, their age-related risk of death from hypertension, while present, remained managed and low enough throughout their life to ensure survival.

# CHAPTER 5: SUMMARY

## 5.1 Summary of research findings

The purpose of this thesis was to develop and evaluate estimation methods to address the computational challenges in both discrete-time and continuous-time models for ILD. The objectives were: (1) to develop and evaluate an algorithm for estimating parameters of ML-TAR models in discrete-time ILD and compare its performance with existing methods; and (2) to construct and evaluate an algorithm for estimating parameters of CT-mMixHMM in continuous-time ILD and compare its performance with existing methods.

Two main reasons explain the relative sparse citation of previous literature in the discussion of the present results. First, regarding the algorithms themselves: the method proposed in this study was developed based on the mean field variational algorithm for the first part [14] and on a sequential variational algorithm [36] for the second part of this thesis. [14] did not conduct analyses with the ML-TAR model and [36] did not evaluate the CT-mMixHMM, and therefore there are no published empirical investigations of the proposed algorithms for ML-TAR model and CT-mMixHMM. Consequently, it is difficult to make comparisons through cross-referencing with existing literature. Second, the both ML-TAR and CT-mMixHMM are a relatively new model. Except for the research by [12] and [35], existing studies focusing on the MT-TAR and CT-mMixHMM respectively remain scarce, making direct comparisons of research results challenging. However, this thesis references pertinent literature in relation to the research findings on the accuracy and computational efficiency of the proposed algorithms compared with traditional algorithms.

In the first part, this thesis proposed an optimization-based approximate Bayesian estimation method for ML-TAR parameters (MFVB), which was designed for regular-time observations and non-linear links between the outcome process and covariates in ILD. The performance of MFVB was compared

with MCMC using simulated data and real-world outcomes (e.g., negative affect scores).

Across varying sample sizes ($M$) and occasions ($T$), MFVB and MCMC achieved similar coverage near 95% as $T$ increased from 30 to 150 and $M$ from 50 to 150, indicating more reliable parameter estimates with more data. MCMC tended to have better coverage at early time points, while MFVB improved with larger samples, becoming comparable to MCMC at later stages. Parameter-specific accuracy varied by condition. Generally, increasing $M$ and $T$ improved accuracy for most parameters, with some parameters (e.g., certain variance components) showing robust accuracy even at smaller samples, while others required larger samples for similar precision. Consistent with previous research findings, the performance of the MCMC and MFVB methods were similar for all simulation conditions [12,14].

Computational efficiency favored MFVB: in a scenario with 150 individuals and 150 time points, MCMC required over six hours, whereas MFVB completed in under 40 minutes on a standard 32 GB RAM, single-core 4.7 GHz system. For larger models (e.g., $M$ = 150, $T$=150), MFVB remained substantially faster (MFVB ≈ 21× faster than MCMC). These findings on the computational efficiency comparison between MFVB and MCMC in case of large models, are consistent with previous research [14].

*Key-findings from real-world data application*

Our empirical analysis on “Physical Activity and Affect Regulation” dataset using the ML-TAR model showed that regulatory strength in affect varies with intensity within individuals, supporting our theoretical expectations and demonstrating the model's relevance. Moreover, the MFVB and MCMC methods showed comparable performance in the analysis of empirical dataset. The ML-TAR model is adaptable to various research questions. Although we examined the effect of physical activity on the threshold variable, the model can also include factors such as daily learning goal attainment, age, and sex.

In the second part, this thesis introduced an optimization-based approximate Bayesian method (Sequential Variational Bayes, SVB) to estimate CT-mixHMM parameters, addressing irregular observation times and dynamic transition among outcomes and covariates in ILD. Its accuracy was evaluated in determining the true number of hidden states (including partial and full absorbing states) and in improving computational speed, using simulated data and real outcomes (SBP/DBP) from irregular visits (Manitoba Follow-Up Study).

Bias, coverage rate, convergence and MAP of the model parameters were estimated under varied individual sizes ($N$=200, 500, 1000), number of time points ($T$=20, 30, 40, 50), and marginal missing data proportions (2%-4%) and various fixed effect sizes in the simulation study. SVB was initialized with a large number of hidden states to recover the true state structure. Results indicated no label switching and stable convergence, with the Evidence Lower Bound (ELBO) plateauing after roughly 120–150 iterations, indicating computational efficiency and reproducibility. SVB reliably recovered initial state probabilities and absorbing-state structure, with wider posterior uncertainty for rare states as expected. Model comparison using Deviance Information Criterion (DIC) favored a six-state solution over five. Larger datasets benefited from fewer initial states to avoid superfluous state allocations. These findings on the detection of optimal number of hidden states with the help of DIC are consistent with previous study [15]. The six-state model aligned with clinically interpretable states such as Normal, Elevated, Hypertension Stage 1 (HTN1), Hypertension Stage 2 (HTN2), On Hypertension Medication (On HTN Med), and Death (an absorbing state).

Transition probability matrices (TPM) estimated by SVB method closely matched true simulated TPMs and outperformed MCMC in capturing absorbing states. SVB method accurately identified Death as a full absorbing state and On HTN Med as a partial absorbing state, whereas MCMC showed inflated posterior variances and failed to enforce absorbing constraints properly. This makes VB methods superior

for modeling transition probabilities in the presence of absorbing states which is consistent with previous studies [15.16].

*Key-findings from real world data applications*

The model's emission distributions for systolic (SBP) and diastolic blood pressure (DBP) showed clear stepwise increases across states, consistent with clinical hypertension categories. Mean SBP ranged from approximately 112 mmHg in Normal to 140 mmHg in HTN2, with DBP increasing similarly. The On Medication state exhibited intermediate means with greater variability, reflecting treatment heterogeneity. The Death state had placeholder missing values (NA) due to lack of physiological measurements. Density plots revealed clustering of Normal and Elevated states at lower pressures, progressive shifts to higher pressures in HTN1 and HTN2, and overlapping distributions in the medication state, indicating clinical complexity in treatment effects which supports the findings of previous research [17].

Age effects on transition rates revealed a stepwise progression from Normal to Elevated, HTN1, HTN2, and then Medication and Death states, with transition rates generally increasing with age for empirical MFUS data application. Forward transitions were more probable than backward ones, consistent with chronic hypertension progression for MFUS participants. Transition rates to severe states and mortality rose markedly with age, highlighting the importance of early detection and management. These findings of age effects on transition rates in different BP states are consistent with previous studies [18,19].

Taken together, both the simulation and empirical results indicated that the SVB provides a scalable and numerically stable approach for joint estimation of multivariate longitudinal processes, latent states, covariate-driven transitions, and informative missingness within a single variational objective. Although variance estimates exhibit mild conservatism, as expected under variational inference, the

framework preserves structural interpretability and offers substantial computational advantages over fully Bayesian MCMC, making it well suited for large-scale longitudinal analyses that are consistent with previous studies [14 -16].

Table 5.1 summarizes how variational Bayes and MCMC methods were assessed based on the study objectives and simulation conditions. MCMC method considered in this research can be implemented in R software, whereas MFVB and R-VGADM method do not have readily available, standard implementations in R, underscoring practical considerations for researchers choosing an estimation approach.

In a nutshell, the thesis demonstrated that the proposed SVB-based CT-mMixHMM and MFVB-based ML-TAR estimation frameworks deliver scalable, accurate, and computationally efficient tools for analyzing high-dimensional, irregularly sampled ILD. They enable robust inference about latent state dynamics, transition structures, and the impact of covariates, with clear implications for application in clinical monitoring and risk stratification in longitudinal health data.

**Table 5.1** Summary of performance of variational Bayes methods across study objectives

| **Study Objective** | **Methods/ Algorithms** | **Sample Size (N)** | **% missing responses due to dropouts** | **Performance** | **Software** |
|---|---|---|---|---|---|
| #1: To develop and assess an algorithm for estimating parameters in ML-TAR models applied to discrete-time ILD, aiming to resolve computational challenges and compare its effectiveness with current estimation techniques; | MCMC<br>MFVB | 1500 (M=50, T=30) to 22,500 (M=150, T=150) | Not Applicable | i. The MFVB algorithm showed comparable performance (in terms of coverage rates, bias, precision and computational speed) to the MCMC method across simulation conditions.<br>ii. The MFVB outperformed MCMC for most of the simulation conditions, especially when the number of individuals and number of time points was large (say M=150, T=150).<br>iii. For larger models (e.g., M = 150), MFVB remained | MCMC can be implemented in R studio packages. However, MFVB cannot be readily available in RStudio. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | substantially faster (MFVB ≈ 21× faster than MCMC). | |
| #2: To construct and evaluate an algorithm for estimating parameters of CT-mMixHMM in continuous-time ILD to address the computational issues and compare its performance with existing estimation methods. | MCMC SVB(R-VGADM) | 4000 (M=200, T=20) to 50,000 (M=1000, T=50) | 2% to 4% | i. The R-VGADM algorithm showed comparable performance (in terms of coverage rates, bias, precision in capturing partial and full absorbing states and computational speed) to the MCMC method across simulation conditions.<br>ii. The R-VGADM outperformed MCMC for most of the simulation conditions, especially when the number of individuals and number of time points was large (say M=1000, T=30).<br>iii. The R-VGADM algorithm effectively recognized Death as | MCMC can be implemented in R studio packages. However, SVB cannot be readily available in RStudio. |

| | | |
|---|---|---|
| | | a fully absorbing state and On HTN Med as a partially absorbing state. In contrast, the MCMC approach exhibited increased posterior variances and did not adequately enforce absorbing constraints. |

Note: Mean-field variational Bayes (MFVB); Markov Chain Monte Carlo (MCMC); Sequential variational Bayes (SVB); Recursive Variational Gaussian Approximation for Dependent Mixture (R-VGADM), *N*: Total sample size, *M*: Number of individuals, *T*: Number of time points

## 5.2 Strengths and Limitations

Although variational inference has been widely studied in machine learning and Bayesian statistics, most existing methods focus on standard latent variable models, discrete-time hidden Markov models, or generalized linear mixed models. Relatively few works address variational inference for intensive longitudinal settings that involve, simultaneously, nonlinear threshold dynamics, continuous-time latent-state evolution, mixed-effects structures, irregular observation intervals, and informative missingness. The present study extends the variational inference literature by developing computationally scalable Bayesian estimation procedures tailored specifically for these complex longitudinal settings. The proposed methods combine continuous-time latent-state modeling, random effects, multivariate outcomes, and state-dependent missingness within unified variational Bayesian frameworks.

This study provides an original contribution to the literature on statistical estimation methods for ILD models by introducing and assessing two approximate Bayesian estimation techniques: Mean Field Variational Bayes (MFVB) and Sequential Variational Bayes (SVB). These approaches are proposed as alternatives to resampling-based methods to address computational challenges encountered during the estimation process. Existing methods such as Markov Chain Monte Carlo (MCMC) and the Laplace approximation often require considerable computational resources or rely on crude approximations that assume a normal distribution for all posteriors [9,10]. MCMC methods often require iterative sampling, leading to time-intensive computations for high-dimensional ILD due to its iterative sampling [12]. By contrast, MFVB and SVB provide efficient alternatives within an optimization framework, delivering substantial speedups while maintaining good accuracy for scalable inference in large-scale data applications [14-16].

The integration of simulated and real-world data to evaluate the proposed methods alongside existing approaches enhances the rigor of this research. A wide variety of data conditions including sample sizes, percentage of missingness due to state, initial number of states and design effects were investigated through the simulation study. This study demonstrated the practical utility of intensive longitudinal datasets drawn from both psychometric and healthcare domains, thereby enhancing the generalizability of our proposed methodologies compared to existing studies on ILD [3, 4, 12, 13, 20, 21]. Practical examples utilized two datasets: the Physical Activity and Affect Regulation study [1] and the Manitoba Follow Up Study [2], both of which offer substantial amounts of ILD. Leveraging the state-dependent missingness in the likelihood function of CT-mMixHMMs while defining the objective functions for our proposed R-VGADM algorithm can reduce bias and improve the precision of the model parameter estimates relative to existing frequentist methods applied to state-dependent missingness [22, 23].

Nonetheless, this research is subject to certain limitations. First, the MFVB can struggle to accurately estimate covariance matrices which align with the findings of [14], though linear response variational Bayes method can correct some variance underestimation [8]; nonetheless, full posterior uncertainty estimation remains limited. Second, the success of Bayesian inference relies on how accurately the analytical model represents the data-generating process. By replicating the data generation process for the simulation study using real-world non-linear and dynamic data structure, including random effects, covariate impacts on transition rates and setting constraints a priori in the transition probability space for specific states, this research aimed to minimize the risk of model misspecification. A data-driven method to test functional form and interaction could be used to reduce overfitting and theoretical misspecification [24]. Third, discrete-time (DT) and continuous-time (CT) models commonly used for ILD rely on simplifying

assumptions such as stationarity, Markovian dynamics, and homogeneous transition probabilities [25, 26]. These assumptions may be violated in real-world ILD, potentially biasing parameter estimates and limiting substantive interpretation [27, 28]. Prior work emphasizes the importance of conducting extensive sensitivity analyses under alternative modeling assumptions. Hence sensitivity analysis could be used to assess robustness and diagnose potential misspecification [29, 30].

Fourth, measurement error is a central concern in ILD. If measurement errors are not explicitly modeled, parameter estimates and state transition inferences can be biased. ILD often involves noise or imprecise (e.g., self-reported mood states, physiological data) [31]. In the ML-TAR, the threshold variable is assumed to be free of measurement error, which may be unrealistic for measures prone to error such as emotion and stress. Incorporating explicit measurement-error modeling (e.g., latent-variable or error-in-variables approaches) could be used to address this issue and improve estimation accuracy [32, 33].

Fifth, ILD from self-reported data are also susceptible to biases such as fatigue bias, where participants drop out or provide lower-quality responses due to exhaustion from frequent data collection, and reactivity bias, where participants change their behavior because they know they are being observed. By considering appropriate mitigation strategy of reducing these biases such as collecting more data early, then reducing frequency, and using massive data collection (e.g., wearable instead of manual logs) to reduce respondent burden, it can improve the accuracy of estimates of our model. Furthermore, Bayesian inference relies heavily on prior distributions. The choice of priors (e.g., informative vs. non-informative) could affect results, and sensitivity analysis can overcome this issue to prior selection process [29,30]. Sixth, like other studies, the simulation conditions investigated were not exhaustive. The number of measurement occasions, and number

of individuals were limited. The simulation conditions in this research were consistent with previous studies of DT and CT models for ILD [3, 4, 12, 13, 35]. Furthermore, the real-world data analysis was primarily undertaken to assess the feasibility of implementing the new estimation algorithm, rather than to emphasize previously unidentified variable associations.

Lastly, while the study aims to improve computational efficiency, it may not fully address the variability inherent in real-world ILD, such as instrument reliability and observer bias. Instrument reliability refers to the consistency and stability of a measurement tool, while observer bias - also known as detection or ascertainment bias - occurs when a researcher's expectations influence their observations, potentially leading to systematic data collection errors and unreliable results. Such biases are particularly problematic in observational studies. Addressing these biases through comprehensive methodological approaches such as incorporating measurement-error model and conducting sensitivity analyses can contribute to the robustness and reliability of the research findings.

### 5.3 Future research directions

There are several topics that warrant further consideration. These include: (1) optimal data driven determination of delay parameters for ML-TAR model in ILD, (2) violations of independence assumptions of parameter space for VB, (3) extending the CT-mMixHMM to higher order Markov process and to other types of missingness mechanism, and (4) developing *R package* for R-VGADM method.

In an ML-TAR model, the optimal selection of delay parameter is important as it determines which past values of threshold variable triggers a regime change. While in the present work, the delay parameter governing regime switching in the ML-TAR model was fixed or selected a priori which implies delay uncertainty was ignored, future research could investigate

data-driven method of delay parameters such as Bayesian model averaging [7] over discrete delay values that could allow competing candidates delays and quantify uncertainty in delay selection.

In this work, while developing variational Bayes (VB) algorithms, our target was to minimize the divergence between the true posterior and an approximation in which the parameters and hidden variables were assumed to be independent. This strong assumption allows for an efficient iterative solution, but it can often lead to poor approximations. Future research could investigate the idea of integrating out the parameters induces dependencies which spread over many latent variables, and thus dependency of any two latent variables is very weak making them conditionally independent [34].

Existing mixed-effect HMMs provide hierarchical modeling of longitudinal processes but typically assume first-order Markov dependence [3, 4]. Higher-order latent state models (e.g., weak or higher-order HMMs) demonstrate methods to capture memory beyond first order but are largely developed in non-mixed contexts [5]. While recent study on Bayesian mixed-effect higher-order HMM work [11] directly addresses combining memory beyond first order with random effects, establishing a foundation for further development in multivariate ILD.

Failing to address missing data can greatly undermine the validity of conclusions drawn from intensive longitudinal studies. Currently, there is limited research on mixed-effect hidden Markov models (HMMs) for longitudinal analysis with missing data [23]. This research has concentrated on state-dependent missingness, particularly situations where data were missing not at random (MNAR) in multivariable longitudinal contexts. Moving forward, further research could explore how other types of missingness - such as missing at random (MAR) and missing completely at random (MCAR) - influence the bias and accuracy of model parameter estimates. Another research direction is to incorporate the missingness indicator as a covariate for initial and

transition probabilities, rather than treating missingness as a dependent sequence.

A thorough simulation framework for CT-mMixHMM and ML-TAR models should extend beyond idealized scenarios to systematically assess robustness when assumptions such as stationarity, Markovian dynamics, perfect regime indicators, and homogeneous transitions are violated. Conducting such simulations is crucial for distinguishing substantive regime-based interpretations from artifacts arising due to model assumptions.

Making computer programs or code for algorithms available and easy to access can increase the prominence of research papers. R package including *BAYSTAR* [6], is available to implement the MCMC method for TAR model, but no equivalent package is available for MFVB with multilevel extension. Moreover, currently, there are several software packages available for standard hidden Markov models and some continuous-time hidden Markov models. For example, packages such as: *mHMMbayes [13]*, *msm [37]*, *momentuHMM [38]*, that can handle certain classes of hidden Markov or multistate models using MCMC estimation. However, to our knowledge, there is currently no readily available software that directly implements the exact framework proposed in this thesis, namely a multivariate continuous-time mixed-effects hidden Markov model with random effects, irregular observation times, absorbing states, and state-dependent missingness estimated through variational Bayesian methods. Most existing software packages are limited in one or more ways. For example: some only support discrete-time HMMs, some do not accommodate multilevel or random-effect structures, some do not support multivariate outcomes, and many become computationally intensive when extended to high-dimensional longitudinal settings. Regarding the computational comparison, the purpose was not to compare against a commercial or optimized software package, but rather to compare two estimation paradigms applied to the same statistical model. We fully acknowledge that future work

could develop an R package for MFVB and R-VGADM, enabling researchers to apply these faster approximate Bayesian methods to large ILD datasets. Therefore, developing an accessible R package for the proposed framework is an important direction for future research.

### 5.4 Conclusions and recommendations

This study has evaluated the performance of the proposed estimation methods for discrete- and continuous-time ILD models, with an emphasis on computational efficiency in comparison to existing methods and identify the proposed algorithms as a promising alternative. The ML-TAR and MixHMM models of ILD with traditional Bayesian methods such as MCMC have been applied in research areas such as psychology and healthcare sector [3, 4, 12]; this is the first study to use the approximate Bayesian methods to address the computational issues in ILD models and propose variational Bayes algorithms which incorporate optimization idea rather than resampling process. Overall, the studies in this thesis demonstrate the potential and complexity of evaluating approximate Bayesian algorithms for discrete- and continuous-time models in ILD and provide guidance for selecting methods to address computational demands in ILD modeling. Finally, by translating these advances into scalable tools, the research holds general and flexible framework for analyzing complex longitudinal data with latent structures, with potential applications extending beyond the specific case studies considered in this thesis. These include a wide range of problems in health and social sciences, such as disease progression modeling, behavioral and social dynamics, where capturing time-varying and unobserved heterogeneity is essential.

## 5.5 References


1. L. Flueckiger, R. Lieb, A. H. Meyer, and J. Mata, “How health behaviors relate to academic performance via affect: An intensive longitudinal study,” *PLoS ONE*, vol. 9, no. 10, 2014, doi: 10.1371/journal.pone.0111080.
2. R. B. Tate, T. E. Cuddy, and F. A. L. Mathewson, "Cohort Profile: The Manitoba Follow-up Study (MFUS)," *Int. J. Epidemiol.*, vol. 44, no. 5, pp. 1528-1536, Oct. 2015, doi: 10.1093/ije/dyu141. PMID: 25064641.
3. R. M. Altman, “Mixed hidden Markov models-An extension of the hidden Markov model to the longitudinal data setting,” *Journal of the American Statistical Association*, vol. 102, pp. 201–210, 2007.
4. Inaba, Y., Shimakawa, A., Miyaoka, E., “Mixed Hidden Markov Models for Clinical Research with Discrete Repeated Measurements”, *Am. J. of Th. & App. Stat.* vol. 6, No.6, 2017, 290-296, doi: 10.11648/j.ajtas.20170606.15
5. Xi, X “Further Application of Higher Order Markov Chain and Development in regime-switching model” Electronic Thesis and Dissertation Repository, https:// ir.lib.uwo.ca/etd/678.
6. Chen, C. W. S., Lin, E. M. H., Liu, F., C., and Richard G., “BAYSTAR v0.2-1.0-An R package for threshold autoregressive model with MCMC”, 2022.
7. Hoeting, J., A., Madigan, D., Raftery, A. E., and Volinsky C.T., “Bayesian Model Averaging: A Tutorial”, *Statistical Science*, 1999, vol. 14, No. 4, 382-417.
8. Giordano R., Broderick, T and Jordan, M. “ Linear Response Methods for Accurate Covariance Estimation from Mean Field Variational Bayes” arziv.org/pdf/1506.04088.
9. Spall J. C. “*Introduction to Stochastic Search, Methods, Estimation, and Application”,*

Wiley. 2003, ISBN: 0-471-33052-3

10. Minka *"Expectation Propagation for Approximate Bayesian Inference",* 2013 page 362-369.
11. Liao, Y., Xiang, Y., Zhao, Z., and Ai, D. " Bayesian mixed effect higher order hidden Markov models with applications to predictive healthcare using electronic health record" *IISE Transaction,* 2025, 57:2, 186-198, doi: 10.1080/24725854.2024.2302368.
12. S. De Haan-Rietdijk, J. M. Gottman, C. S. Bergeman, and E. L. Hamaker, "Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation," *Psychometrika*, vol. 81, no. 1, pp. 217–241, 2016, doi: 10.1007/s11336-014-9417-x.
13. Aarts, E and Mildiner, S. " mHMMbayes – An R package for multilevel Hidden Markov Models using Bayesian Estimation, version 1.1.1.
14. D. M. Hughes, M. García-Fiñana, and M.P. Wand. "Fast approximate inference for multivariate longitudinal data," *Biostatistics.* Dec 12;24(1):177-192, 2022. doi: 10.1093/biostatistics/kxab021. PMID: 33991420; PMCID: PMC9748580.
15. C. A. McGrory and D. M. Titterington, "Variational Bayesian Analysis for Hidden Markov Models", *Aust. N.Z. J. Stat.,* 51(2), 227-244, 2009, doi: 10.1111/j.1467-842X.2009.00543.x.
16. P. Jiangwei, Zhang. Boqian and R. Vinayak, "Collapsed variational Bayes for Markov jump process", *NIPS*, 2017, CA. USA.
17. V. A. Shaffer, P. Wegier, K. D. Valentine, J. L. Belden, S. M. Canfield, M. Popescu, L. M. Steege, A. Jain, and R. J. Koopman, "Use of Enhanced Data Visualization to Improve Patient Judgments about Hypertension Control", *Med Decis Making*. 40(6):785-796, 2020. doi: 10.1177/0272989X20940999.

18. T. A. Gowan, M. D. Tringali, J. A. Hostetler, J. Martin, L. I. Ward-Geiger, and J. M. Johnson, “A hidden Markov model for estimating age-specific survival when age and size are uncertain”. *Ecology*. 102(8):e03426, 2021, doi: 10.1002/ecy.3426.
19. J. Przybilla, P. Ahnert, H. Bogatsch, F. Bloos, F. M. Brunkhorst, M. Bauer, M. Loeffler, M. Witzenrath, N. Suttorp, and Scholz M. “Markov State Modelling of Disease Courses and Mortality Risks of Patients with Community-Acquired Pneumonia”, *J Clin Med*. 5;9(2):393, 2020, doi: 10.3390/jcm9020393.
20. J. H. L. Oud and M. J. M. H. Delsing, "Continuous time modeling of panel data by means of SEM," in *Longitudinal research with latent variables*, K. van Montfort, J. H. L. Oud, and A. Satorra, Eds., Springer, 2010, pp. 201–244.
21. M. C. Voelkle, J. H. L. Oud, E. Davidov, and P. Schmidt, "An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia," *Psychological Methods*, vol. 17, no. 2, pp. 176-192, Apr. 2012, doi: 10.1037/a0027543.
22. F. Bortolucci and A. Farcomeni, “A shared-parameter continuous-time hidden Markov and survival model for longitudinal data with informative dropout,” *Statistical Methods in Medical Research*, vol. 38, pp. 1056–1073, 2019.
23. S. Maarten, and V. Ingmar, “State-dependent missingness in hidden Markov models, with an application to drop-out in a clinical trial”, *Psychometrika*, 90, 476-507, 2025, doi:10.1017/psy.2024.3.
24. A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin, “Bayesian Data Analysis” (3rd ed.). 2013, Chapman and Hall/CRC. https://doi.org/10.1201/b16018.

25. O. Ryan, R. M. Kuiper, and E. L. Hamaker, “A continuous-time approach to intensive longitudinal data: What, why, and how?”, *Continuous time modeling in the behavioral and related sciences*. Cham: p. 27–54, 2017.
26. C. Crayen, M. Eid, T. Lischetzke, and J.K. Vermunt, “A continuous-time mixture latent state–trait Markov model for experience sampling data: Application and evaluation”, *Psychol Methods*. 20(3):386–405, 2015.
27. P. Pankowska, BFM. Bakker, D. L. Oberski, and D. Pavlopoulos, “How linkage error affects hidden Markov model estimates: A sensitivity analysis”. *J Surv Stat Methodol*. 8(3):483–512,2020.
28. A. Y. Mitrophanov, A. Lomsadze, and M. Borodovsky, “Sensitivity of hidden Markov models”, *J Appl Probab*. 42(3):632–42,2005.
29. S. Renooij , “Efficient sensitivity analysis in hidden Markov models”, In: *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann; p. 440–7,2001.
30. S. Bonhomme, and M. Weidner, “Minimizing sensitivity to model misspecification”, *Econometrica*. 89(3):1225–60,2021.
31. N. Bolger, and J. P. Laurenceau, “Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research”, New York: Guilford Press; 2013.
32. T. T. L. Chong, H. Chen, TN. Wong, and IKM. Yan, “Estimation and inference of threshold regression models with measurement errors”, Stud Nonlinear Dyn Econom. 22(2):1–27,2017.
33. R. J. Carroll, D. Ruppert, LA. Stefanski, and CM. Crainiceanu, “Measurement Error in Nonlinear Models: A Modern Perspective”. 2nd ed. Boca Raton: CRC Press; 2006.

34. Wang, P., and P. Blunsom, "Collapsed Variational Bayesian Inference for Hidden Markov Models", *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS),* 2013, Scottsdale, AZ, USA.

35. D. J. Raffa and J. A. Dubin, "Multivariate longitudinal data analysis with mixed effect hidden Markov models," *Biometrics*, 2015, doi: 10.1111/biom.12096.

36. A. Bao, D. Gunawan, and A. Z. Mangion, "R-VGAL: A sequential variational Bayes algorithm for generalized linear mixed models," *Statistics and Computing*, vol. 34, Art. no. 110, pp. 1–15, 2024.

37. C. Jackson, "Multi-State Models for Panel Data: The msm Package for R". *Journal of Statistical Software*, 2011, *38*(8), 1–28. https://doi.org/10.18637/jss.v038.i08.

38. B.T., McClintock, and T. Michelot, "momentuHMM: R package for generalized hidden Markov models of animal movement". *Methods in Ecology and Evolution*, 2018, 9(6), 1518-1530.

## Appendix A: Contributions of Authors

For the first paper of this thesis, DJ and AR conceived the study. AR and DJ conducted the analysis and prepared the draft manuscript. All authors reviewed and approved the final version of the manuscript.

For the second paper of this thesis, AR and DJ conceived the study. AR and DJ conducted the analysis and prepared the draft manuscript. All authors reviewed and approved the final version of the manuscript.

## Appendix B: Ethics Approval and Data Access

The presented study findings and interpretations reflect the authors' perspectives. The data was provided to the investigators under specific data-sharing agreements of HREB. The researcher does not own the original source data and, as such, cannot be provided to a public repository. The original source data are available through the submission of appropriate ethics and data access approval forms to the Health Research Ethics Board (HREB) of the University of Manitoba (see https://umanitoba.ca/research/opportunities-support/ethics-compliance for more details).

## Appendix C: No use for Artificial Intelligence Training