Bayesian Dynamic EFA with Hidden Markov Models

Master Thesis (M.Sc.)
in
Quantitative Data Science Methods
Psychometrics, Econometrics and Machine Learning
Faculty of Economics and Social Sciences
at the University of Tübingen
Methods Center

submitted by
Yeong Hwangbo
from Tübingen
(Matriculation number: 6147172)

Submitted in Tübingen
September 30, 2025

1. Supervisor: Prof. Dr. Holger Brandt

1. Reviewer: Prof. Dr. Holger Brandt
2. Reviewer: Prof. Dr. Augustin Kelava

Declaration of Academic Integrity

Hereby, I declare that I have composed the presented paper independently on my own and without any other resources than the ones indicated. All thoughts taken directly or indirectly from external sources are properly denoted as such.

This thesis has neither been previously submitted as a whole or in any significant part as part of any other examination process, and the thesis has not been published as a whole or in any significant part, and that the copy submitted in electronic file form is identical in content to the bound copies submitted.

Yeong Hwangbo

Name

Tübingen, September 30, 2025

Place, Date

Signature

Acknowledgements

I would like to express my gratitude to Prof. Holger Brandt for his support and guidance, and for always being willing to help. I am grateful for your feedback and advice, from which I learned a lot throughout this project. Special thanks to my family who always supported me and believed in me with patience.

Abstract

In intensive longitudinal data (ILD) collected multiple times from a large number of sample, changes in underlying factor structures are more likely to occur over time. Such dimensionality changes arise from unobserved heterogeneity (e.g., reconceptualization of latent constructs, changes in response pattern) and are likely to emerge in different ways across individuals. These are modeled using a dynamic latent class structural equation modeling (DLC-SEM) with latent states that follow hidden markov models (HMM). We extend the DLC-SEM framework in two respects: exploratory factor analysis (EFA) and regularization. EFA is specified as a measurement model to explore underlying factor structures and dimensionality changes over time, and to address the unstable nature of factor structures for psychological constructs (e.g., Hamilton Rating Scale for Depression). In conjunction with EFA, cross-loadings are penalized via shrinkage priors to obtain a sparse model and prevent overfitting, which often occurs in complex models such as DLC-SEM. The adjusted DLC-SEM model is tested for convergence under different conditions with varying MCMC parameters in a simulation study, and model performance is evaluated. The model is applied to the Hamilton Rating Scale for Depression as an empirical example to evaluate model performance, particularly in detecting changes in factor structure.

Keywords: dynamic latent class structural equation modeling, exploratory factor analysis, intensive longitudinal data, Bayesian estimation

Contents

1 Introduction **1**

1.1 Outline 2

2 Theoretical Background **5**

2.1 Dynamic Latent Class Structural Equation Models 5

2.1.1 The Between-level Models 7

2.1.2 The Within-level Models 7

2.1.3 The Markov Switching Model 9

2.1.4 Random Effects 9

2.1.5 Submodels 9

2.2 Bayesian Regularized Exploratory Factor Analysis 10

3 Simulation Study **15**

3.1 Method 15

3.1.1 Model and Prior Specification 15

3.1.2 Data Generation 20

3.1.3 Procedure 21

3.2 Results 22

3.2.1 Convergence Diagnostics 22

3.2.2 Model Comparison 27

3.2.3 Performance Evaluation 32

4 Empirical Example: Hamilton Rating Scale for Depression **35**

4.1 Method 35

4.1.1 Data 35

4.1.2 Procedure 36

4.2 Results 37

4.2.1 Exploratory Factor Analysis 37

4.2.2 Model Selection 41

4.2.3 Parameter Estimates 45

5 Discussion **49**

5.1 Limitations and Future Directions 50

References **53**

Appendix A **57**

Appendix B **61**

Appendix C **65**

Appendix D **67**

Appendix E **71**

Chapter 1

Introduction

Technological advancement in electronic devices (e.g., smartphones) enabled collections of longitudinal data repeatedly measured over specific periods in large samples (Asparouhov et al., 2017). An example of such type of data is ecological momentary assessment (EMA Moskowitz & Young, 2006) data, also known as experience sampling methods (ESM). EMA data are measured multiple times throughout the course of a study (e.g., 6 - 7 times a day Moskowitz & Young, 2006), providing more detailed and accurate real-time measurements, and ultimately facilitating statistical modeling of intensive longitudinal data (ILD) (Stinson et al., 2022).

In time-intensive longitudinal data, it is more likely that measurement models change over time (Vogelsmeier et al., 2019). If that is the case, measurement invariance is violated and thus comparisons of constructs become invalid (Vogelsmeier et al., 2019). The most likely reason for such dimensionality changes is unobserved heterogeneity in intra-individual trajectories. Previous studies revealed different forms of unobserved heterogeneity that resulted in changes in factor structures. Vogelsmeier et al. (2019) proposed Latent Markov Factor Analysis (LMFA) to detect measurement model changes due to response styles. More specifically, extreme response styles caused by distracting situations or loss of motivation are strong factors that directly affect measurement quality in time-intensive longitudinal data (Moors, 2003). Furthermore, Vogelsmeier et al. (2019) demonstrated dimensionality shift of factor models with empirical ESM data of persistent anhedonia, where three states were defined by three different factor models that assess the same set of items. Unobserved heterogeneity in the intra-individual process appeared as an effect of interventions, as not everyone was assumed to switch, and switchers were not assumed to transit between states in the same order, therefore, patterns of state switches were heterogeneous across subjects (Vogelsmeier et al., 2019).

Response styles/behaviors associated with dimensionality changes were further studied by Roman et al. (2024). DLC-SEM was used to detect inattentive responses

(e.g., random responses, choosing the middle categories in all items) by means of latent states, which were defined by two states that represent compliant responses and inattentive responses, respectively.

In research on patients' conceptualization of the alliance by Flückiger et al. (2022), patients' understanding of the alliance concept was revealed by investigating patients' state transitions from a 4-factor structure to a 2-factor structure. It can also be perceived as dimensionality changes caused by response patterns, as unobserved heterogeneity, because the conceptualization of patients was determined by examining whether patients' responses are represented by 4-factor or 2-factor structures over time (Flückiger et al., 2022).

Kelava et al. (2022) reported university student dropout in math, using the DLC-SEM framework. Student dropout was predicted by specifying two latent states: no intention to quit and intention to quit that is indicated by higher scores of factors such as afraid to fail and stress.

The previous findings discussed above demonstrate that the DLC-SEM provides a flexible modeling framework, especially for detecting unobserved heterogeneity that involves underlying factor structure changes by means of dynamic latent states. In this thesis, we introduce the following extensions to the DLC-SEM framework: (a) EFA model is specified as a measurement model, instead of CFA model, and (b) EFA model is regularized by applying shrinkage priors. One advantage of using EFA is that it allows for exploring unknown underlying factor structure changes, including the number of factors and cross-loadings, which cannot be done in CFA because of its restrictive nature (Vogelsmeier et al., 2019). This is especially useful for specifying models for scales that numerous factor solutions exist (e.g., Hamilton Rating Scale for Depression). By allowing for cross-loadings, it enables exploring an unstable factor structure as well as detecting potential measurement model changes over time. With EFA models specified in each state, the necessity for sparsity is growing particularly for complex models such as the DLC-SEM, as noted by Kelava and Brandt (2019). We obtain simpler factor models by specifying shrinkage priors on cross-loadings. By doing so, we can avoid overfitting and obtain more interpretable and parsimonious models.

1.1 Outline

The aim of this work is to examine the capability of the DLC-SEM with two extensions applied (i.e., Bayesian dynamic EFA with hidden markov models) with respect to detecting dimensionality changes and to inspect general performance of the model (e.g., state switch recovery). To this end, the remainder of this work is structured as follows: we start by providing relevant theoretical background on the

DLC-SEM modeling frameworks and theoretical concepts of regularized EFA with examples of shrinkage priors. In Chapter 3, the model is tested under different conditions varying with respect to MCMC parameters to investigate behaviors of the model for convergence and to discover the best choice of MCMC parameters that enhance convergence. The application of the model with an empirical dataset is presented in Chapter 4, exhibiting behaviors of the model to capture changes in dimensionality. This thesis ends with highlighting the key findings and discussing the limitations of this work with future directions.

Chapter 2

Theoretical Background

This chapter provides theoretical background on the modeling framework and techniques used in the analyses that will be discussed later. We present a Dynamic Latent Class Structural Equation Modeling, and then a regularized exploratory factor analysis with Bayesian estimation. We also discuss how these two methods are combined and applied in the analyses.

2.1 Dynamic Latent Class Structural Equation Models

A dynamic latent class structural equation modeling (DLC-SEM) proposed by Kelava and Brandt (2019) is a comprehensive statistical modeling framework that integrates two existing frameworks: Dynamic Structural Equation Modeling (DSEM Asparouhov et al., 2018) and Dynamic Latent Class Analysis (DLCA Asparouhov et al., 2017).

The DSEM is developed based on structural equation models, multilevel models, and time-series models. The combined modeling techniques allow for the modeling of dynamics of latent and observed variables, as well as the separation of time-specific and person-specific effects. For example, time dependence between consecutive observations is modeled using the autoregressive process (Asparouhov et al., 2018). In ILD, such time-series components are a part of within-level models, as those are time- and person-dependent components. Person-specific and time-specific effects are modeled as components of between-level models, where each effect consists of a structural model and a measurement model (Asparouhov et al., 2018). Person-specific and time-specific random effects are explained in respective structural models on the between-level (Kelava & Brandt, 2019). To describe intra-individual changes, the autoregressive process can also be applied to latent variables to describe dynamics of latent variables (Kelava & Brandt, 2019).

With these dynamic latent variable models, within-person dynamics can be further explained by inter-individual differences and time-specific effects (Kelava & Brandt, 2019).

The DLCA is formulated as a combination of the DSEM mixture model with Hidden Markov Models (HMM Altman, 2007) and the multilevel latent transition model (Asparouhov et al., 2017). Based on DSEM models, a latent class variable is now added and measured for each individual at each time point (Kelava & Brandt, 2019). Unobserved heterogeneity or changes of intra-individual trajectories over time are captured by latent classes (Vogelsmeier et al., 2019). Switches between different latent classes follow a Hidden Markov process, which means a latent class membership at time point t depends on the previous latent class membership at time point $t-1$. Because latent class memberships change over time, these dynamic latent classes can be referred to as latent states (Vogelsmeier et al., 2019). Whether to stay within a state or not is determined by person-specific transition probabilities, forming a transition matrix. Person-specific transition probabilities are estimated as random effects on the between-level, and inter-individual differences such as personality traits are used to predict transition probabilities between latent states (Kelava & Brandt, 2019).

The DSEM and the DLCA are complementary in that DLCA accounts for unobserved heterogeneity in individual trajectories using latent states, which is not addressed in DSEM, and DSEM separates time-specific and person-specific random effects on the between-level, which is not done in DLCA (Kelava & Brandt, 2019). By addressing these aspects, the DLC-SEM framework integrates DSEM and DLCA into a single framework as a comprehensive approach. Nonlinear effects (e.g., splines or interactions) are now part of the model, allowing flexibility in modeling.

The DLC-SEM begins with the decomposition of the observed scores, which are indicated by $Y_{it} = (Y_{i1t}, Y_{i2t}, ..., Y_{iJt})'$, a $J \times 1$ vector of multivariate responses to J items for individual i at time point t. Due to the nature of longitudinal studies, the number of measurements might differ across individuals, and thus, time points t are denoted as $t = 1, ..., T_i$. Following the notation of Kelava and Brandt (2019), Y_{it} is decomposed into three components

$$Y_{it} = Y_{1it} + Y_{2i} + Y_{3t}, \tag{2.1}$$

where Y_{2i} and Y_{3t} are individual-specific and time-specific components on the between-level. Y_{1it} is the deviation of the item score of subject i at time point t from Y_{2i} and Y_{3t}, and is a part of the within-level models. Each component is defined by the structural equation models that consist of a measurement model

and a structural model.

2.1.1 The Between-level Models

On the between-level, the individual-specific component Y_{2i} takes the following form

$$Y_{2i} = \nu_2 + \Lambda_2 \eta_{2i} + K_2 X_{2i} + \epsilon_{2i} \tag{2.2}$$

$$\eta_{2i} = \alpha_2 + B_2 \eta_{2i} + \Omega_2 h_2(\eta_{2i}) + \Gamma_2 X_{2i} + \zeta_{2i}. \tag{2.3}$$

As Y_{2i} accounts for person-specific contributions that are time-invariant, η_{2i} and X_{2i} are a vector of individual-specific, time-independent covariates and latent variables, respectively. An example of X_{2i} is race or gender, and η_{2i} could be intelligence score (e.g., IQ) fixed at baseline (Kelava & Brandt, 2019). Nonlinear effects are specified as $h_2(\eta_{2i})$, as a vector of functions of η_{2i}. If interactions are assumed, $h_2(\eta_{2i})$ can be written as a vector of products of latent variables $vech(\eta_{2i}\eta_{2i}\prime)$ (Kelava & Brandt, 2019). The remaining variables are defined as follows: ϵ_{2i} and ζ_{2i} are multivariate residuals with zero means, and ν_2, Λ_2, K_2, α_2, B_2, Ω_2 and Γ_2 are fixed effects matrices.

The time-specific contributions that are individual-independent, such as average time trends, are modeled by the time-specific component Y_{3t} on the between-level (Kelava & Brandt, 2019). The Y_{3t} model is given by the following equations

$$Y_{3t} = \nu_3 + \Lambda_3 \eta_{3t} + K_3 X_{3t} + \epsilon_{3t} \tag{2.4}$$

$$\eta_{3t} = \alpha_3 + B_3 \eta_{3t} + \Omega_3 h_3(\eta_{3t}) + \Gamma_3 X_{3t} + \zeta_{3t}. \tag{2.5}$$

X_{3t} and η_{3t} are a vector of time-specific, individual-invariant covariates (e.g., seasonal events that affect all individuals) and latent variables, respectively. $h_3(\eta_{3t})$ is a vector of functions of η_{3t}, which is included if nonlinear effects are assumed in the data. The ϵ_{3t} and ζ_{3t} are vectors of multivariate residuals with zero means. ν_3, Λ_3, K_3, α_3, B_3, Ω_3 and Γ_3 are fixed effects matrices.

2.1.2 The Within-level Models

On the within-level, intra-individual changes in developmental trajectories are modeled by Y_{1it}, and a latent state variable and time-series components are now

included. The within-level models are as follows

$$[Y_{1it}|S_{it}=s] = \nu_{1s} + \sum_{l=0}^{L} \Lambda_{1ls}\eta_{1i(t-l)} + \sum_{l=0}^{L} R_{ls}Y_{1i(t-l)} + \sum_{l=0}^{L} K_{1ls}X_{1i(t-l)} + \epsilon_{1it} \quad (2.6)$$

$$\begin{aligned}[\eta_{1it}|S_{it}=s] = \alpha_{1s} &+ \sum_{l=0}^{L} B_{1ls}\eta_{1i(t-l)} + \sum_{l=0}^{L}\sum_{l'=0}^{L'} \Omega_{1ll's}h_{1ll'}(\eta_{1i(t-l)}, \eta_{1i(t-l')}) \\ &+ \sum_{l=0}^{L} Q_{ls}Y_{1i(t-l)} + \sum_{l=0}^{L} \Gamma_{1ls}X_{1i(t-l)} + \zeta_{1it}.\end{aligned} \quad (2.7)$$

S_{it} is a latent state variable for an individual i at time t. It is a categorical variable such that $S_{it} = 1, ..., K$, where K is the number of states. Y_{1it} and η_{1it} now depend on the values of S_{it}, and thus the within-level equations are state-specific. Furthermore, time-series models enable the modeling of lagged effects from the past l time points (i.e., autoregressive models (AR)), which can be included not only in observed variables but also in latent variables such as η_{1it}. If residuals are assumed to have lagged effects in the data, they can be modeled by moving average models (MA). X_{1it} is a vector of person- and time-specific covariates (e.g., weights or blood pressure), and η_{1it} is a vector of person- and time-specific latent variables such as a latent factor for deviation of IQ test score from average IQ across persons and across time points (Kelava & Brandt, 2019). Nonlinear effects are described via $h_{1ll'}(\eta_{1i(t-l)}, \eta_{1i(t-l')})$, which is a vector of nonlinear functions of latent variables at different lags (Kelava & Brandt, 2019). Similar to the between-level models, ϵ_{1it} and ζ_{1it} are multivariate residual vectors with zero means, and $\nu_{1s}, \Lambda_{1ls}, R_{ls}, K_{1ls}, \alpha_{1s}, B_{1ls}, \Omega_{1ll's}$ and Γ_{1ls} are fixed or random effects matrices.

Thus far, for simplicity, we assumed the continuous observed vector Y_{it}. If $[Y_{1jit}|S_{it}=s]$, j-th element of the vector $[Y_{1it}|S_{it}=s]$, is a categorical variable with categories $1, ..., m_j$, it can be converted to a normally distributed latent variable $[Y^*_{1jit}|S_{it}=s]$ with threshold parameters $\tau_{j1s}, ..., \tau_{j(m_j-1)s}$. The equation is then given by

$$[Y_{1jit}=m|S_{it}=s] \Leftrightarrow \tau_{j(m-1)s} \leq [Y^*_{1jit}|S_{it}=s] < \tau_{jms}, \quad (2.8)$$

where $\tau_{j0s} = -\infty$ and $\tau_{j(m_j)s} = \infty$ for all latent states $s = 1, ..., K$. In this work, all the item vectors Y_{it} denote continuous variables.

2.1.3 The Markov Switching Model

The person- and time-specific and latent state variable S_{it} follows the HMM process. Its transition probabilities are given by

$$P(S_{it} = d|S_{i(t-1)} = c) = \frac{exp(\alpha_{itdc})}{\sum_{k=1}^{K} exp(\alpha_{itkc})}. \tag{2.9}$$

The transition probabilities are individual- and time-dependent and are defined for all pairs of latent states, yielding a transition matrix P_{it} for each person at each time point. α_{itdc} are person- and time-specific random effects. For identification purposes, $\alpha_{itKc} = 0$.

2.1.4 Random Effects

At the within-level, the parameters are allowed to vary across subjects, time, or both. It can be specified in the within-level equations by introducing indices i and t for the parameters (e.g., $\Lambda_{1ilts}, B_{1ilts}$ for random loadings). These random effects at within-level p_{it} are decomposed as

$$p_{it} = p_{2i} + p_{3t}, \tag{2.10}$$

where p_{2i} is an individual-specific random effects that is estimated by the structural model η_{2i} on the between-level. p_{3t} is a time-specific random effect that is modeled as a part of the time-specific latent vector η_{3t}.

This holds for the within-level residual variances $Var(\epsilon_{1it})$ and $Var(\zeta_{1it})$, that is, residual variances ν_{it} are allowed to have random effects. It can be used, for example, to separate heteroscedasticity into individual-specific and time-specific sources (Kelava & Brandt, 2019). The ν_{it} can be decomposed as follows

$$\nu_{it} = exp(p_{2i} + p_{3t}), \tag{2.11}$$

where p_{2i} is a subject-specific random effect, and p_{3t} is a time-specific random effects. These random effects are again explained by the higher level structural equations η_{2i} and η_{3t}, respectively.

2.1.5 Submodels

The models described in the previous sections are the general framework that is used to investigate changes in intra-individual trajectories over time. Depending on the research questions or research hypotheses for the data, many variations can be derived from the general framework as special cases. Examples of special cases

include Multilevel Markov Switching Autoregressive Model (MMSAR Asparouhov et al., 2017), Measurement Error AR(1) (Asparouhov et al., 2018, MEAR(1)), two-level dynamic SEM models, and two-level dynamic LCA models (Kelava & Brandt, 2019). One of the special cases that might be interesting for psychologists is the two-class regime switching model for latent factors proposed by Asparouhov et al. (2017), because latent factors are often variables of interest in psychological studies (Asparouhov et al., 2017). The model can be written as

$$\begin{aligned} Y_{pit} &= \nu_{pi} + \Lambda_p \eta_{it} + \epsilon_{pit} \\ \eta_{it} &= \mu_{S_{it}} + \beta_{S_{it}} \eta_{i(t-1)} + \zeta_{it,S_{it}} \\ P(S_{it} = 1 \mid S_{i(t-1)} = j) &= \frac{\exp(\alpha_{ij})}{1 + \exp(\alpha_{ij})}. \end{aligned} \tag{2.12}$$

Y_{pit} is a $p \times 1$ vector of observed variables for individual i and time t where $p = 1, ..., 4$. Because η_{it} follows a two-class regime switching model with an AR(1) process, the parameters in the structural model are state-specific. The transition probabilities are person-specific and are predicted by inter-individual differences. ν_{pi} are four random intercepts for four observed variables.

The DLC-SEM framework can be estimated using both Bayesian methods and frequentist methods. In this work, we focus on Bayesian methods with MCMC simulation. Bayesian analyses are performed by specifying priors that represent researchers' subjective beliefs or assumptions about parameters. This is reflected by the choice of prior distributions, which are informative, weakly informative, or non-informative. A detailed explanation of the estimation procedure, including prior specification, is provided in the following chapters.

2.2 Bayesian Regularized Exploratory Factor Analysis

As the DLC-SEM framework allows flexible modeling of many features such as intra-individual changes, inter-individual differences, lagged effects, unobserved heterogeneity captured by HMM, and non-linear effects, the models might be overly complex (Kelava & Brandt, 2019). Such models are computationally intensive and involve too many parameters, which result in overfitted models (Kelava & Brandt, 2019). To avoid the problems arising from the high complexity of the models, spare models can be considered to obtain simpler structures with fewer parameters and more interpretable results. This can be achieved by regularized estimation that penalizes parameters with small effects exactly to zero (e.g., Lasso) or towards zero (e.g., Ridge) while leaving parameters with large effects. In Bayesian estimation,

sparsity is achieved via shrinkage priors.

With the need for sparsity, we employ Bayesian regularized exploratory factor analysis to model intensive longitudinal data, in conjunction with the DLC-SEM framework. The models that we will use are based on DLC-SEM models with HMM, where the measurement models for each state allow for cross-loadings, and thus EFA models. In other words, EFA models with different numbers of factors define each latent state, and dynamic changes in underlying factor structures of EFA models (i.e., switches between latent states) are modeled by the HMM process by estimating transition probabilities given the latent state at the previous time point.

The reason for using EFA rather than CFA is to utilize the advantage of EFA, that is, EFA enables discovery of underlying factor structures by means of cross-loadings, while CFA is not capable of it because of restrictions. Moreover, EFA becomes more advantageous in situations where a large number of factor models are suggested as factor solutions for a single scale (e.g., Hamilton Rating Scale for Depression).

Bayesian approaches to EFA enable a one-step procedure for EFA where factor extraction and parameter estimation are estimated simultaneously, which is in contrast to traditional approaches where all the steps are performed sequentially, with arbitrary decisions made by researchers involved in each step (Conti et al., 2014). The arbitrariness of choices in the steps (e.g., the number of factors to be retained) and indeterminacy in factor rotation (i.e., choice of factor rotation methods) are resolved in the unified procedure for EFA because the dimensionality of a factor structure are treated as a parameter and determined by a prior, jointly estimating factor loadings, and sparsity and interpretability of factor structures are addressed by shrinkage priors without going through the rotation step (Conti et al., 2014).

The main purpose of applying regularization to EFA models in this work is to avoid cross-loadings, and it can be achieved by specifying shrinkage priors on cross-loadings. To this end, we utilize the approach proposed by Lu et al. (2016) to perform Bayesian regularized estimation. Lu et al. (2016) suggested a one-step Bayesian variable selection approach that combines selection of factor structure and final parameter estimation into a single step. To describe the idea of the approach, consider the factor model

$$y_i = \Lambda \eta_i + \epsilon_i, \tag{2.13}$$

where y_i is a $p \times 1$ vector of observed indicators with sample size $i = 1, ..., N$. Λ is a $p \times q$ factor loading matrix where λ_{jk} is the factor loading of item j on

factor k. η_i is a $q \times 1$ vector of latent factors, and ϵ_i is a $p \times 1$ vector of residuals that follows a multivariate normal distribution with zero means. The idea of the approach is that factor loadings are categorized into two groups—main loadings and cross-loadings—and that priors are specified on each type separately. Main-loadings are elements in the loading matrix Λ that take non-zero values significantly distant from zero (Lu et al., 2016). The rest of the elements of Λ are classified as cross-loadings that are expected to be close to zero. Lu et al. (2016) used normal priors on main-loadings, with variances that are assumed to be sufficiently large, so as to have values deviate from zero. For all cross-loadings, Lu et al. (2016) proposed using the Spike-and-Slab prior (Mitchell & Beauchamp, 1988) to shrink cross-loadings toward zeros. Following the notation of Lu et al. (2016), it is given by

$$\lambda_{jk} \sim (1 - \gamma_{jk})N(0, \sigma_{jk}^2) + \gamma_{jk}N(0, c_{jk}^2). \tag{2.14}$$

It is a mixture distribution where the first component is defined as a *spike* that shrinks and selects variables, reflecting a belief that $\lambda_{jk} = 0$, and the second component models a *slab* part that is often specified as a wider distribution for non-zero loadings (Brandt et al., 2025). σ_{jk}^2 takes a value that is either zero or near zero (e.g., 0.01), yielding a point mass at zero or a normal distribution centered around zero, respectively (Lu et al., 2016). The former case results in a prior that is a mixture of a discrete value and a normal distribution (Brandt et al., 2025). As with σ_{jk}^2, c_{jk}^2 is the scaling parameter of the slab, and is assumed to be larger than σ_{jk}^2 to ensure the non-zero loading (Brandt et al., 2025). γ_{jk} follows Bernoulli distribution or Beta distribution. If γ_{jk} is Bernoulli distributed, it plays the role of an indicator, taking value either 0 or 1, with $\gamma_{jk} = 0$ indicating spike (i.e., λ_{jk} set to zero) and $\gamma_{jk} = 1$ representing slab part (i.e., non-zero λ_{jk}) (Brandt et al., 2025). If γ_{jk} follows a Beta distribution, it functions as a weight or probability of λ_{jk} being non-zero (Brandt et al., 2025).

In addition to the spike-and-slab prior, there are many different priors that can be used as shrinkage priors. We describe two of them, which were applied in the analyses in the later chapters: Bayesian lasso prior (Park & Casella, 2008) and Horseshoe prior (Carvalho et al., 2009).

The Bayesian lasso is the Bayesian analogue of Lasso regularization (Tibshirani, 1996). Adopting the notation of Brandt et al. (2025), the Bayesian lasso for a

factor loading λ_{jk} is specified as

$$\begin{aligned} \lambda_{jk} &\sim Laplace(\sigma/\gamma_{(jk)}) \\ \gamma^2_{(jk)} &\sim Gamma(a_1, b_1) \\ \sigma^{-2} &\sim Gamma(a_2, b_2). \end{aligned} \tag{2.15}$$

λ_{jk} follows a double exponential (Laplace) distribution with zero mean. σ^{-2} and $\gamma^2_{(jk)}$ are the gamma-distributed scale parameters that are controlled by hyper-parmeters a and b. Specifically, $\gamma^2_{(jk)}$ is a shrinkage factor that controls the overall degree of shrinkage, which means the same shrinkage strength is applied to all loadings. However, as observed in the index of $\gamma^2_{(jk)}$, shrinkage can be applied to each loading differently, which indicates extension of Bayesian lasso to Bayesian adaptive lasso (**leng**). Compared to the spike-and-slab prior, Bayesian lasso may yield less parsimonious factor structure, because the spike-and-slab prior shrinks loadings exactly to zero by means of a point mass (Brandt et al., 2025). Moreover, the light tail of the Laplace distribution assigns high probability to small values, while larger values are unlikely. This implies that Bayesian lasso efficiently shrinks not only smaller values but larger values also, which might lead to over-shrinkage.

The horseshoe prior is one of the global-local type priors, composed of a global shrinkage part and a local shrinkage part. With the notation taken from Brandt et al. (2025), the horseshoe prior is given by

$$\begin{aligned} \lambda_{jk} &\sim N(0, \tau^2\gamma^2_{jk}) \\ \gamma_{jk} &\sim C^+(0, 1) \\ \tau &\sim C^+(0, a), \end{aligned} \tag{2.16}$$

where τ represents a global shrinkage part following the Half-Cauchy distribution with zero mean and scale a. Overall strength of shrinkage of λ_{jk} is controlled by τ, which is applied to all loadings. γ_{jk}, as a local shrinkage parameter, affects each loading differently. In contrast to Bayesian lasso, the heavy-tailed Half-Cauchy distribution of γ_{jk} allows loadings with large effects to preserve their effects, while the tall spike at zero shrinks small loadings very close to zero (Carvalho et al., 2009). This provides theoretical unbiasedness of larger values (Brandt et al., 2025). Taken together, global shrinkage τ works in conjunction with γ_{jk}, that is, τ encourages sparsity, while γ_{jk} prevents loadings with large effects from being over-shrinkage, providing adaptive regularization.

Chapter 3

Simulation Study

In this chapter, we illustrate the methods and results of the simulation studies using DLC-SEM models with Bayesian regularized EFA. We first present the simulation setup, including the specification of models and priors used for simulation analyses, followed by the data generation process and the procedure of the analyses. In the chapter 3.2, we evaluate the convergence of models via convergence statistics and scatter plots, and examine how closely models recovered true values. Next, we evaluate performances in recovering state parameters.

3.1 Method

3.1.1 Model and Prior Specification

To perform the simulation study, we started by assuming the structure of a model. The model was expressed using the DLC-SEM framework with regularized EFA. Let Y_{it} be 9×1 vector of observed item responses for individuals $i = 1, ..., N$ with $N = 100$ or 200 and time points $t = 1, ..., 15$. Y_{it} is decomposed into two components:

$$Y_{it} = Y_{1it} + Y_{2i}, \tag{3.1}$$

where Y_{2i} denotes person-specific components that is time-invariant (i.e., inter-individual differences), and Y_{1it} represents fluctuations in intra-individual trajectories.

The resulting Y_{it} was assumed to switch between two latent states, where state 1 represented a two-factor EFA model, and state 2 represented a three-factor EFA model. It is important to note that the factor models were specified for each state separately, because we assumed that the underlying factor structure would change over time. Accordingly, the model was interpreted independently for each state. Both EFA models were based on the same 9 items measured across 15 occasions.

The Between-level Models

On the between-level, we assumed person-specific random intercepts for latent factors. Such random effects are modeled as η_{2i}, a latent variable at the between-level. Y_{2i} can be written as

$$\begin{aligned} Y_{2i} &= \Lambda_2 \eta_{2i} \\ \eta_{2i} &= \alpha_2 + \zeta_{2i}, \end{aligned} \tag{3.2}$$

where Λ_2 is a factor loading matrix, η_{2i} represents a vector of latent factors with random effects included. More precisely, η_{2i} consists of α_2, which is a vector of fixed intercept for latent factors, and ζ_{2i}, which models person-specific random intercepts. Since α_2 was nearly zero, α_2 was constrained to zero for identification and simplicity. As noted earlier, two different EFA models were specified independently for each state to represent the model assumption, and therefore the equations 3.2 hold only within a state, and all the parameters in the equation were coded as state-dependent variables, adding the index s.

Since α_2 was set to zero, the prior for η_{2i} can be defined by prior for ζ_{2i}. The normal prior was specified for ζ_{2i}:

$$\zeta_{2i} \sim N(0, \Phi_2), \tag{3.3}$$

where means of the random intercepts were set to zero. Φ_2 is a covariance matrix of random intercepts and is defined for each state.

The Within-level Models

The person- and time-variant factor structure and the time series models were modeled on the within level. The within-level measurement models were defined by the factor structure, which is specified as

$$[Y_{1it}|S_{it} = s] = \nu_{1s} + \Lambda_2 \eta_{1it} + \epsilon_{1it}. \tag{3.4}$$

The model is state-dependent, with a two-factor EFA specified for $S = 1$ and a three-factor EFA specified for $S = 2$. Λ_2 is a state-specific factor loading matrix that contains cross-loadings. ϵ_{1it} is a state-specific residual for Y_{1it}.

Note that the same Λ_2 was used as in the between-level model, and within each state it is assumed to be invariant across individual i and time points t. In other words, Λ_2 is state-specific but constant across levels, persons, and time for each state. Thus, in such cases it is possible to write Λ_2 and random effects jointly as $\Lambda_2(\eta_{1it} + \zeta_{2i})$. For simplicity of the model, the intercept of the factor model was constrained to zero (i.e., $\nu_{1s} = 0$).

For each item $j = 1, ..., 9$, state-dependent Y_{1jit} follows the normal distribution, which is given by

$$[Y_{1jit}|S_{it} = s] \sim N(\mu_{Y_{1jit}}, \sigma^2_{\epsilon_{1j}}), \tag{3.5}$$

with the mean $\mu_{Y_{1jit}}$ and the variance $\sigma^2_{\epsilon_{1j}}$. The mean $\mu_{Y_{1jit}}$ is explained by the factor structure $\Lambda_2\eta_{1it}$, where η_{1it} is state-specific.

As the factor loading matrix Λ_2 contains cross-loadings, regularization was applied to shrink loadings with negligible effects toward zero while retaining those with large effects using a Bayesian variable selection approach by Lu et al. (2016). That is, priors were specified separately for primary loadings and cross-loadings. We set the following form for Λ_2:

$$\begin{bmatrix} \lambda_{11} & \lambda_{12}{}^* \\ \lambda_{21} & \lambda_{22}{}^* \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \\ 0 & \lambda_{92} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \lambda_{11} & \lambda_{12}{}^* & 0 \\ \lambda_{21} & 0 & \lambda_{23}{}^* \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{bmatrix}, \tag{3.6}$$

where the matrices on the left and the right represent the factor structures for states 1 and 2, respectively. λ_{jk} denotes the factor loading of item j on factor k, and cross-loadings are indicated by $\lambda_{jk}{}^*$.

For all the primary loadings, weakly informative normal priors were applied as follows:

$$\lambda_{jk} \sim N(0.6, 1). \tag{3.7}$$

The mean was set to 0.6 as specified in the data generation process that will be discussed in the next section.

To apply priors on cross-loadings, we tested three different shrinkage priors: Bayesian lasso, Horseshoe, and spike-and-slab. Of all considered priors, the horseshoe prior outperformed the other priors and therefore was selected for the analyses. The horseshoe prior for cross-loadings is given by

$$
\begin{aligned}
{\lambda_{jk}}^* &\sim N(0, 1/\gamma_{jk}^2) \\
\gamma_{jk} &\sim C^+(0, 1/\tau) \\
\tau &\sim C^+(0, 1),
\end{aligned}
\tag{3.8}
$$

where each cross-loading follows a normal distribution with zero mean. The precision of the cross-loadings is determined by γ_{jk} and τ hierarchically: The half-cauchy distributed local shrinkage γ_{jk} is controlled by global shrinkage τ with scale 1. This allows adaptive shrinkage of cross-loadings, preserving cross-loadings with significantly large effects while maintaining sparsity.

The latent factors in each state were assumed to follow an AR(1) process. It was specified in the within-level structural model as follows:

$$
[\eta_{1it}|S_{it} = s] = \alpha_{1s} + B_{1s}\eta_{1i(t-1)} + \zeta_{1it}, \tag{3.9}
$$

where α_s is a state-specific intercept, and ζ_{1it} is state-specific residual for η_{1it}. B_{1s} represents 2×2 and 3×3 matrices of autoregressive coefficients for factors in state 1 and state 2 respectively. As parameter estimates for cross-lagged effects were nearly zero, we constrained cross-lagged effects to zero (i.e., non-diagonal elements of B_{1s}) and estimated diagonal elements only (i.e., lagged effects).

Equation 3.9 was centered around the mean of η_{1it} at $t = 1$. We defined $\mu_{\eta_{1i1}}$ as $\alpha_2 + \zeta_{2i}$. With η_{1it} mean-centered, Equation 3.9 can be specified for the first two time points as follows:

$$
\begin{aligned}
[\eta_{1i1}|S_{i1} = s] &= \alpha_2 + \zeta_{2i} + \zeta_{1i1} \\
[\eta_{1i2}|S_{i2} = s] &= \alpha_2 + \zeta_{2i} + B_{1s}(\eta_{1i1} - \alpha_2 - \zeta_{2i}) + \zeta_{1i2}
\end{aligned}
\tag{3.10}
$$

for $t = 1$ and $t > 2$ respectively. Instead of specifying α_{1s}, the centered model above was used for the analysis.

The normal prior was applied to η_{1it}

$$
[\eta_{1it}|S_{it} = s] \sim N(\mu_{\eta_{1it}}, \Phi_1), \tag{3.11}
$$

with the mean $\mu_{\eta_{1it}}$ and the covariance matrix of latent factors Φ_1. Since η_{1it} is state-dependent, its mean and covariance are defined for each state.

For AR coefficients, the following prior was chosen:

$$
B_{1s} \sim unif(-1, 1), \tag{3.12}
$$

where diagonal elements of the state-specific matrix B_{1s} are sampled from the

uninformative uniform distribution ranging from -1 to 1.

In JAGS, all the variances and the covariance matrices defined earlier are specified as precision. Therefore, priors for the variances and the covariance matrices were defined on their inverses.

The precision for random intercepts Φ_2^{-1} and the precision for factors Φ_1^{-1} were specified for each state. The following non-informative Wishart priors were applied to the precisions with the degree of freedom 2 in state 1,

$$\begin{aligned} \Phi_2^{-1} &\sim Wishart(\Phi_0^{-1}, 2) \\ \Phi_1^{-1} &\sim Wishart(\Phi_0^{-1}, 2) \end{aligned} \tag{3.13}$$

and with the degree of freedom 3 in state 2,

$$\begin{aligned} \Phi_2^{-1} &\sim Wishart(\Phi_0^{-1}, 3) \\ \Phi_1^{-1} &\sim Wishart(\Phi_0^{-1}, 3), \end{aligned} \tag{3.14}$$

where Φ_0^{-1} denotes 2×2 and 3×3 identity matrices in state 1 and 2 respectively.

For state-dependent residual precision for Y_{1it}, uninformative gamma prior was used

$$\sigma_{\epsilon_{1j}}^{-2} \sim Gamma(1, 1) \tag{3.15}$$

for $j = 1, ..., 9$.

The Markov Switching Model

The transition probabilities of latent state S_{it} that follows the HMM process are collected in a transition matrix P_{it} for each person i at each time point t. The transition probabilities were individual- and time-specific, and thus were modeled at the within-level. The transition matrix P_{it} can be specified as following 2×2 matrix

$$P_{it} = \begin{bmatrix} P_{11it} & P_{12it} \\ P_{21it} & P_{22it} \end{bmatrix}, \tag{3.16}$$

where P_{11it} represents probability of remaining in state 1 at time t and P_{12it} represents probability of transition to state 2 at time t to from state 1. P_{21it} and P_{22it} are considered if state at $t-1$ is 2. Similarly, P_{21it} and P_{22it} are the probability of staying in state 2 and the probability of switching from state 2 to state 1.

S_{it}, the latent state variable for subject i at time t, is sampled the transition

matrix P_{it}, which is modeled via categorical distribution:

$$S_{it} \sim Categorical(P(S_{it} = s|S_{i(t-1)} = s')) \text{ for } s, s' = 1, 2. \tag{3.17}$$

Depending on the state at the previous time point $t-1$, different elements of P_{it} are considered. If an individual i was in state 1 at $t-1$, S_{it} is chosen between P_{11it} and P_{12it}. Similarly, if in state 2, S_{it} is drawn from P_{21it} and P_{22it}. For simplicity, we assumed that all individuals start in state 1 at $t = 1$, that is, $S_{i1} = 1$.

To predict the probability of remaining in state 1 at time point t, we used the information from latent factors in state 1 at $t-1$. This is modeled via logit, where the linear regression was formed by the two factors in state 1: $\eta_{11i(t-1)}$ and $\eta_{12i(t-1)}$. The model can be expressed as:

$$logit(P_{11it}) = \beta_0 + \beta_1\eta_{11i(t-1)} + \beta_2\eta_{12i(t-1)} \tag{3.18}$$

$$P_{12it} = 1 - P_{11it}, \tag{3.19}$$

with P_{12it} determined by subtracting P_{11it} from 1. β_1 and β_2 indicate regression coefficients for logit model, and β_0 indicates an intercept. For β_k $(k = 0, 1, 2)$, the following normal prior was used:

$$\beta_k \sim N(0, 0.5) \tag{3.20}$$

with zero mean and a variance of 0.5.

For simplicity and model stabilization, P_{21it} and P_{22it} were fixed to 0.99 and 0.01, respectively. In other words, once an individual i enters state 2, it is unlikely to switch back to state 1. These values are identical across all individuals i and time points t. The model can be written as follows:

$$\begin{aligned} P_{21it} &= 0.01 \\ P_{22it} &= 1 - P_{21it} = 0.99. \end{aligned} \tag{3.21}$$

3.1.2 Data Generation

Based on the models assumed in the previous section, data were generated for 9 items across 15 measurement occasions and 100 individuals. For the between-level model, we set the mean of person-specific random intercepts for factors (ζ_{2i}) to zero. The factor intercepts (α_2) were generated with a mean of zero. The variance of α_2 was set to 0.3 with factors uncorrelated with one another.

In within-level, both primary loadings and cross-loadings were fixed to 0.6. This gives the following loading matrix Λ_2

$$\begin{bmatrix} 0.6 & 0.6^* \\ 0.6 & 0.6^* \\ 0.6 & 0 \\ 0.6 & 0 \\ 0 & 0.6 \\ 0 & 0.6 \\ 0 & 0.6 \\ 0 & 0.6 \\ 0 & 0.6 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0.6 & 0.6^* & 0 \\ 0.6 & 0 & 0.6^* \\ 0.6 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.6 \end{bmatrix}, \tag{3.22}$$

where the left and right matrices are loading matrices for states 1 and 2, respectively. In the analysis, the first loadings on each factor were fixed to 1 for identification and stabilization of the model.

The residual variance of Y_{1jit} for items j was generated by subtracting squared loadings of item j from 1 (i.e., $1 - {\lambda_{jk}}^2$). The AR coefficients were set to 0.5 with no cross-lagged effects assumed.

From this process, two- and three-factor models were obtained. Next, these two models were combined under the assumption that 20% of individuals switch from state 1 (two-factor model) to state 2 (three-factor model), which means that for the switchers (20% of individuals), their data was replaced by data from the three-factor model from their switch time points sampled in the data generation process. Switchers were randomly drawn from values 1 - 100 with the switch probability of 0.2, and then for those sampled, their switch time points were randomly drawn from 1 - 15, yielding 15 persons labeled as switchers.

By applying this process, we generated a combined model with $N = 200$, which is identical to the model with $N = 100$ in all aspects, except for the sample size. For the model $N = 200$, 44 individuals transitioned to state 2 from state 1.

3.1.3 Procedure

The simulation study aims to evaluate the performance of models and to explore optimal choices of numbers of iterations and thinning, and sample size for convergence. The simulation procedure run in the analyses is described as follows:

1. Generate a combined model according to the data generation process 3.1.2 for $N = 100$ and $N = 200$ each.

2. Examine Gelman-Rubin-Brooks plots (Brooks & Gelman, 1998) for each of the

two models and determine the number of iterations needed for convergence.

3. Run the model that vary with respect to: (a) the sample size $N = 100$ or $N = 200$, (b) the number of iterations/burn-in $50,000/25,000$ or $25,000/12,500$, and (c) the number of thinning interval 1 or 2, which yields $2 \times 2 \times 2 = 8$ conditions. All conditions were run for four chains.

4. Diagnose convergence of the models via $\hat{R}$ statistic (Gelman & Rubin, 1992) and effective sample size (ESS), and evaluate performance using RMSE and accuracy in a confusion matrix

The model and priors were implemented in JAGS, using MCMC sampling in R (Version 4.4.2). All analyses for the simulation study were performed using R packages `R2jags` (Version 0.8-9) and `coda` (Version 0.19-4.1).

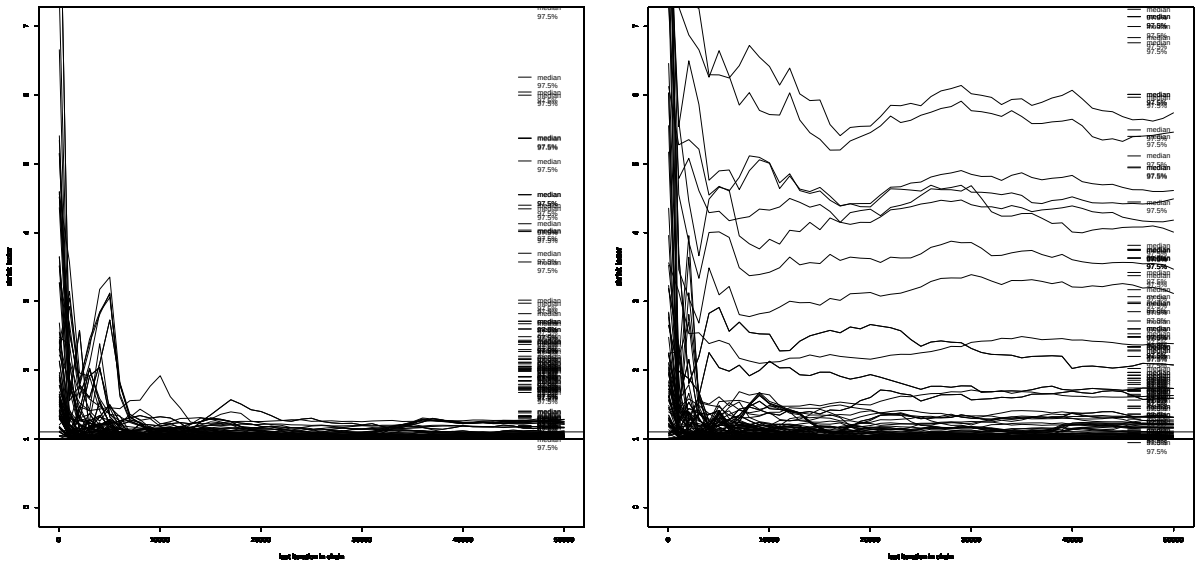
3.2 Results

3.2.1 Convergence Diagnostics

Determination of Length of Chain for Convergence

In Step 1, the models with $N = 100$ and $N = 200$ were generated. Before fitting the assumed model with the priors as specified in 4.2.2 to the generated data from Step 1, we examined evolutions of $\hat{R}$ with Gelman-Rubin-Brooks plots (reference needed) for each of the two models in order to explore the number of iterations needed for convergence. This was done by increasing the number of iterations (50,000 steps) per chain (4 chains) and setting the burn-in period to 1. Figure 3.1a and Figure 3.1b describe how $\hat{R}$ changes over 50,000 iterations for all the parameters under $N = 100$ and $N = 200$, respectively, where the x-axis represents the length of a chain and y-axis represents a shrink factor, which are referred to as $\hat{R}$. In the Figure 3.1a for the model with $N = 100$, fluctuations in $\hat{R}$ were observed until 20,000 iterations, after which $\hat{R}$ for most of the parameters were stabilized. Figure 3.1b indicated non-convergence for some parameters, with no decrease in $\hat{R}$ observed throughout the sampling process. The failure of convergence was mostly due to loading parameters, which will be addressed in the next section. Except for these parameters, the rest of the parameters converged after 15,000 iterations. Although a few of the parameters did not converge in the model with $N = 200$, the model was retained for comparison with the model with $N = 100$ and for analysis purposes.

Based on the result of Figure 3.1, we determined the length of a chain to 50,000 with half of the sample (25,000 samples) discarded (i.e., burn-in), as most of



(a) Gelman-Rubin plot for $N = 100$ (b) Gelman-Rubin plot for $N = 200$

Figure 3.1: Gelman-Rubin plots

the parameters were found to be converged after 50,000 iterations in both models. To check whether longer iterations improve the convergence of the model, we varied the length of a chain and ran the model with 25,000 iterations, yielding four conditions. Next, the four models were thinned by 1 and 2 to examine the impact of the thinning interval on effective sample size. The convergence of the resulting 8 models is assessed in the next section.

Convergence Assessment for All Parameters

Convergence for all parameters under all conditions was evaluated using $\hat{R}$ and effective sample size (ESS). Table 3.1 shows $\hat{R}$ and effective sample sizes for each condition. To diagnose non-convergence of parameters, we chose a lenient criterion for $\hat{R}$ threshold $\hat{R} > 1.2$ as suggested by Roman et al. (2024). Because complex models with larger sample sizes, such as DLC-SEM, require substantial time for implementations, setting a less stringent $\hat{R}$ threshold of 1.2 is more practical and feasible than running models with longer chains.

Overall, the model $N = 100$ indicated better convergence in terms of all convergence statistics than the model $N = 200$, as it showed significantly smaller $\hat{R}$ and larger ESS. Of all conditions within $N = 100$ the models with 50,000 iterations thinned by 1 and 2 were found to outperform the others. The findings contradict the expectation that in models with a larger sample size, parameter estimates are more likely to converge and yield more stable values. Within each sample size, the models exhibited similar patterns: Thinning did not increase ESS, and running

Table 3.1: Convergence statistics for all parameters

Model	$\hat{R}$			ESS	
Iterations/Burn-in/ Thinning	Median	Max	$\hat{R} > 1.2$	Min	Median
N = 100					
50,000 / 25,000 / 1	1.02	1.20	0	18	210
50,000 / 25,000 / 2	1.02	1.20	0	18	210
25,000 / 12,500 / 1	1.04	1.35	9	15	94
25,000 / 12,500 / 2	1.04	1.35	9	15	95
N = 200					
50,000 / 25,000 / 1	1.03	6.02	21	4	96
50,000 / 25,000 / 2	1.03	6.02	21	4	98
25,000 / 12,500 / 1	1.07	6.26	33	4	60
25,000 / 12,500 / 2	1.07	6.26	33	4	61

Note. Median, maximum values of $\hat{R}$ and the number of parameters with $\hat{R}$ larger than 1.2. Minimum and median values of ESS.

models with longer iterations improved convergence, lowering $\hat{R}$ values. In this work, however, we expected that increasing the length of iterations for the model $N = 200$ would not achieve better convergence because as illustrated in 3.1b, $\hat{R}$ of some problematic parameters remained stable throughout the sampling process.

Because of non-convergence of some loadings in model $N = 200$, maximum $\hat{R}$ were very high across all conditions in $N = 200$, and accordingly, the number of parameters with $\hat{R} > 1.2$ were larger, and minimum ESS and median ESS were lower than those in $N = 100$. Although the remaining parameters in the models $N = 200$ showed intermediate convergence (median $\hat{R} = 1.03$ and 1.07), interpretations of parameters should be done with caution, since a relatively small median ESS does not ensure reliable estimates. For the models $N = 100$, despite a small minimum ESS of 18, the model with 50,000 iterations seemed to converge as none of the $\hat{R}$ exceeded the threshold of 1.2. While in the model with 25,000 iterations, 9 parameters did not converge, the model indicated showed better convergence with respect to all measures than the model with the same condition in $N = 200$. Considering all the conditions, the results indicated that convergence improved in smaller sample sizes with longer iterations, which were demonstrated by increased $\hat{R}$ and decreased ESS. Thus, in this dynamic EFA with two latent states setup, a model with a sample size of 100 with 50,000 iterations may be the optimal choice for convergence. Parameter estimates with convergence statistics of

the model $N = 100$ with 50,000 iterations and thin = 1 are listed in Appendix 5.1.

Convergence Assessment for Loadings

After evaluation of the convergence of all parameters across conditions, convergence was examined in detail for loadings, which are parameters of interest in this study. Table 3.2 summarizes the averaged posterior mean with corresponding standard deviation, mean $\hat{R}$, and mean ESS of all types of loadings, averaged across each of primary loadings, cross-loadings, and the remaining loadings in each condition.

As observed in the previous section, longer iterations improved convergence, indicated by lowered $\hat{R}$ and increased ESS, while thinning did not make any difference in convergence statistics. Despite the convergence issue of condition $N = 200$, models $N = 100$ and $N = 200$ yielded the similar estimates for each type of loadings, and in all conditions, the models recovered parameters well: the mean of primary and cross-loadings were significantly higher than that of the remaining loadings. As we set the true loadings to 0.6 with the first primary loadings on each factor constrained to 1, the minimum posterior mean of 0.723 indicated that the model successfully identified primary loadings, cross-loadings the remaining loadings (maximum posterior mean = 0.051). In the models $N = 200$, relatively higher $\hat{R}$ and smaller ESS resulted from some of the problematic parameters ($\hat{R} = 6.02$, see Table 3.1), which were revealed to be loading parameters for S = 2 (three-factor EFA model) regardless of the type of loadings. The identical pattern was observed for all conditions in $N = 200$ with the following loadings: primary loadings for items 2, 3, 5, and 6 in S = 2, cross-loadings and remaining loadings for items 1 - 8 in S = 2, and some of the remaining loadings on factor 3. Interestingly, most of them were loadings on the first and second factors. Moreover, in the model with iterations 25,000, non-convergence of loading for items 1 - 4 was found in S = 1 (two-factor EFA), which loaded on factor 2. For the models $N = 100$, all loading parameters in S = 1 and S = 2 were converged with highest $\hat{R} = 1.20$ and sufficiently large ESS. The findings for convergence of loadings corroborated the previous result that the model with $N = 100$ and 50,000 iterations outperformed the other models in parameter recovery and convergence statistics. A table of full parameter estimates with convergence statistics of the model $N = 200$ with 50,000 iterations and thin = 1 is provided in Appendix 5.1.

To delve into the convergence issue of loadings, the models with $N = 200$ were further examined using trace plots and density plots. As an example, trace plot and density plot for the problematic loading that showed the highest $\hat{R}$ in the model $N = 200$ with iterations of 50,000 and thinning of 1 are depicted in Figure 3.2. It is a cross-loading of item 5 and measured by factor 1 in state 2. The left panel of

Table 3.2: Convergence statistics for loading parameters

Model Iterations/Burn-in/ Thinning	Primary loading			Cross-loading			Remaining loading		
	$M(SD)$	Mean $\hat{R}$	Mean ESS	$M(SD)$	Mean $\hat{R}$	Mean ESS	$M(SD)$	Mean $\hat{R}$	Mean ESS
N = 100									
50,000 / 25,000 / 1	0.888 (0.184)	1.03	1977.92	0.745 (0.250)	1.09	232.00	0.003 (0.165)	1.05	453.39
50,000 / 25,000 / 2	0.888 (0.184)	1.03	1490.08	0.745 (0.250)	1.09	227.00	0.003 (0.165)	1.05	426.09
25,000 / 12,500 / 1	0.876 (0.183)	1.05	265.54	0.723 (0.235)	1.09	47.25	0.002 (0.161)	1.04	389.35
25,000 / 12,500 / 2	0.876 (0.183)	1.05	260.85	0.723 (0.235)	1.09	47.00	0.002 (0.161)	1.04	408.96
N = 200									
50,000 / 25,000 / 1	0.912 (0.137)	1.69	198.69	1.002 (0.198)	1.41	54.00	0.051 (0.175)	1.92	75.74
50,000 / 25,000 / 2	0.912 (0.137)	1.68	197.92	1.002 (0.198)	1.41	54.00	0.051 (0.175)	1.93	74.00
25,000 / 12,500 / 1	0.907 (0.133)	1.71	199.85	0.975 (0.222)	1.97	9.25	0.047 (0.178)	2.05	69.78
25,000 / 12,500 / 2	0.907 (0.133)	1.71	202.69	0.975 (0.222)	1.97	9.25	0.047 (0.178)	2.05	71.78

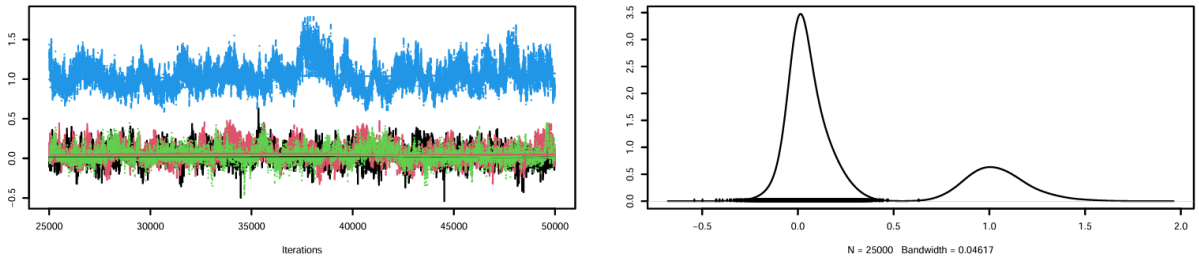


Figure 3.2: Trace plot (left) and density plot (right) of the loading that indicated the highest $\hat{R}$ ($\lambda^*_{5,1,S_2}$).

Figure 3.2 describes the trace of the 4 chains run for the model, where the x-axis represents chain lengths and the y-axis represents values of the parameter, which should be 0. No trend or fluctuations were observed, and all chains were stationary. However, the problem stemmed from the blue chain: the three chains overlapped around 0, while the blue chain was not mixed with the others, exhibiting higher values. This implies poor mixing of chains. The right panel of Figure 3.2 illustrates the density of the parameter, where the x-axis represents values that the parameter can take on, and the y-axis represents a probability. The parameter indicated a bimodal distribution, which should be a unimodal distribution centered around zero. The same pattern as Figure 3.2 was found for all of the non-converged loadings.

3.2.2 Model Comparison

Effects of Thinning Interval

Comparisons of models under the same conditions with different thinning intervals enable investigation of the effects of the thinning interval on convergence. Figure 3.3 illustrates the impact of thinning on ESS for the models $N = 100$ and $N = 200$ in the first and the second rows, respectively. The top-left and top-right panels describe the thinning effects of 50,000 and 25,000 iterations, respectively, and both showed similar results that thinning did not significantly increase ESS, as most of the parameters were on the identity line. The similar pattern was observed for $N = 200$ on the bottom-left panel for the model with 50,000 iterations and the bottom-right panel for the model with 25,000 iterations in 3.3. The wider range of ESS in the top-left and the bottom-left panels indicated that the models with longer chains produced higher ESS, as demonstrated in Table 3.1.

Effects of Chain Length

To investigate the impact of increasing chain length on convergence, models with different numbers of iterations (50,000 and 25,000) were compared with respect

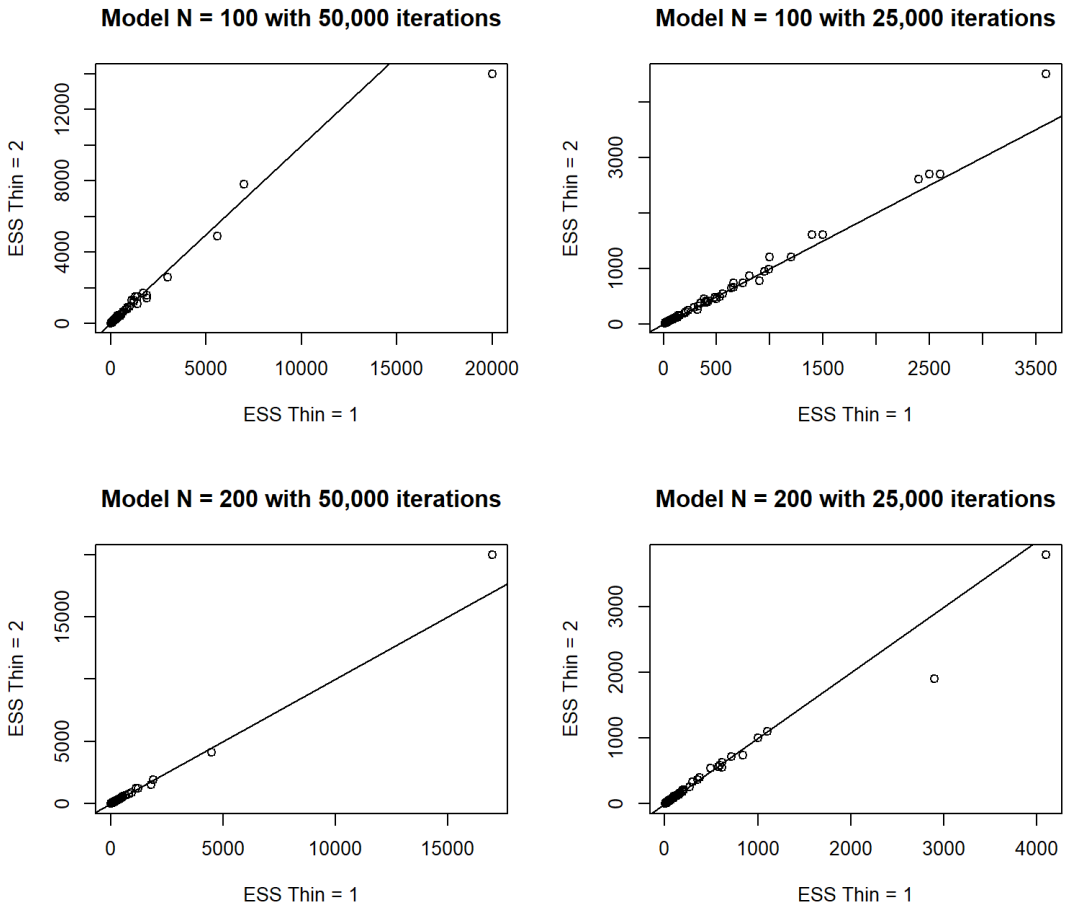


Figure 3.3: Effects of different thinning intervals on ESS

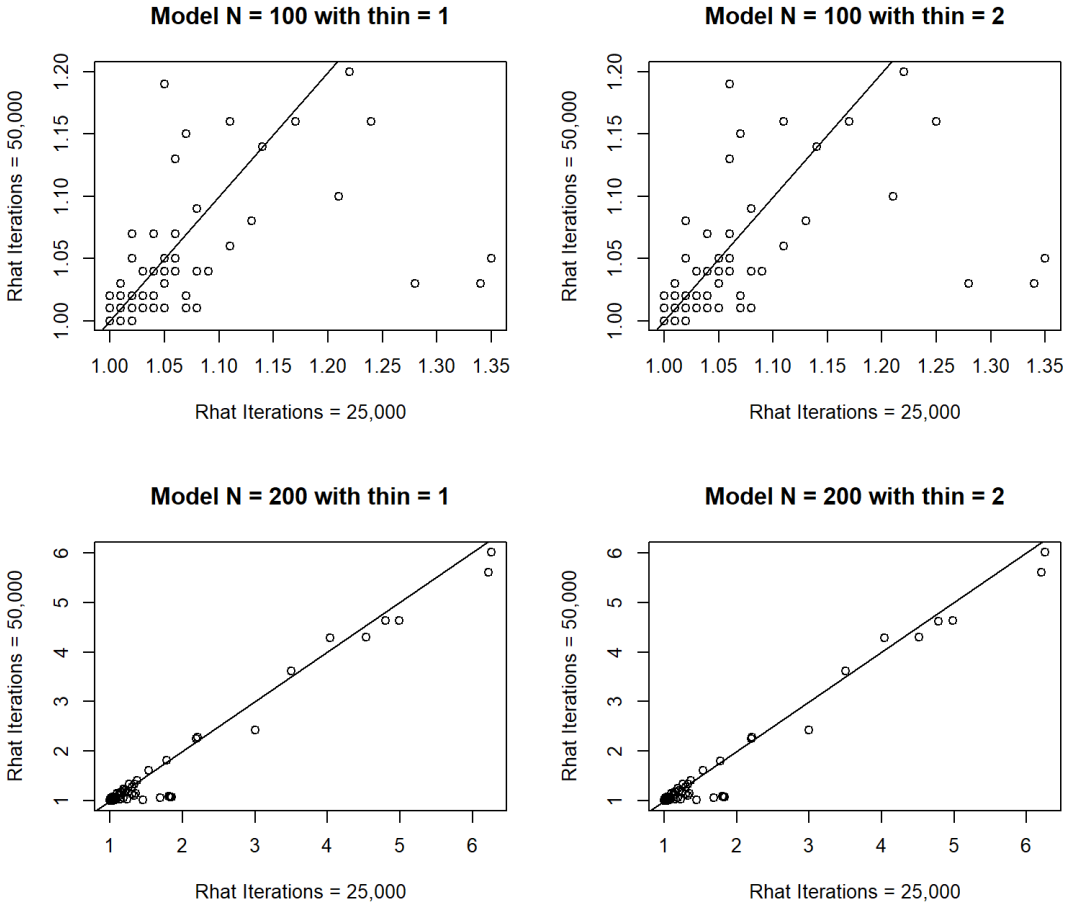


Figure 3.4: Effects of different chain lengths on $\hat{R}$

to $\hat{R}$. Figure 3.4 depicts how $\hat{R}$ changes with increased length of iterations. The top-left and top-right panels in 3.4 show the plots for the models $N = 100$ thinned by 1 and 2, respectively. As suggested that thinning did not work for convergence, two plots indicated good convergence and were exactly the same with the pattern that most of the parameters were centered around the identity line with a few of them rightward deviated from the line. This implies that parameters with higher $\hat{R}$ (e.g., $\hat{R} = 1.3$) exhibited lower $\hat{R}$ when running longer iterations. Similar results were observed in the bottom-left and bottom-right panels for the models $N = 200$ with thin 1 and 2, respectively in 3.4, which is no difference was found between the models thinned by 1 and 2. Although it was not clearly observed in the second row of 3.4 due to very high $\hat{R}$ for loading parameters, a few of the remaining parameters shifted to the right of the line around 1.5 - 2 of $\hat{R}$ on the x-axis showed $\hat{R}$ closer to 1 when length of chain was increased to 50,000. This finding is demonstrated more clearly with convergence metrics in Table 3.1.

Since it was revealed in Table 3.1 that running a model with increased iterations improved not only $\hat{R}$ but ESS as well, the effects of longer chains on ESS were

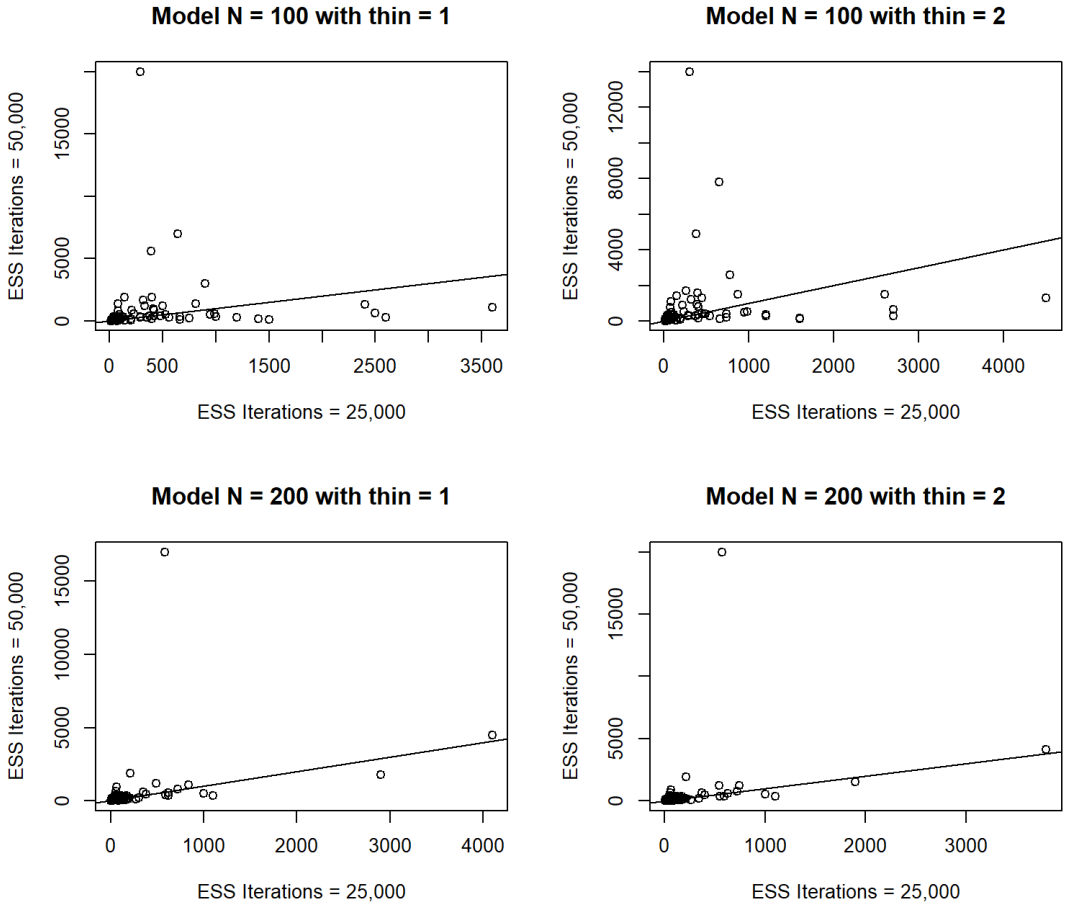


Figure 3.5: Effects of different chain lengths on ESS

further investigated. Figure 3.5 displays scatter plots that compare models with 25,000 iterations on the x-axis to 50,000 iterations on the y-axis. The first row presents the models $N = 100$, and the second row presents the models $N = 200$. The left panel depicts the models with thinning 1, and the right panel depicts the models with thinning 2. Comparisons of the plots on the left panels with the ones on the right panels showed exactly the same results that thinning did not increase ESS, which was expected from the previous findings. For both models $N = 100$ and $N = 200$ (the first and the second rows), identical patterns were observed: Higher ESS values for the models with longer iterations were expressed by the identity lines strongly tilted to the x-axes and a wider range of ESS on the y-axes (0 - 15,000) while the range of x-axes was relatively narrow (0 - 4,000). The results again supported the previous findings that longer iteration efficiently improved $\hat{R}$ as well as ESS under all conditions.

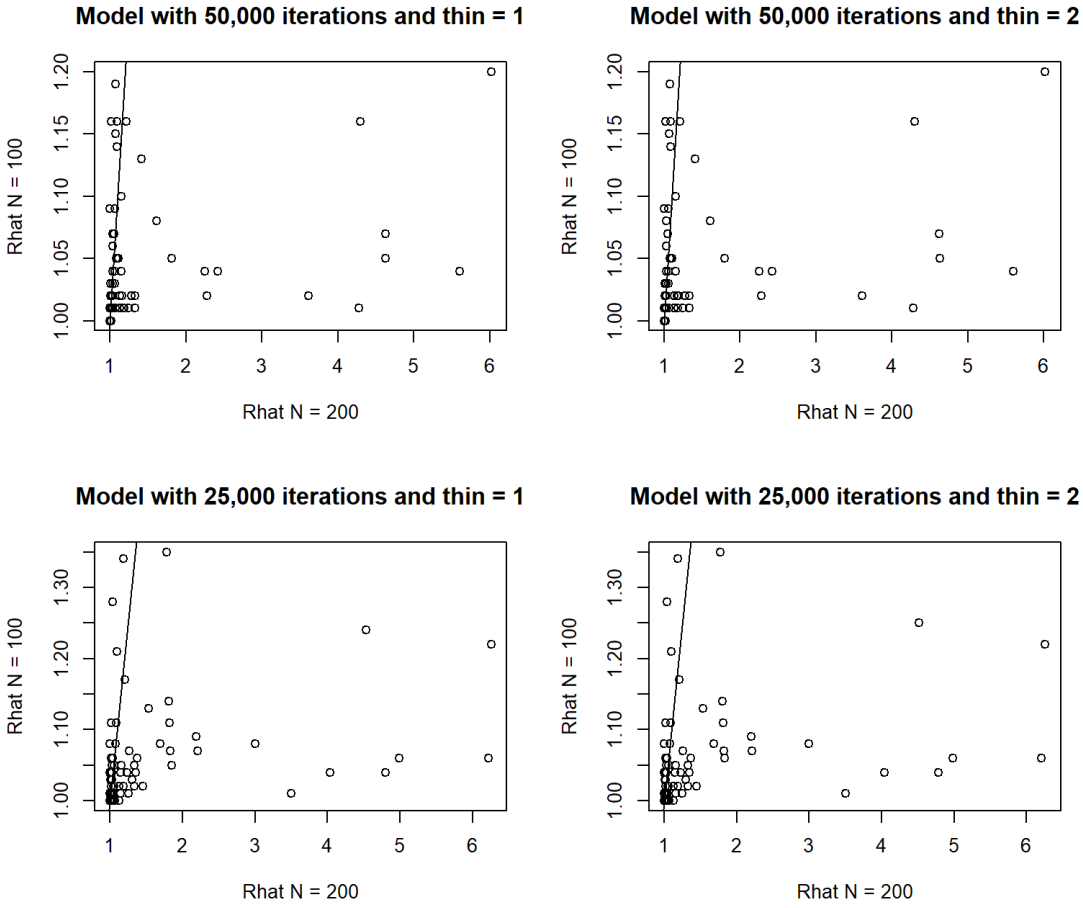


Figure 3.6: Effects of different sample sizes on $\hat{R}$

Effects of Sample Size

The models with the same number of iterations/burn-in and thinning were compared for $N = 100$ and $N = 200$ to examine the impact of a larger sample size with respect to $\hat{R}$ on convergence. Figure 3.6 shows how $\hat{R}$ changes with the increase of sample size $N = 100$ to $N = 200$. The first row illustrates models with 50,000 iterations, and the second row illustrates models with 25,000 iterations. The scatter plots on the left panels (thinning = 1) and on the right panels (thinning = 2) were identical, which indicated no effect of thinning intervals on convergence. In all plots in Figure 3.6, $\hat{R}$ for models $N = 200$ on the x-axes ranged from 1 - 6, while $\hat{R}$ ranged from 1 - 1.3 for models $N = 100$ on the y-axes, providing evidence of non-convergence of models with larger sample size. A few points scattered around $\hat{R}$ value 1 - 6 and the inclined identity lines to the y-axes with very steep slope corroborated the failure of convergence for models $N = 200$. Thus, from the scatter plots comparing $N = 100$ and $N = 200$, it was evident that the models with $N = 100$ indicated better convergence.

3.2.3 Performance Evaluation

In addition to the investigation on recovery of loading parameters, recovery of state parameters was assessed under all conditions, and performance of the models was evaluated via confusion matrix and RMSE. More precisely, the models' capabilities of correctly classifying predicted states were evaluated using confusion matrices, and RMSE was used to quantify deviations of estimated switching time points from true values. Table 3.3 lists confusion matrices for all conditions, where rows represent true states of individuals at the last time point $S = 1$ and $S = 2$ (i.e., actual switchers), and columns represent predicted states of individuals $\hat{S} = 1$ and $\hat{S} = 2$ (i.e., predicted switchers). Comparison of tables on the left panels to tables on the right panels indicated that varying thinning intervals had no effect on the prediction of states. Under the condition $N = 100$ (Table (a), (b), (c), and (d)), running with longer chain did not make any difference in accuracy of state recovery, yielding the same results across all models $N = 100$, whereas in model $N = 200$ (Table (e), (f), (g), and (h)), models with different iterations showed slightly different results. To measure the precision of prediction, the accuracy and specificity of confusion matrices across all conditions were summarized in Table 3.4, with performance metrics for predicted switching time points. Accuracy is defined as the sum of True Positives (TP) and True Negatives (TN) divided by the total sample size. Specificity is defined as TN divided by actual switchers S = 2 (i.e., TN + False Positives (FP)). As performance metrics, we considered specificity along with accuracy, because we aim to specifically examine the models' abilities to recover true switchers with high precision, which represents correct classification of true switchers into predicted switchers (i.e., a large number of TN).

Both the models $N = 100$ and $N = 200$ gave similar levels of accuracy (0.89 and 0.915, respectively). For the models with $N = 100$, a low level of specificity (0.267) was observed, due to the small number of predicted switchers (four switchers), which should be 15. This means that only four switchers were correctly classified as switchers and the other were classified as non-switchers, while for the models $N = 200$, 33 switchers out of 44 (true switchers) were correctly predicted with a specificity of 0.795 for the model with 50,000 iterations and 0.75 for 25,000 iterations. To summarize, despite the convergence issue of the models $N = 200$, all models showed higher performance in state parameter recovery, which was expected as models with more data typically produce more stable predictions.

Next, performance in the prediction of switching time points was assessed via RMSE. Table 3.4 lists RMSE and mean of predicted switching time points for each condition, and mean of true switching time points for data $N = 100$ and $N = 200$ extracted from the data generation process. As with the prediction of states at

Table 3.3: Confusion matrices

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	85	0
$S=2$	11	4

(a) Model N = 100 with iterations = 50,000, thin = 1

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	85	0
$S=2$	11	4

(b) Model N = 100 with iterations = 50,000, thin = 2

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	85	0
$S=2$	11	4

(c) Model N = 100 with iterations = 25,000, thin = 1

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	85	0
$S=2$	11	4

(d) Model N = 100 with iterations = 25,000, thin = 2

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	148	8
$S=2$	9	35

(e) Model N = 200 with iterations = 50,000, thin = 1

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	148	8
$S=2$	9	35

(f) Model N=200 with iterations = 50,000, thin = 2

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	150	6
$S=2$	11	33

(g) Model N = 200 with iterations = 25,000, thin = 1

	$\hat{S}=1$	$\hat{S}=2$
$S=1$	150	6
$S=2$	11	33

(h) Model N = 200 with iterations = 25,000, thin = 2

Note. $S = 2$ (i.e., actual switchers) represents true states of individuals at the last time point, and $\hat{S} = 2$ represents predicted switchers.

the last time points, thinning intervals did not affect the estimation of switching time points. For the models with $N = 100$ and $N = 200$, RMSE was slightly higher in models with longer iterations (0.169 and 1.689, respectively) compared to the models under the same condition with 25,000 iterations (0.116 and 1.518, respectively). The models $N = 100$ produced lower RMSE (0.169 and 0.116) than the models with $N = 200$ (1.689 and 1.518). However, RMSE for $N = 100$ was calculated from only four switchers, the values were not reliable as suggested by low specificity (0.267). Despite the small number of predicted switchers, the means of predicted switching points (8.818 and 8.715) were centered around the true mean value (8.45). The same pattern was observed for the models $N = 200$, where means of estimated switching point (8.703 and 8.244) were close to the true mean (8.19), from which deviated 1.689 and 1.518 time points on average (for the model with 50,000 and with 25,000 iterations, respectively).

In summary, the simulation study demonstrated that among all conditions, the model $N = 100$ with longer iterations showed the best convergence statistics (the

Table 3.4: Accuracy and specificity of states and RSME of switching time points

Model	$\hat{S}$		$\hat{S}_t$		S_t
Iterations/Burn-in Thinning	Accuracy	Specificity	RMSE	M	M
N = 100					
50,000/25,000/1	0.890	0.267	0.169	8.818	8.45
50,000/25,000/2	0.890	0.267	0.169	8.818	
25,100/12,500/1	0.890	0.267	0.116	8.715	
25,100/12,500/2	0.890	0.267	0.116	8.715	
N = 200					
50,000/25,000/1	0.915	0.795	1.689	8.703	8.19
50,000/25,000/2	0.915	0.795	1.689	8.703	
25,100/12,500/1	0.915	0.750	1.518	8.244	
25,100/12,500/2	0.915	0.750	1.518	8.244	

Note. Accuracy and specificity between estimated states $\hat{S}$ and true states S, and RMSE between estimated average switching time points $\hat{S}_t$ and true switching time points S_t. M indicates mean values.

lowest $\hat{R}$ and the highest ESS), while no effects of different thinning intervals were found across models. Moreover, it was found in all conditions that running longer iterations significantly improved the convergence of the models as well as accuracy and specificity for the estimation of state transitions. In the following chapter, a model with empirical data was implemented with the MCMC parameters found to be optimal for convergence in the simulation study (i.e., 50,000 iterations with 25,000 burn-in).

Chapter 4

Empirical Example: Hamilton Rating Scale for Depression

In order to reveal how DLC-SEM with regularized EFA performs and behaves on real-world datasets, we apply the model to an empirical dataset: the Hamilton Rating Scale for Depression. The results from the simulation study are used to determine MCMC parameters to improve convergence of the model. This chapter is structured as follows: we provide a detailed explanation of the dataset used, then perform EFA to discover the most appropriate model for the data at hand. This chapter ends with an interpretation of the model selected, focusing on the model's capability to capture unobserved heterogeneity that leads to changes in underlying factor structures.

4.1 Method

4.1.1 Data

Measures

The Hamilton Depression Rating Scale (HAMD), also known as HDRS, is the most widely used rating scale to assess depression symptoms. The HAMD, published by Hamilton (1960), was devised for patients to assess the effects of antidepressants and has been used as a gold standard for assessment of depression severity (Bagby et al., 2004; Hamilton, 1960). The HAMD is a clinician-rated depression scale, which means it is administered by clinicians through structured or semi-structured interviews (Gibbons et al., 1993). In other words, information necessary for scoring the scale is obtained and recorded by clinicians (Hamilton, 1960). Many versions of the HAMD exist with different numbers of items, for example, the original scale with 17 items, the shortest version with 6 items (HDRS-6), and HDRS-21 (21

items). In this work, the HAMD with 24 items was used. Each item of the HAMD assesses symptoms of depression (e.g., suicide, weight loss, and insomnia). Items are assigned to factors such as an anxiety, a general depression, a sleep, and a somatic symptoms depending on versions and methods of factor analysis used (Seemüller et al., 2023), because numerous studies on factor structures for the HAMD suggest different factor solutions ranging from 3- to 8-factor models, including bi-factor solution (e.g., Bagby et al., 2004; Gibbons et al., 1993; Goldberger et al., 2011; Hamilton, 1960; Nixon et al., 2020; Pancheri et al., 2002; Shafer, 2006). The items of HAMD-24 are rated on three-, four-, and five-point Likert scales, with $'0'$ being absence of symptoms and $'4'$ being incapacitating symptoms. The responses are summed over all items, and levels of severity of depression are indicated by higher total scores.

Participants

220 participants were asked to complete HAMD on a weekly basis for 13 weeks. The number of participants decreased over the sessions, from 220 in week 1 to 177 in week 12, and increased to 195 in the last session. The participants consisted of 144 females and 76 males. Age of the participants ranged from 60 - 92 with a mean age of 73.01 ($SD = 7.77$).

4.1.2 Procedure

With the optimal condition for convergence revealed in the simulation study, we applied DLC-SEM with regularized EFA to the HAMD dataset. EFA was conducted for the first and the last time point to explore the dimensionality of the HAMD over time. On the basis of EFA results, a model was formulated and tested with competing models using DIC. Finally, parameter estimates were interpreted for the selected model.

The HAMD data used for analyses was scaled using the mean and standard deviation at $t = 1$. For the model formulation, we employed the same model and prior specifications presented in the simulation study 4.2.2. All the analyses were implemented in JAGS in R (Version 4.4.2), using R packages `R2jags` (Version 0.8-9), `psych` (Version 2.4.12), and `GPArotation` (Version 2024.3-1).

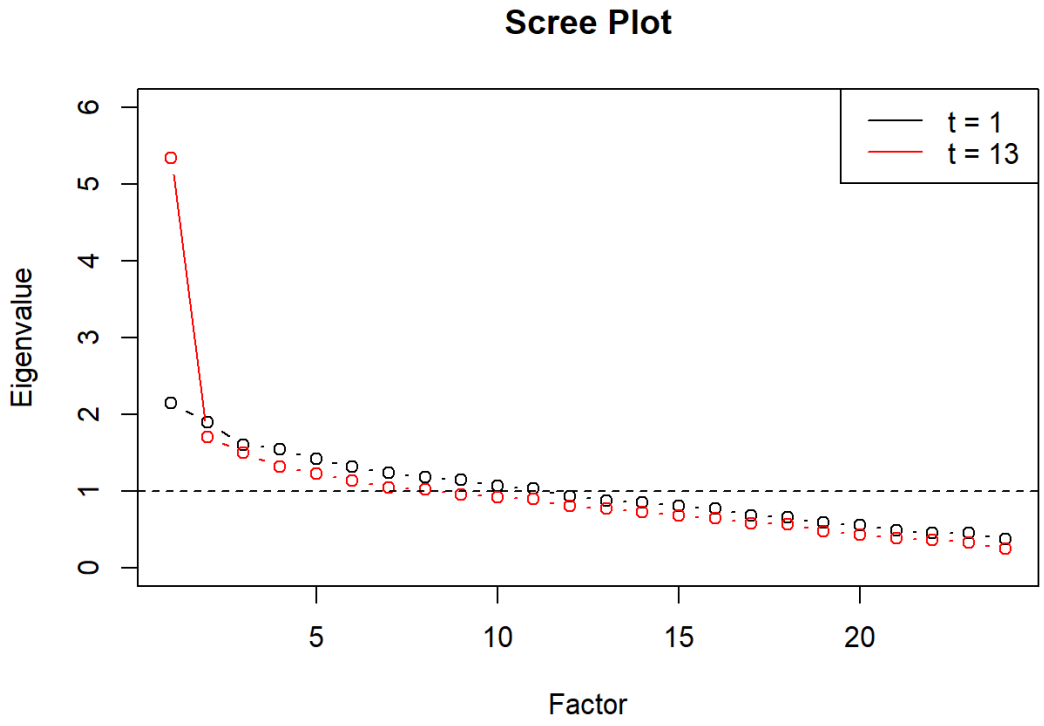


Figure 4.1: Scree plot for $t = 1$ and $t = 13$

4.2 Results

4.2.1 Exploratory Factor Analysis

Examination of Dimensionality Changes Over Time

Before fitting the DLC-SEM model to HAMD data, it is necessary to specify an EFA model for each state that best describes the HAMD data. Therefore, as a first step, we performed EFA to explore the underlying factor structure of HAMD.

To determine the number of factors, we examined the eigenvalues of HAMD for the first ($t = 1$) and the last time points ($t = 13$). Figure 4.1 illustrates the scree plot of HAMD for the two time points, where the x-axis represents eigenvalues and the y-axis represents the number of factors. In $t = 1$, variances were explained by a relatively larger number of factors than in $t = 13$, whereas in $t = 13$, a substantial portion of the variances were explained by the first factor. These findings imply that the dimensionality of HAMD factor models changes over time.

Assumption for Unobserved Heterogeneity

The changes in dimensionality found in Figure 4.1 were modeled via latent states. Adopting the DLC-SEM framework with two latent states in the section 4.2.2,

models for $t = 1$ and for $t = 13$ were specified in $S = 1$ and $S = 2$ respectively, describing changes in HAMD factor structures over time. We assumed that the unobserved heterogeneity that causes dimensionality changes was reconceptualization (i.e., changes in definitions of latent constructs), which is one of the three types of response shift proposed by Schwartz and Sprangers (1999). Response shift refers to changes in one's self-evaluation of a target construct Schwartz and Sprangers (1999). As a meta-construct, response shift consists of three inter-related concepts: (a) recalibration, which is a change in one's interpretation or internal standards of response category values (Oort, 2005), (b) reprioritization, also known as beta change, a change in the importance of items or in a respondent's values (e.g., increased factor loading value Golembiewski et al., 1975; Oort, 2005), and (c) reconceptualization, also referred to as gamma change, a redefinition of target constructs Schwartz and Sprangers (1999). Among those, we focused on reconceptualization because it was demonstrated as an indicative of factor structure changes (Golembiewski et al., 1975; Oort, 2005). That is, one's conceptualization of latent variables is reflected by changes in patterns and/or magnitude of loadings over time (Schwartz et al., 2004).

Response shift is the concept for respondents. We applied the concept to clinicians and regarded it as a rater's behavior, as clinicians interpret and record answers provided by patients. Since the HAMD heavily relies on expertise and skills of interviewers (Hamilton, 1960; Potts et al., 1990), it is not unreasonable to assume that response shift (e.g., a rater's subjective interpretation or redefinition of items or answers) could occur even for a trained clinician with psychiatric backgrounds.

Other possible sources of unobserved heterogeneity that lead to dimensionality changes might include response style (RS Vogelsmeier et al., 2019), differential rater functioning (DRIFT Rudner, 1991; Wolfe et al., 2001), such as a rater's fatigue, halo effect, or leniency/stringency error. However, we considered that changes in an interviewer's understanding of a target construct over time were the most plausible cause of dimensionality changes, more specifically, convergence of factor structures as demonstrated by Oort (2005).

EFA with Factor Rotations

To determine the number of factors to retain, we conducted the parallel analysis for $t = 1$ and $t = 13$ separately. 3-factor and 6-factor solutions were suggested for $t = 1$ and $t = 13$, respectively. With this, our assumptions are formulated: The dimensionality of the HAMD changes from 6 factors to 3 factors over time due to a rater's reconceptualization of the HAMD factors. Figures for the result of

parallel analyses for the two time points can be found in Appendix 5.1.

Items that did not load on any factors at both time points or that are a factor itself were deleted, resulting in 17 items. Based on the result of parallel analysis, EFA was performed for $t = 1$ and $t = 13$ separately. Table 4.1 shows the 3-factor solution for $t = 13$. The oblimin rotation was used for EFA. All loadings below 0.3 were deleted except for the items 'Insomnia (Late)' (0.280) and the cross-loaded item 'Diurnal Variation' (0.442 and -0.398). Those items were retained because both were associated with sleep disturbance and circadian rhythm and appeared to be good proxies of the respective factor. The clear 3-factor structure was observed for $t = 13$, where each factor can be labeled as 'Depression/Anxiety', 'Somatic Symptoms', and 'Insomnia', with the labels of the factors determined by the contents of the items being measured. Although several factor solutions for HAMD were proposed with no consensus on a unified solution, patterns that are consistent with prior studies were observed. According to Bagby et al. (2004), a combination of 'Mood', 'Suicide', and 'Psychic Anxiety' was found to load on the same factor in previous studies several times, and a combination of 'Mood', 'Guilt', and Suicide' as well , which might be indicative of a general depression factor. Bagby et al. (2004) also found the insomnia items consistently on the same factor, indicating 'Sleep Disturbance' factor, while in this study, the factor was labeled as 'Insomnia', including the item 'Diurnal Variation'. It is interesting to note that the three factors extracted were very similar to the four-factor solution found by Shafer (2006), where 'Depression', 'Anxiety', 'Insomnia', and 'Somatic' were suggested.

For $t = 1$, EFA was performed with the target rotation method in order to reflect our hypothesized factor pattern (Darton, 1980): Factor structures of HAMD converge over time from the 6-factor model ($S = 1$) to the 3-factor model ($S = 2$) due to a rater's redefinition of HAMD factors. Therefore, a target matrix was specified based on the 3-factor solution for $t = 13$ by stretching the 3-factor solution to the 6-factor, so that the 6-factor solution is rotated towards the 3-factor loading pattern specified below:

Table 4.1: EFA 3-factor solution

	Depression /Anxiety (Factor 1)	Somatic symptoms (Factor 2)	Insomnia (Factor 3)
Mood	**0.81**	0.01	0.08
Guilt	**0.51**	-0.06	-0.06
Suicide	**0.48**	-0.03	0.04
Work Activities	**0.54**	0.27	0.08
Anxiety (Psychic)	**0.67**	0.00	-0.15
Genital	**0.45**	0.14	-0.05
Helplessness	**0.46**	0.16	0.18
Hopelessness	**0.71**	0.00	0.04
Worthlessness	**0.59**	-0.10	-0.07
Anxiety (Somatic)	0.03	**0.37**	0.11
Somatic (Gastro)	0.03	**0.62**	0.10
Somatic (General)	0.17	**0.45**	0.15
Weight	-0.05	**0.67**	-0.17
Insomnia (early)	0.20	0.08	**0.43**
Insomnia (Middle)	0.34	-0.04	**0.32**
Insomnia (Late)	0.34	-0.04	**0.28**
Diurnal Variation	0.44	0.04	**-0.40**

Note. The loadings larger than 0.30 are highlighted in boldface.

$$\begin{bmatrix} ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ ? & 0 & 0 & 0 & 0 & 0 \\ 0 & ? & 0 & 0 & 0 & 0 \\ 0 & ? & 0 & 0 & 0 & 0 \\ 0 & ? & 0 & 0 & 0 & 0 \\ 0 & ? & 0 & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 & 0 \end{bmatrix}, \tag{4.1}$$

where entries '?' are assumed to be substantially large non-zero loadings. Each row corresponds to the items in Table 4.1 in the same order. The target matrix kept the clear 3-factor structure while loadings below 0.3 were forced to zero. Equation 4.1 describes our assumption on factor structure for $t = 1$ that all items load on the first three factors with substantially large loadings, while the factors 4 - 6 consist of zero or negligible loadings (i.e., cross-loadings) and are subsequently collapsed into the factors 1 - 3.

Table 4.2 shows the EFA result with target rotation for $t = 1$. Compared to the 3-factor model for $t = 13$, some of the loadings shifted and were assigned to factors 4 - 6 that should contain values near zero, cross-loaded more than one factor, or did not load on any factor. This yielded a more diffuse loading pattern than the target matrix, making it harder to interpret than in $t = 13$.

4.2.2 Model Selection

Competing Models

The identical model and prior specifications presented in were applied, but with slight adjustments for model structure in order to represent our assumption on change in factor structures: The latent factors were shared across the two states, not specified independently for each state. That is, the same factors were applied across the two states that do not depend on the state of person i at time t, and

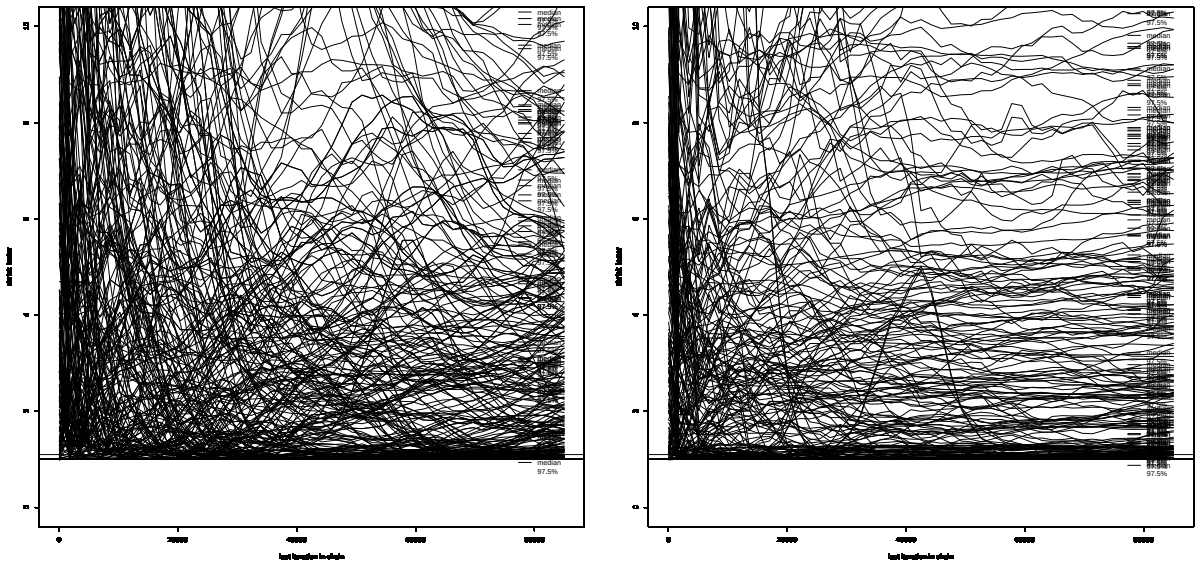
Table 4.2: EFA 6-factor solution

	Depression /Anxiety (Factor 1)	Somatic symptoms (Factor 2)	Insomnia (Factor 3)	Factor 4	Factor 5	Factor 6
Mood	**0.51**	0.06	**-0.40**	-0.14	-0.12	**-0.37**
Guilt	**0.36**	-0.04	-0.11	**0.55**	**0.45**	0.16
Suicide	0.27	0.09	-0.05	-0.16	0.13	-0.01
Work Activities	0.01	-0.12	0.06	0.20	-0.17	**-0.38**
Anxiety (Psychic)	0.21	0.16	-0.19	0.28	**-0.31**	0.18
Genital	0.15	-0.09	-0.02	-0.25	0.27	-0.29
Helplessness	**0.33**	0.07	0.25	0.20	-0.21	-0.20
Hopelessness	**0.57**	-0.00	0.25	-0.28	-0.13	0.20
Worthlessness	**0.37**	-0.21	0.17	-0.06	0.09	0.25
Anxiety (Somatic)	-0.12	0.08	-0.23	-0.02	-0.25	0.27
Somatic (Gastro)	0.03	**0.57**	-0.03	-0.03	-0.14	-0.17
Somatic (General)	-0.15	0.03	0.06	-0.01	0.01	-0.23
Weight	-0.01	**0.89**	0.11	0.01	0.21	0.09
Insomnia (early)	0.07	0.01	-0.15	-0.20	-0.07	0.04
Insomnia (Middle)	-0.02	-0.06	-0.12	-0.03	0.29	0.03
Insomnia (Late)	-0.05	-0.04	**-0.46**	-0.06	0.03	0.03
Diurnal Variation	0.12	-0.09	-0.22	0.17	-0.09	-0.00

Note. The loadings larger than 0.30 are highlighted in boldface.

thus to specify a three-factor model for $S = 2$, the items load on only the first three factors. By doing so, it allows for the identical interpretation of factors across the states and accounts for a collapse of the factors over time. Apart from the adjustment for the model structure, stronger shrinkage was specified for the prior for cross-loadings to improve convergence.

To ensure that the model (Model A) best describes the HAMD data, two competing models were suggested. The first competing models (Model B) assumed the state-specific factors conditioning on states for subject i at time t as in where interpretations of the factors were no longer identical, and the second (Model C) was the same as Model A, but individual- and time-specific states were sampled backwards. Whereas in Model A, states at $t = 1$ were forced to 1 for all individuals and depended on states at previous time points (S_{t-1}), Model C assumed $S = 2$ for the last time point and states at time t (S_t) were estimated using the states at next time points (S_{t+1}) to check whether a subject i started from state 1 at the



(a) Gelman-Rubin plot for correlated DLC-SEM

(b) Gelman-Rubin plot for uncorrelated DLC-SEM

Figure 4.2: Gelman-Rubin plots

beginning of the session. The three models were compared via deviance information criterion (DIC Spiegelhalter et al., 2002).

Model Comparison and Selection

Model A was selected with the lower DIC (124548.1) than Model B (3830857) and C (205085.1). Model B produced the highest DIC, supporting that the application of the same factors is reasonable for this data. However, Model A showed non-convergence, we tried the same model with the uncorrelated factors by applying a univariate normal distribution to each factor, because most off-diagonal elements of the factor covariance matrix yielded values of zero or nearly zero. Figure 4.2 illustrates Gelman-Rubin plots to diagnose convergence of the models. Figure 4.2a and 4.2b assess convergence of Model A and Model A with uncorrelated factors, respectively, where x-axes represent chain length, and y-axes represent $\hat{R}$. The models were run for 85,000 iterations for two chains.

Both models strongly indicated the failure of convergence, and the same patterns were observed that for many parameters $\hat{R}$ did not show downward trends, remaining high throughout the sampling process, while relatively more parameters converged in 4.2b. The model with uncorrelated factors produced significantly lower DIC 101923.7 than the correlated factor model (124548.1). Thus, the uncorrelated factor model was preferred over the correlated factor model, despite the failure of convergence. As Kruschke (2014, p. 178) stated that MCMC chains of models with higher complexity are more likely to be problematic, the failure of

Table 4.3: DIC for competing models

Model	DIC
DLC-SEM with correlated factors	124548.1
DLC-SEM with uncorrelated factors	101923.7
Correlated 6-factor DSEM	112319.3
Correlated 3-factor DSEM	111491.4
Uncorrelated 6-factor DSEM	119171.5
Uncorrelated 3-factor DSEM	111766.6

convergence of the two models may be attributable to the high complexity of the DLC-SEM models that involve a large number of parameters as well as complex model structures. Moreover, the non-convergence of DLC-SEM models might be indicative of inadequately defined latent states (Faleh et al., 2025). Thus, we tested models with no latent states (i.e., DSEM). Accordingly, our assumption on unobserved heterogeneity was changed: The factor structure of HAMD was assumed to be stable over time. In other words, it was assumed that changes in factor structures due to a rater's reconceptualization of factors did not occur. DSEM models varied across (a) correlation of factors and (b) the number of factors, resulting in $2 \times 2 = 4$ models: uncorrelated 6-factor DSEM, uncorrelated 3-factor DSEM, correlated 6-factor DSEM, and correlated 3-factor DSEM. Table 4.3 lists DIC values of all six models including the two DLC-SEM models with (un)correlated factors.

Among all models, DLC-SEM with uncorrelated factors yielded the lowest DIC, followed by correlated 3-factor DSEM and uncorrelated 3-factor DSEM. The 6-factor models indicated larger DIC than their respective 3-factor models, providing evidence that 3-factor models fit the HAMD data better and are more suitable for the data. This finding was corroborated by convergence statistics of 6- and 3-factor DSEM models. The 6-factor DSEM models did not converge, which were indicated by convergence statistics: median $\hat{R} = 1.26$, maximum $\hat{R} = 12.41$, and median ESS = 16 for correlated 6-factor DSEM, and median $\hat{R} = 1.24$, maximum $\hat{R} = 12.28$, and median ESS = 18 for uncorrelated 6-factor DSEM. In contrast, two 3-factor DSEM models showed good convergence: maximum $\hat{R} = 1.03$, minimum ESS = 95, and median ESS = 630 for correlated 3-factor DSEM, and maximum $\hat{R} = 1.01$, minimum ESS = 350, and median ESS = 3800 for uncorrelated 3-factor DSEM.

Correlated and uncorrelated 3-factor DSEM models were similar with respect

to loading parameters, in addition to good convergence and lower DIC. The means with standard deviation of primary loadings and cross-loadings were $M = 0.739$ ($SD = 0.054$) and $M = 0.16$ ($SD = 0.102$) for the correlated 3-factor DSEM, respectively, and the uncorrelated 3-factor DSEM model yielded $M = 0.695$ ($SD = 0.045$) for primary loadings and $M = 0.127$ ($SD = 0.071$) for cross-loadings. Both models produced substantially larger primary loadings than zeros, and cross-loadings were closer to zeros, indicating good representation of the data. Although correlated 3-factor DSEM indicated the lowest DIC among the four DSEM models, uncorrelated 3-factor DSEM was selected. It might be more reasonable to assume that depression factors are correlated each other as reported in previous research on the HAMD (e.g., Pancheri et al., 2002; Seemüller et al., 2023). However, given that uncorrelated 3-factor DSEM fitted the data as good as correlated 3-factor DSEM, we selected uncorrelated 3-factor DSEM for comparison with uncorrelated DLC-SEM and for analysis purposes. The selection of uncorrelated 3-factor DSEM as a final model confirms the assumption that dimensionality of HAMD is stable over time and therefore latent state switches are not needed for this dataset.

4.2.3 Parameter Estimates

Parameter Estimates of Uncorrelated 3-factor DSEM

As revealed in the previous section, all parameters of uncorrelated 3-factor DSEM converged with substantially high ESS, and the model yielded larger primary loadings and relatively smaller cross-loadings close to zero. In this section, we examine and interpret all parameter estimates of the uncorrelated 3-factor DSEM model. Due to space limitations, a table that lists parameter estimates is provided in Appendix 5.1. The AR coefficients indicated high values, which are $M = 1.00$ ($SD = 0.00$), $M = 0.76$ ($SD = 0.05$), and $M = 0.67$ ($SD = 0.11$) for 'Depression/Anxiety', 'Somatic Symptoms', and 'Insomnia', respectively. The first factor showed the higher lagged effect, indicating that the depression factor at $t - 1$ was a strong predictor of itself at the subsequent time point t. The mean of residual variances was high ($M = 0.706$, $SD = 0.024$), ranging from 0.44 to 1.01. The variances of the factors were small, with $M = 0.07$, $SD = 0.003$. The similar pattern was found for the variances of person-specific random intercepts for each factor ($M = 0.15$, $SD = 0.027$), which indicates that the variability of individuals' factor scores across all time points were moderate.

Parameter Estimates of Uncorrelated DLC-SEM

Although the lack of convergence of uncorrelated DLC-SEM produced rather unstable parameter estimates, as demonstrated in Figure 4.2b, parameters were interpreted for the purpose of subsequent analysis. Parameter estimates are listed in Appendix 5.1 for the reason of space limitations. In $S = 1$, cross-loadings on factors 4 - 6 ($\lambda^*_{j,4,S_1}$, $\lambda^*_{j,5,S_1}$, $\lambda^*_{j,6,S_1}$) were nearly zeros, while cross-loadings on the first three factors ($\lambda^*_{j,1,S_1}$, $\lambda^*_{j,2,S_1}$, $\lambda^*_{j,3,S_1}$) produced significantly larger values, supporting that sparse loading pattern for factors 4 - 6. In other words, the items loaded and cross-loaded on the first three factors with significantly large values, where primary loadings ($M = 0.459$, $SD = 0.125$) and cross-loadings ($M = 0.223$, $SD = 0.177$) were not distinguishable, yielding the unclear loading pattern, whereas cross-loadings on factors 4 - 6 showed sparse structure with negligible values. If the model converged, this might be indicative of the collapse of factor structures to the 3-factor solution. The residual variance was high in both states, $M = 0.691$ with $SD = 0.054$ for $S = 1$, and $M = 0.595$ with $SD = 0.058$ for $S = 2$, where smaller residual variances in $S = 2$ were expected due to a collapse of the factors for a successfully converged model.

The regression coefficients for HMM ranged from moderate to high (0.36 - 1.79), with the depression factor ($M = 1.79$, $SD = 0.27$) at $t - 1$ contributed most to the prediction of the transition from $S = 1$ to $S = 2$, indicating that the first factor was a strong predictor of state transition. The autoregressive effects for each of the six factors were mostly high (0.78 - 0.96), except for the estimates for the first factor ($M = 0.19$, $SD = 0.07$). Compared to factor 1, factors 2 - 6 were more predicted by themselves at $t - 1$.

Transitions from state 1 to 2 were examined using a state switch plot. The left panel of Figure 4.3 illustrates estimated individual- and time-specific probabilities of being in $S = 2$ in gray lines (i.e., probabilities of transition to $S = 2$) with average probabilities in a black line. The probabilities showed a rapidly increasing trend in the first 4 - 6 sessions, after which the trend became stable. This implies that all individuals starting from $S = 1$ at the beginning, most of the transitions occurred in the early phase of sessions. The average transition probability remained rather stable at around a probability of 0.5 in the later stage of sessions, as proved by state estimates extracted from the data: Among all 220 individuals, actual switchers were 115 with a transition probability of 52.27%. That is, about half of the individuals switched to state 2, and the others remained in state 1. If we assume reliable estimates with good convergence statistics, it could be interpreted that about 50% of clinicians redefined their understanding of the HAMD factors (i.e., reconceptualization response shift), resulting in dimensionality changes to a

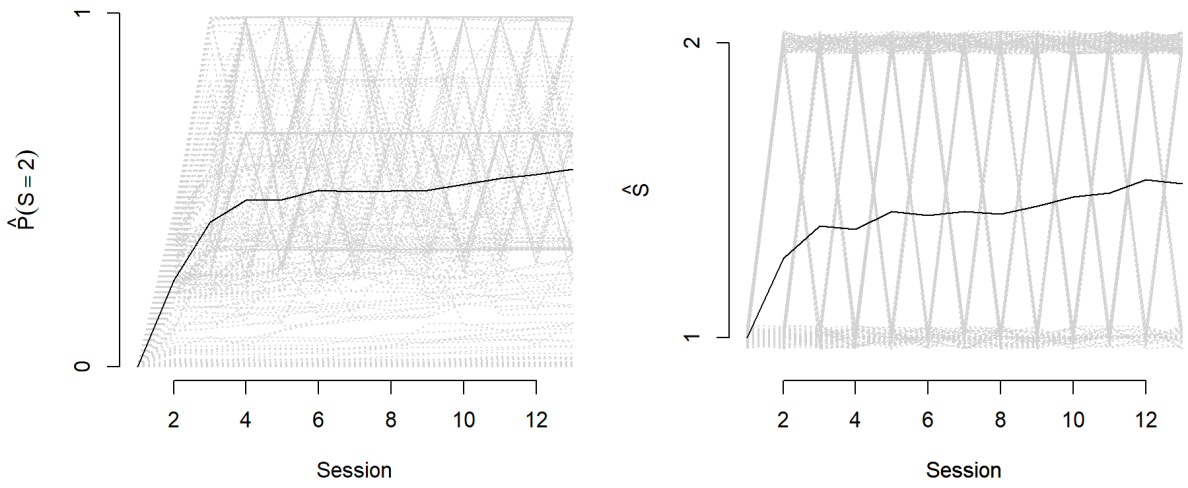


Figure 4.3: Illustration of estimated transition probabilities ($\hat{P}(S = 2)$) of being in $S = 2$ (left panel) and estimated states ($\hat{S}$) over time (right panel). Individual values in gray lines and averaged values in black lines.

clear 3-factor model.

The right panel of Figure 4.3 exhibits estimated states (i.e., latent state memberships) for subject i at measurement occasions t in gray lines, with estimated states averaged across all individuals represented in a black line. The average $\hat{S}$ increased with $\hat{P}(S = 2)$ in a similar pattern: A steep increase in $\hat{S}$ was observed in the initial stage of sessions, as indicated by the estimated average switching time point of 6.12, and exhibited a stable trend until the end of sessions. $\hat{S}$ at around 1.5 again indicated that transitions to $S = 2$ occurred with a probability of 0.5. It is noteworthy that downward gray lines going to the right were observed, implying possibilities of individuals switching back from $S = 2$ to $S = 1$. As the probability was fixed to 0.01, this was not expected and might be indicative of poor convergence of the model.

Chapter 5

Discussion

The primary objective of this work was to investigate how well the model performs in detecting changes in dimensionality due to unobserved heterogeneity using the DLC-SEM framework with regularized EFA. To this end, we performed the simulation study under different MCMC parameter settings and sample sizes to examine the optimal choice of MCMC parameters that achieves convergence. The study revealed that for our model, better convergence was achieved in the smaller sample size ($N = 100$) with longer chain and burn-in (50,000 and 25,000, respectively), indicated by lower $\hat{R}$ and higher ESS. Varying thinning intervals (1 or 2) had no effect on any of the conditions. All models with $N = 200$ indicated a very high value of maximum $\hat{R}$ (e.g., 6.02), stemming from some problematic loadings found mostly in $S = 2$ (3-factor model). Although the models $N = 200$ did not converge, all of the models successfully recovered primary loadings and cross-loadings, and loading patterns. The models were compared via scatter plots that depict the effects of thinning intervals, chain lengths, and sample size. The same results as the simulation study further support our findings that larger numbers of iterations efficiently improve convergence statistics $\hat{R}$ and ESS as well.

Next, we assessed recovery of state parameters under all conditions. In line with the previous finding in the simulation study, running models with longer iterations yielded higher accuracy in specificity across all conditions. However, specificity for models with the smaller sample was very low, which resulted from a very small number of correctly classified switches (4 switchers of 15), whereas switchers were classified with high specificity for the models $N = 200$ (33 switchers of 44). RMSE indicated that average switching time points were compared with the true value. RMSE for models $N = 100$ was very low, however, it was not a reliable estimate due to low specificity. Nevertheless, estimates yielded values close to the true parameters under all conditions.

With the capability of the model revealed, we empirically demonstrated it using the HAMD data. Investigation of eigenvalues at the first and the last time points

indicated a change in the number of factors over time. We performed EFA separately for each time point with the assumption that an interview's change in definition or understanding of the HAMD factors (i.e., reconceptualization) causes changes in underlying factor structure. That is, we assumed that reconceptualization led to the convergence of the factor structure from 6 to 3. Competing models were tested via DIC. DLC-SEM model with uncorrelated factors was selected despite the poor convergence, together with uncorrelated 3-factor DSEM, for analysis purposes. As uncorrelated 3-factor DSEM was selected with good convergence statistics, this could be interpreted that transitions of states are not assumed for the HAMD data, and implies that a rather homogeneous underlying factor structure across time is expected.

In spite of non-convergence of uncorrelated DLC-SEM, the model exhibited that around 50% of the individuals actually switched, most of them switched in the early stage of sessions (the first 4 - 6 sessions) with only a few transitions occurring in the later sessions. Stated differently, about half of the clinicians changed their understanding of the HAMD factors, leading to the collapse of factor structure from 6 to 3. However, it should be pointed out that in Figure 4.3, switches from $S = 2$ to $S = 1$ were observed several times despite the probability of 0.01. This might be due to non-convergence of estimates, and thus it should be interpreted with caution.

5.1 Limitations and Future Directions

The findings in this work point to the need for well-specified, improved models with respect to convergence diagnostics and model performance in capturing underlying dimension changes. As revealed by previous studies (Faleh et al., 2025; Kruschke, 2014), the likelihood of failure of convergence increases in highly complex models that contain many effects and numerous parameters, and in the situation where latent states do not provide a good representation of the data at hand. Therefore, we deemed the main cause of the non-convergence as the misspecification of our model. Even though uncorrelated 3-factor DSEM that assumes no state transitions was favored by the data with good convergence, it is plausible to assume the presence of an alternative model that better represents the HAMD data by considering different assumptions on unobserved heterogeneity (i.e., assuming different factor models for each state), because changes in dimensionality of the HAMD were clearly indicated by the eigenvalues at $t = 1$ and $t = 13$. Thus, in order to improve the model, the following suggestions may be taken into consideration in future research.

First, a censored distribution (e.g., truncated normal distribution) can be used

to ensure the order of values are preserved, so as to facilitate identification of discrete latent states (Kelava et al., 2022). For instance, a vector of state-specific between-level factor intercepts (denoted as α_{21s} in Kelava et al. (2022)) can be constrained by applying truncated normal prior to $\Delta\alpha_{21j(s=2)} = \alpha_{21j(s=2)} - \alpha_{21j(s=1)}$ for all factors j, so that state 2 is defined by higher scores than state 1. This can be adopted considering the context of a study, while in this work, it was not implemented as we did not assume higher or lower factor scores for a specific state.

Another aspect that might be considered to improve the model is that several parameters (e.g., α_2, ν_{1s}) were dropped in this work for the following reasons (a) estimates were mostly zeros or very close to zero, (b) parameters were constrained to zero for identification of the model, or (c) parameters were excluded from the beginning of the analyses for model simplicity and faster convergence (e.g., intercepts of EFA model). Considering the example of Kelava et al. (2022) discussed above, for latent states that are distinguished by values of factor scores, higher intercepts of factors themselves might be a strong indicative of latent states. As demonstrated, investigation on parameters excluded in this thesis might provide meaningful insight into latent states, and thus inclusion of parameters should be carefully evaluated, taking into account the context and assumptions on latent states.

Next to that, adding covariates as predictors might be one of the ways to improve model fit, if it is considered to be needed for the data at hand. In intensive longitudinal data, patients could be affected by time-varying covariates during the course of treatments (e.g., seasonal events that affect all individuals), denoted as X_{3t} in the DLC-SEM framework, or individual-specific covariates such as personality or traits (X_{2i}) that can be used to predict individual transition probabilities. Such covariates should be able to account for some portions of Y that are not explained by factor models nor by the effects the model involves (e.g., autoregressive effects of factors). Additionally, a non-linear term could be included if it is supported by theories/assumptions of the dataset at hand and by a better model fit than competing models (Kelava & Brandt, 2019).

Depending on data and research questions, different time series models could be considered. In this work, AR models with order of 1 were applied for each factor, while other time series models (e.g., MA(1), MEAR(1), or AR(2)) can also be suggested for our model. To decide the sufficient models, patterns of the autocorrelation function (ACF) are investigated. Typically, ACF illustrates an exponential decay for the AR(1), whereas the ARMA(1,1) decays slower (Asparouhov et al., 2018). For the decision between MEAR(1) and ARMA(1,1), Asparouhov et al. (2018) suggested the use of the MEAR(1) model over ARMA(1,1) for the reason of easier interpretability of the model, as measurement error is common in the social

sciences.

The HAMD data we used for the analyses consists of 17 items that differ in response categories, because the 24-item HAMD, which was our initial data, was constructed with items measured on different Likert-type response scales. More specifically, our data, after item deletion in the EFA process, resulted in 17 items: 8 items with a 3-point scale, 2 items with a 4-point scale, and 7 items with a 5-point scale. More than half of the items were measured on 3- and 4-point scales, which are not regarded as continuous (Sullivan & Artino, 2013). Since different versions of the HAMD consist of different sets of items with various response categories (e.g., HAMD-17 consists of 17 items rated on 3- and 5-point Likert scales), future research should examine response categories for each item before conducting analysis (e.g., using DLC-SEM with categorical variables). In case the model is not identified, further theoretical investigation of model identification might be meaningful to address the convergence problem and model misspecification.

Finally, in addition to respecifying the model, the model can also be improved by addressing the non-convergence problem. As discussed in Figure 3.2 in the simulation study, poor mixing of the chains was observed. To obtain well-mixed chains, an initial value of the chain can be specified manually. In our case, higher values of $\hat{R}$ were mostly found from loadings in $S = 2$ for model $N = 200$, exhibiting the non-convergence of a similar set of items across all conditions in $N = 200$. As problematic parameters are already known, one can fix a plausible initial value for parameters (for loadings, in this case) (Thero & Hernvann, 2024). By doing so, faster model convergence can be achieved with better mixing of chains (Thero & Hernvann, 2024).

References

Altman, R. M. (2007). Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association*, *102*(477), 201–210. https://doi.org/10.1198/016214506000001086

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 257–269. https://doi.org/10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? *American Journal of Psychiatry*, *161*(12), 2163–2177. https://doi.org/10.1176/appi.ajp.161.12.2163

Brandt, H., Chen, S. M., & Bauer, D. J. (2025). Bayesian penalty methods for evaluating measurement invariance in moderated nonlinear factor analysis. *Psychological Methods*, *30*(3), 482–512. https://doi.org/10.1037/met0000552

Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling Sparsity via the Horseshoe. *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, 73–80. Retrieved September 25, 2025, from https://proceedings.mlr.press/v5/carvalho09a.html

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics*, *183*(1), 31–57. https://doi.org/10.1016/j.jeconom.2014.06.008

Darton, R. A. (1980). Rotation in Factor Analysis. *The Statistician*, *29*(3), 167. https://doi.org/10.2307/2988040

Faleh, R., Morelli, S., Andriamiarana, V., Roman, Z. J., Flückiger, C., & Brandt, H. (2025, August 18). *Dynamic Latent Class Structural Equation Modeling:*

A Hands-On Tutorial for Modeling Intensive Longitudinal Data. arXiv: 2508.12983 [stat]. https://doi.org/10.48550/arXiv.2508.12983

Flückiger, C., Horvath, A. O., & Brandt, H. (2022). The evolution of patients' concept of the alliance and its relation to outcome: A dynamic latent-class structural equation modeling approach. *Journal of Counseling Psychology*, *69*(1), 51–62. https://doi.org/10.1037/cou0000555

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gibbons, R. D., Clark, D. C., & Kupfer, D. J. (1993). Exactly what does the Hamilton depression rating scale measure? *Journal of Psychiatric Research*, *27*(3), 259–273. https://doi.org/10.1016/0022-3956(93)90037-3

Goldberger, C., Guelfi, J., & Sheehan, D. (2011). Assessment of Anxiety in Clinical Trials with Depressed Patients Using the Hamilton Depression Rating Scale. *Psychopharmacology Bulletin*, *44*(3), 34–50. Retrieved September 25, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC5044547/

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1975). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, *12*(2), 133–157. https://doi.org/10.1177/002188637601200201

Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery & Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56

Kelava, A., & Brandt, H. (2019). A Nonlinear Dynamic Latent Class Structural Equation Model. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 509–528. https://doi.org/10.1080/10705511.2018.1555692

Kelava, A., Kilian, P., Glaesser, J., Merk, S., & Brandt, H. (2022). Forecasting Intra-individual Changes of Affective States Taking into Account Inter-individual Differences Using Intensive Longitudinal Data from a University Student Dropout Study in Math. *Psychometrika*, *87*(2), 533–558. https://doi.org/10.1007/s11336-022-09858-6

Kruschke, J. (2014, November 11). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press. https://www.sciencedirect.com/book/9780124058880/doing-bayesian-data-analysis

Lu, Z.-H., Chow, S.-M., & Loken, E. (2016). Bayesian Factor Analysis as a Variable-Selection Problem: Alternative Priors and Consequences. *Multivariate Behavioral Research*, *51*(4), 519–539. https://doi.org/10.1080/00273171.2016.1168279

Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032. https://doi.org/10.2307/2290129

Moors, G. (2003). Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach. Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined. *Quality and Quantity*, *37*(3), 277–302. https://doi.org/10.1023/A:1024472110002

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, *31*(1), 13–20. Retrieved September 25, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC1325062/

Nixon, N., Guo, B., Garland, A., Kaylor-Hughes, C., Nixon, E., & Morriss, R. (2020). The bi-factor structure of the 17-item Hamilton Depression Rating Scale in persistent major depression; dimensional measurement of outcome (T. D. Tran, Ed.). *PLOS ONE*, *15*(10), e0241370. https://doi.org/10.1371/journal.pone.0241370

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*(3), 587–598. https://doi.org/10.1007/s11136-004-0830-y

Pancheri, P., Picardi, A., Pasquini, M., Gaetano, P., & Biondi, M. (2002). Psychopathological dimensions of depression: A factor study of the 17-item Hamilton depression rating scale in unipolar depressed outpatients. *Journal of Affective Disorders*, *68*(1), 41–47. https://doi.org/10.1016/S0165-0327(00)00328-1

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337

Potts, M. K., Daniels, M., Burnam, M., & Wells, K. B. (1990). A structured interview version of the Hamilton Depression Rating Scale: Evidence of reliability and versatility of administration. *Journal of Psychiatric Research*, *24*(4), 335–350. https://doi.org/10.1016/0022-3956(90)90005-B

Roman, Z. J., Schmidt, P., Miller, J. M., & Brandt, H. (2024). Identifying Dynamic Shifts to Careless and Insufficient Effort Behavior in Questionnaire Responses; a Novel Approach and Experimental Validation. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(5), 775–793. https://doi.org/10.1080/10705511.2024.2304816

Rudner, L. M. (1991). Reducing Errors Due to the Use of Judges. *Practical Assessment, Research, and Evaluation*, *3*(1). https://doi.org/10.7275/w4a1-cb66

Schwartz, C. E., & Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine (1982)*, *48*(11), 1531–1548. https://doi.org/10.1016/s0277-9536(99)00047-7

Schwartz, C. E., Sprangers, M. A., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology & Health*, *19*(1), 51–69. https://doi.org/10.1080/0887044031000118456

Seemüller, F., Schennach, R., Musil, R., Obermeier, M., Adli, M., Bauer, M., Brieger, P., Laux, G., Gaebel, W., Falkai, P., Riedel, M., & Möller, H.-J. (2023). A factor analytic comparison of three commonly used depression scales (HAMD, MADRS, BDI) in a large sample of depressed inpatients. *BMC Psychiatry*, *23*(1), 548. https://doi.org/10.1186/s12888-023-05038-7

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*, *62*(1), 123–146. https://doi.org/10.1002/jclp.20213

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Stinson, L., Liu, Y., & Dallery, J. (2022). Ecological Momentary Assessment: A Systematic Review of Validity Research. *Perspectives on Behavior Science*, *45*(2), 469–493. https://doi.org/10.1007/s40614-022-00339-w

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Thero, H., & Hernvann, P.-Y. (2024). *How to deal with convergence problems*. https://cran.r-project.org/web/packages/EcoDiet/vignettes/convergence_problems.html

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vogelsmeier, L. V. D. E., Vermunt, J. K., Van Roekel, E., & De Roover, K. (2019). Latent Markov Factor Analysis for Exploring Measurement Model Changes in Time-Intensive Longitudinal Studies. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 557–575. https://doi.org/10.1080/10705511.2018.1554445

Wolfe, E. W., Moulder, B. C., & Myford, C. M. (2001). Detecting differential rater functioning over time (DRIFT) using a Rasch multi-faceted rating scale model. *Journal of Applied Measurement*, *2*(3), 256–280.

Appendix A

Table 1: Posterior estimates of model $N = 100$ with 50,000 iterations and thin = 1

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
β_0	5.20	0.45	4.38	6.16	1.05	55
β_1	-0.35	0.69	-1.72	0.97	1.01	270
β_2	-0.19	0.73	-1.64	1.22	1.00	1400
b_{1,S_1}	0.58	0.04	0.49	0.66	1.00	7000
b_{2,S_1}	0.53	0.04	0.45	0.60	1.09	33
b_{1,S_2}	0.38	0.30	-0.27	0.85	1.02	170
b_{2,S_2}	0.35	0.28	-0.28	0.80	1.01	330
b_{3,S_2}	0.21	0.36	-0.44	0.90	1.07	46
$\lambda_{2,1,S_1}$	1.02	0.04	0.94	1.10	1.00	860
$\lambda_{3,1,S_1}$	1.07	0.07	0.93	1.21	1.01	200
$\lambda_{4,1,S_1}$	0.94	0.06	0.82	1.07	1.00	890
$\lambda_{6,2,S_1}$	1.02	0.06	0.91	1.15	1.01	300
$\lambda_{7,2,S_1}$	1.11	0.06	1.00	1.24	1.01	380
$\lambda_{8,2,S_1}$	0.99	0.06	0.87	1.11	1.01	280
$\lambda_{9,2,S_1}$	0.95	0.06	0.84	1.07	1.00	1200
$\lambda_{2,1,S_2}$	0.52	0.34	-0.09	1.28	1.07	69
$\lambda_{3,1,S_2}$	0.63	0.40	-0.17	1.45	1.16	22
$\lambda_{5,2,S_2}$	0.64	0.38	-0.10	1.42	1.08	41
$\lambda_{6,2,S_2}$	0.85	0.43	-0.07	1.65	1.04	71
$\lambda_{8,3,S_2}$	1.08	0.23	0.67	1.56	1.01	20000
$\lambda_{9,3,S_2}$	0.72	0.20	0.35	1.13	1.01	1400
$\lambda^*_{1,2,S_1}$	0.97	0.11	0.71	1.19	1.16	33
$\lambda^*_{2,2,S_1}$	1.01	0.11	0.74	1.23	1.14	42

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Table 1: Posterior estimates of model $N = 100$ with 50,000 iterations and thin = 1 (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{3,2,S_1}$	-0.04	0.11	-0.32	0.16	1.19	24
$\lambda^*_{4,2,S_1}$	-0.01	0.09	-0.24	0.17	1.15	31
$\lambda^*_{5,1,S_1}$	-0.02	0.05	-0.14	0.08	1.01	290
$\lambda^*_{6,1,S_1}$	0.03	0.05	-0.07	0.15	1.02	160
$\lambda^*_{7,1,S_1}$	-0.04	0.06	-0.17	0.06	1.01	420
$\lambda^*_{8,1,S_1}$	0.02	0.05	-0.09	0.12	1.01	400
$\lambda^*_{9,1,S_1}$	0.03	0.05	-0.07	0.13	1.01	260
$\lambda^*_{1,2,S_1}$	0.59	0.53	-0.11	1.69	1.04	73
$\lambda^*_{2,2,S_1}$	-0.08	0.28	-0.72	0.45	1.02	160
$\lambda^*_{3,2,S_1}$	-0.06	0.24	-0.67	0.37	1.01	330
$\lambda^*_{4,1,S_1}$	-0.02	0.19	-0.43	0.37	1.05	61
$\lambda^*_{5,1,S_1}$	0.12	0.23	-0.25	0.66	1.20	18
$\lambda^*_{6,1,S_1}$	0.15	0.27	-0.20	0.87	1.04	220
$\lambda^*_{7,1,S_1}$	0.04	0.19	-0.31	0.49	1.02	260
$\lambda^*_{8,1,S_1}$	-0.00	0.20	-0.43	0.44	1.02	5600
$\lambda^*_{9,1,S_1}$	0.07	0.20	-0.24	0.60	1.03	310
$\lambda^*_{1,3,S_2}$	-0.10	0.29	-0.77	0.45	1.04	76
$\lambda^*_{2,3,S_2}$	0.41	0.25	-0.03	0.92	1.02	780
$\lambda^*_{3,3,S_2}$	0.02	0.18	-0.37	0.42	1.05	57
$\lambda^*_{4,3,S_2}$	-0.04	0.16	-0.41	0.26	1.02	190
$\lambda^*_{5,3,S_2}$	0.02	0.17	-0.35	0.37	1.16	21
$\lambda^*_{6,3,S_2}$	0.03	0.17	-0.31	0.41	1.01	570
$\lambda^*_{7,2,S_2}$	0.01	0.19	-0.41	0.41	1.01	350
$\lambda^*_{8,2,S_2}$	-0.04	0.19	-0.50	0.34	1.01	270
$\lambda^*_{9,2,S_2}$	-0.01	0.18	-0.45	0.35	1.02	350
$\sigma^2_{\epsilon_{1,1,S_1}}$	0.30	0.02	0.27	0.34	1.01	530
$\sigma^2_{\epsilon_{1,2,S_1}}$	0.29	0.02	0.25	0.32	1.01	330
$\sigma^2_{\epsilon_{1,3,S_1}}$	0.63	0.03	0.58	0.69	1.01	510
$\sigma^2_{\epsilon_{1,4,S_1}}$	0.64	0.03	0.59	0.70	1.01	500
$\sigma^2_{\epsilon_{1,5,S_1}}$	0.67	0.03	0.62	0.73	1.00	1900
$\sigma^2_{\epsilon_{1,6,S_1}}$	0.63	0.03	0.58	0.68	1.00	1300

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Table 1: Posterior estimates of model $N = 100$ with 50,000 iterations and thin = 1 (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\sigma^2_{\epsilon_{1,7,S_1}}$	0.56	0.02	0.52	0.61	1.00	1700
$\sigma^2_{\epsilon_{1,8,S_1}}$	0.68	0.03	0.63	0.74	1.00	3000
$\sigma^2_{\epsilon_{1,9,S_1}}$	0.64	0.03	0.59	0.69	1.00	1100
$\sigma^2_{\epsilon_{1,1,S_2}}$	0.56	0.24	0.22	1.15	1.04	77
$\sigma^2_{\epsilon_{1,2,S_2}}$	0.56	0.19	0.28	1.00	1.02	210
$\sigma^2_{\epsilon_{1,3,S_2}}$	0.84	0.26	0.43	1.44	1.07	39
$\sigma^2_{\epsilon_{1,4,S_2}}$	0.62	0.21	0.31	1.12	1.01	660
$\sigma^2_{\epsilon_{1,5,S_2}}$	0.65	0.20	0.35	1.11	1.00	1200
$\sigma^2_{\epsilon_{1,6,S_2}}$	0.79	0.24	0.42	1.37	1.01	320
$\sigma^2_{\epsilon_{1,7,S_2}}$	0.82	0.24	0.44	1.38	1.00	1900
$\sigma^2_{\epsilon_{1,8,S_2}}$	0.52	0.20	0.23	0.99	1.01	230
$\sigma^2_{\epsilon_{1,9,S_2}}$	0.54	0.16	0.29	0.91	1.01	990
$\Phi_{1,S_1}[1,1]$	0.16	0.02	0.12	0.20	1.03	110
$\Phi_{1,S_1}[1,2]$	0.01	0.02	-0.02	0.05	1.09	53
$\Phi_{1,S_1}[2,2]$	0.16	0.02	0.13	0.19	1.02	120
$\Phi_{1,S_2}[1,1]$	0.34	0.18	0.11	0.80	1.04	72
$\Phi_{1,S_2}[2,1]$	0.03	0.11	-0.16	0.28	1.02	240
$\Phi_{1,S_2}[3,1]$	0.06	0.12	-0.12	0.36	1.13	29
$\Phi_{1,S_2}[2,2]$	0.23	0.11	0.09	0.52	1.01	230
$\Phi_{1,S_2}[3,2]$	-0.01	0.08	-0.17	0.16	1.10	31
$\Phi_{1,S_2}[3,3]$	0.27	0.14	0.10	0.64	1.16	21
$\Phi_{2,S_1}[1,1]$	0.16	0.03	0.10	0.23	1.01	370
$\Phi_{2,S_1}[2,1]$	0.01	0.02	-0.03	0.06	1.02	160
$\Phi_{2,S_1}[2,2]$	0.12	0.02	0.08	0.18	1.01	570
$\Phi_{2,S_2}[1,1]$	0.54	0.45	0.12	1.78	1.05	56
$\Phi_{2,S_2}[2,1]$	0.08	0.25	-0.35	0.63	1.05	92
$\Phi_{2,S_2}[3,1]$	0.12	0.35	-0.47	1.01	1.06	52
$\Phi_{2,S_2}[2,2]$	0.41	0.32	0.11	1.22	1.03	120
$\Phi_{2,S_2}[3,2]$	0.01	0.24	-0.48	0.55	1.03	88
$\Phi_{2,S_2}[3,3]$	0.71	0.46	0.18	1.95	1.03	110

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Appendix B

Table 2: Posterior estimates of model $N = 200$ with 50,000 iterations and thin = 1

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
β_0	3.85	0.20	3.49	4.26	1.08	37
β_1	-0.53	0.36	-1.24	0.20	1.01	280
β_2	0.23	0.32	-0.39	0.85	1.00	930
b_{1,S_1}	0.52	0.04	0.44	0.60	1.00	1900
b_{2,S_1}	0.52	0.03	0.45	0.58	1.00	4500
b_{1,S_2}	0.37	0.14	0.07	0.60	1.13	28
b_{2,S_2}	0.25	0.15	-0.07	0.52	1.33	12
b_{3,S_2}	0.40	0.12	0.16	0.61	1.05	61
$\lambda_{2,1,S_1}$	1.01	0.03	0.95	1.07	1.00	780
$\lambda_{3,1,S_1}$	0.98	0.05	0.87	1.08	1.03	100
$\lambda_{4,1,S_1}$	1.02	0.06	0.90	1.13	1.02	140
$\lambda_{6,2,S_1}$	1.02	0.04	0.94	1.11	1.01	340
$\lambda_{7,2,S_1}$	0.99	0.04	0.91	1.08	1.01	220
$\lambda_{8,2,S_1}$	0.99	0.04	0.91	1.07	1.01	240
$\lambda_{9,2,S_1}$	0.94	0.04	0.86	1.02	1.01	200
$\lambda_{2,1,S_2}$	0.58	0.48	-0.44	1.12	4.63	4
$\lambda_{3,1,S_2}$	0.59	0.35	-0.16	1.01	4.30	4
$\lambda_{5,2,S_2}$	1.01	0.18	0.62	1.33	1.61	9
$\lambda_{6,2,S_2}$	0.88	0.23	0.35	1.25	2.25	6
$\lambda_{8,3,S_2}$	0.98	0.12	0.76	1.24	1.02	200
$\lambda_{9,3,S_2}$	0.86	0.12	0.64	1.12	1.01	340
$\lambda^*_{1,2,S_1}$	1.05	0.09	0.81	1.22	1.09	45
$\lambda^*_{2,2,S_1}$	1.00	0.09	0.76	1.16	1.09	46

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Table 2: Posterior estimates of model $N = 200$ with 50,000 iterations and thin = 1 (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{3,2,S_1}$	-0.01	0.08	-0.23	0.14	1.07	52
$\lambda^*_{4,2,S_1}$	-0.03	0.08	-0.27	0.10	1.07	57
$\lambda^*_{5,1,S_1}$	-0.01	0.04	-0.09	0.09	1.03	92
$\lambda^*_{6,1,S_1}$	0.01	0.04	-0.06	0.12	1.02	170
$\lambda^*_{7,1,S_1}$	0.04	0.04	-0.03	0.14	1.03	110
$\lambda^*_{8,1,S_1}$	0.01	0.04	-0.07	0.10	1.02	190
$\lambda^*_{9,1,S_1}$	-0.00	0.04	-0.08	0.08	1.01	320
$\lambda^*_{1,2,S_2}$	1.15	0.41	0.60	2.03	2.42	5
$\lambda^*_{2,2,S_2}$	0.03	0.46	-0.54	1.05	3.61	5
$\lambda^*_{3,2,S_2}$	0.21	0.40	-0.25	1.12	4.28	4
$\lambda^*_{4,1,S_2}$	0.35	0.38	-0.03	1.15	4.63	4
$\lambda^*_{5,1,S_2}$	0.30	0.45	-0.10	1.26	6.02	4
$\lambda^*_{6,1,S_2}$	0.20	0.46	-0.29	1.20	5.60	4
$\lambda^*_{7,1,S_2}$	0.00	0.12	-0.28	0.25	1.28	15
$\lambda^*_{8,1,S_2}$	-0.03	0.13	-0.29	0.27	1.33	13
$\lambda^*_{9,1,S_2}$	0.02	0.09	-0.16	0.22	1.03	170
$\lambda^*_{1,3,S_2}$	-0.02	0.25	-0.45	0.50	1.15	33
$\lambda^*_{2,3,S_2}$	0.81	0.20	0.43	1.23	1.03	120
$\lambda^*_{3,3,S_2}$	0.00	0.16	-0.33	0.38	1.11	81
$\lambda^*_{4,3,S_2}$	-0.01	0.15	-0.31	0.35	1.16	71
$\lambda^*_{5,3,S_2}$	0.04	0.15	-0.23	0.42	1.21	65
$\lambda^*_{6,3,S_2}$	0.06	0.15	-0.19	0.45	1.24	34
$\lambda^*_{7,2,S_2}$	-0.04	0.11	-0.29	0.19	1.13	28
$\lambda^*_{8,2,S_2}$	0.04	0.12	-0.22	0.30	1.19	20
$\lambda^*_{9,2,S_2}$	0.02	0.09	-0.15	0.23	1.03	200
$\sigma^2_{\epsilon_{1,1,S_1}}$	0.28	0.01	0.26	0.31	1.00	1800
$\sigma^2_{\epsilon_{1,2,S_1}}$	0.27	0.01	0.24	0.29	1.00	17000
$\sigma^2_{\epsilon_{1,3,S_1}}$	0.62	0.02	0.58	0.67	1.01	380
$\sigma^2_{\epsilon_{1,4,S_1}}$	0.63	0.02	0.59	0.68	1.02	110
$\sigma^2_{\epsilon_{1,5,S_1}}$	0.66	0.02	0.62	0.70	1.00	1100
$\sigma^2_{\epsilon_{1,6,S_1}}$	0.66	0.02	0.62	0.70	1.00	1200

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Table 2: Posterior estimates of model $N = 200$ with 50,000 iterations and thin = 1 (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\sigma^2_{\epsilon_{1,7,S_1}}$	0.63	0.02	0.59	0.67	1.01	370
$\sigma^2_{\epsilon_{1,8,S_1}}$	0.66	0.02	0.62	0.70	1.01	510
$\sigma^2_{\epsilon_{1,9,S_1}}$	0.65	0.02	0.61	0.69	1.01	230
$\sigma^2_{\epsilon_{1,1,S_2}}$	0.24	0.05	0.16	0.34	1.03	96
$\sigma^2_{\epsilon_{1,2,S_2}}$	0.31	0.05	0.22	0.41	1.01	340
$\sigma^2_{\epsilon_{1,3,S_2}}$	0.57	0.06	0.46	0.69	1.03	80
$\sigma^2_{\epsilon_{1,4,S_2}}$	0.71	0.07	0.58	0.86	1.01	460
$\sigma^2_{\epsilon_{1,5,S_2}}$	0.57	0.06	0.46	0.69	1.01	540
$\sigma^2_{\epsilon_{1,6,S_2}}$	0.61	0.06	0.49	0.74	1.02	180
$\sigma^2_{\epsilon_{1,7,S_2}}$	0.58	0.07	0.45	0.74	1.02	120
$\sigma^2_{\epsilon_{1,8,S_2}}$	0.67	0.07	0.53	0.81	1.07	42
$\sigma^2_{\epsilon_{1,9,S_2}}$	0.70	0.07	0.58	0.84	1.01	400
$\Phi_{1,S_1}[1,1]$	0.16	0.02	0.13	0.19	1.03	90
$\Phi_{1,S_1}[2,1]$	-0.00	0.02	-0.03	0.04	1.06	76
$\Phi_{1,S_1}[2,2]$	0.17	0.01	0.14	0.19	1.01	370
$\Phi_{1,S_2}[1,1]$	0.23	0.08	0.12	0.43	1.06	140
$\Phi_{1,S_2}[2,1]$	-0.03	0.08	-0.25	0.07	2.28	6
$\Phi_{1,S_2}[3,1]$	0.03	0.06	-0.11	0.13	1.41	11
$\Phi_{1,S_2}[2,2]$	0.17	0.06	0.10	0.32	1.18	26
$\Phi_{1,S_2}[3,2]$	0.01	0.04	-0.06	0.10	1.15	48
$\Phi_{1,S_2}[3,3]$	0.20	0.05	0.12	0.30	1.02	170
$\Phi_{2,S_1}[1,1]$	0.14	0.02	0.10	0.19	1.01	320
$\Phi_{2,S_1}[1,2]$	-0.01	0.02	-0.04	0.03	1.02	160
$\Phi_{2,S_1}[2,2]$	0.14	0.02	0.11	0.19	1.01	620
$\Phi_{2,S_2}[1,1]$	0.21	0.09	0.09	0.45	1.10	35
$\Phi_{2,S_2}[2,1]$	-0.00	0.08	-0.20	0.12	1.81	7
$\Phi_{2,S_2}[3,1]$	0.02	0.06	-0.09	0.14	1.03	220
$\Phi_{2,S_2}[2,2]$	0.19	0.06	0.09	0.34	1.03	120
$\Phi_{2,S_2}[3,2]$	0.02	0.05	-0.07	0.12	1.06	62
$\Phi_{2,S_2}[3,3]$	0.19	0.07	0.09	0.34	1.01	630

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_{1,S_l}, covariance matrix for random effects Φ_{2,S_l}. The first primary loading of each factor were fixed to 1.

Appendix C

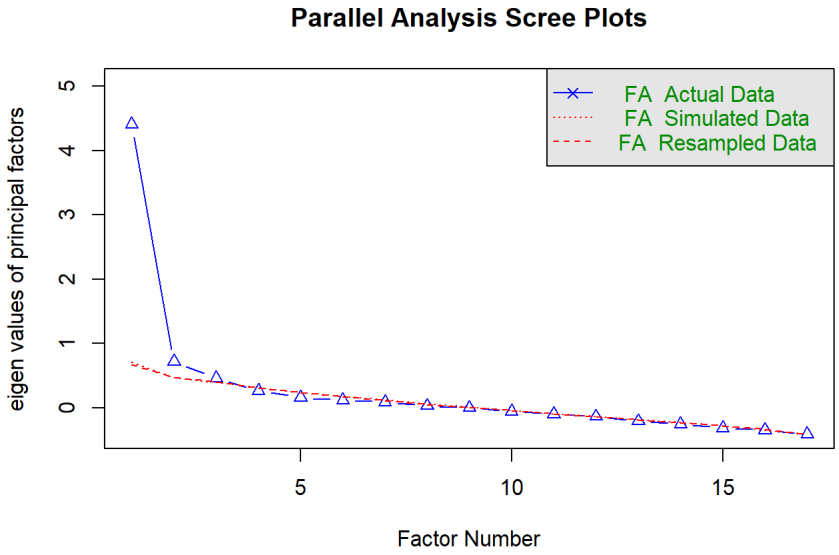


Figure 1: Parallel analysis for $t = 13$

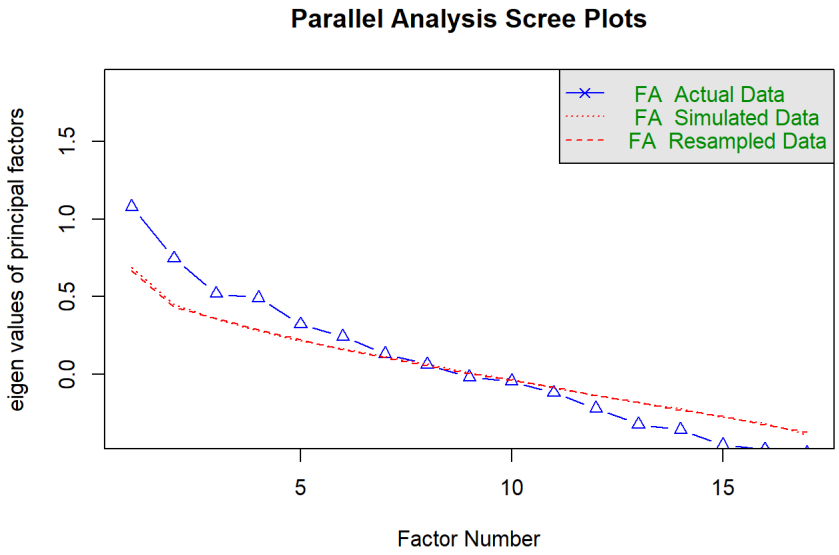


Figure 2: Parallel analysis for $t = 1$

Appendix D

Table 3: Posterior estimates of uncorrelated 3-factor DSEM model

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
b_1	1.00	0.00	0.99	1.00	1.00	25000
b_2	0.76	0.05	0.66	0.85	1.00	9200
b_3	0.67	0.11	0.44	0.85	1.01	400
$\lambda_{2,1}$	0.61	0.03	0.56	0.66	1.00	980
$\lambda_{3,1}$	0.49	0.01	0.46	0.52	1.00	3800
$\lambda_{4,1}$	0.84	0.03	0.80	0.90	1.00	1900
$\lambda_{5,1}$	0.59	0.02	0.55	0.62	1.00	2300
$\lambda_{6,1}$	0.26	0.02	0.22	0.29	1.00	2800
$\lambda_{7,1}$	0.57	0.01	0.55	0.60	1.00	8200
$\lambda_{8,1}$	0.39	0.01	0.37	0.42	1.00	21000
$\lambda_{9,1}$	0.57	0.02	0.53	0.61	1.00	880
$\lambda_{11,2}$	2.01	0.11	1.80	2.24	1.00	12000
$\lambda_{12,2}$	0.68	0.09	0.51	0.86	1.00	10000
$\lambda_{13,2}$	1.21	0.08	1.07	1.37	1.00	1300
$\lambda_{15,3}$	0.89	0.08	0.74	1.05	1.00	1000
$\lambda_{16,3}$	0.55	0.06	0.42	0.67	1.00	2800
$\lambda_{17,3}$	0.07	0.06	-0.05	0.18	1.00	2300
$\lambda^*_{1,2}$	0.10	0.12	-0.08	0.39	1.00	990
$\lambda^*_{2,2}$	-0.61	0.11	-0.83	-0.40	1.00	2000
$\lambda^*_{3,2}$	0.02	0.06	-0.09	0.15	1.00	4500
$\lambda^*_{4,2}$	0.66	0.11	0.44	0.88	1.00	3400
$\lambda^*_{5,2}$	-0.03	0.07	-0.18	0.10	1.00	3300
$\lambda^*_{6,2}$	0.29	0.09	0.12	0.47	1.00	1200
$\lambda^*_{7,2}$	-0.21	0.09	-0.37	-0.03	1.00	1100

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_k, primary loadings $\lambda_{j,k}$, cross-loadings $\lambda^*_{j,k}$, residual variance $\sigma^2_{\epsilon_{1,j}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 3: Posterior estimates of uncorrelated 3-factor DSEM model (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{8,2}$	0.09	0.08	-0.03	0.25	1.00	1300
$\lambda^*_{9,2}$	-0.62	0.09	-0.80	-0.43	1.00	2600
$\lambda^*_{10,1}$	0.13	0.02	0.09	0.16	1.00	11000
$\lambda^*_{11,1}$	0.21	0.03	0.16	0.26	1.00	5000
$\lambda^*_{12,1}$	0.48	0.02	0.44	0.52	1.00	37000
$\lambda^*_{13,1}$	0.19	0.02	0.15	0.22	1.00	5500
$\lambda^*_{14,1}$	0.19	0.02	0.15	0.22	1.00	1200
$\lambda^*_{15,1}$	0.25	0.02	0.21	0.29	1.00	1600
$\lambda^*_{16,1}$	0.16	0.02	0.13	0.20	1.00	2000
$\lambda^*_{17,1}$	0.28	0.01	0.25	0.31	1.00	23000
$\lambda^*_{1,3}$	0.93	0.11	0.72	1.15	1.01	790
$\lambda^*_{2,3}$	-0.46	0.10	-0.66	-0.26	1.00	10000
$\lambda^*_{3,3}$	0.09	0.07	-0.03	0.23	1.00	870
$\lambda^*_{4,3}$	-0.09	0.10	-0.30	0.07	1.01	650
$\lambda^*_{5,3}$	0.05	0.07	-0.07	0.20	1.00	1300
$\lambda^*_{6,3}$	0.70	0.07	0.58	0.84	1.00	41000
$\lambda^*_{7,3}$	0.48	0.07	0.34	0.63	1.00	1200
$\lambda^*_{8,3}$	0.41	0.07	0.27	0.56	1.00	1200
$\lambda^*_{9,3}$	-0.24	0.09	-0.42	-0.05	1.00	7000
$\lambda^*_{10,3}$	0.09	0.07	-0.03	0.23	1.00	2800
$\lambda^*_{11,3}$	0.04	0.11	-0.18	0.27	1.01	350
$\lambda^*_{12,3}$	0.32	0.08	0.16	0.47	1.00	1100
$\lambda^*_{13,3}$	-0.12	0.08	-0.28	0.02	1.01	420
$\lambda^*_{14,2}$	0.09	0.10	-0.06	0.31	1.00	1100
$\lambda^*_{15,2}$	0.02	0.09	-0.14	0.23	1.00	950
$\lambda^*_{16,2}$	0.44	0.08	0.28	0.61	1.00	1100
$\lambda^*_{17,2}$	-0.02	0.06	-0.14	0.09	1.00	18000
$\sigma^2_{\epsilon_{1,1}}$	0.70	0.03	0.65	0.75	1.00	32000
$\sigma^2_{\epsilon_{1,2}}$	0.77	0.03	0.72	0.83	1.00	5500
$\sigma^2_{\epsilon_{1,3}}$	0.50	0.02	0.47	0.53	1.00	22000
$\sigma^2_{\epsilon_{1,4}}$	0.89	0.03	0.83	0.95	1.00	7500
$\sigma^2_{\epsilon_{1,5}}$	0.79	0.02	0.75	0.84	1.00	42000

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_k, primary loadings $\lambda_{j,k}$, cross-loadings $\lambda^*_{j,k}$, residual variance $\sigma^2_{\epsilon_{1,j}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 3: Posterior estimates of uncorrelated 3-factor DSEM model (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\sigma^2_{\epsilon_{1,6}}$	0.60	0.02	0.56	0.64	1.00	8800
$\sigma^2_{\epsilon_{1,7}}$	0.55	0.02	0.52	0.59	1.00	100000
$\sigma^2_{\epsilon_{1,8}}$	0.71	0.02	0.67	0.75	1.00	11000
$\sigma^2_{\epsilon_{1,9}}$	0.62	0.02	0.58	0.66	1.00	11000
$\sigma^2_{\epsilon_{1,10}}$	0.81	0.02	0.76	0.86	1.00	30000
$\sigma^2_{\epsilon_{1,11}}$	0.38	0.02	0.35	0.43	1.00	2300
$\sigma^2_{\epsilon_{1,12}}$	1.01	0.03	0.95	1.07	1.00	100000
$\sigma^2_{\epsilon_{1,13}}$	0.44	0.02	0.41	0.47	1.00	4700
$\sigma^2_{\epsilon_{1,14}}$	0.73	0.02	0.68	0.77	1.00	6200
$\sigma^2_{\epsilon_{1,15}}$	0.85	0.03	0.79	0.91	1.00	3900
$\sigma^2_{\epsilon_{1,16}}$	0.84	0.03	0.79	0.89	1.00	29000
$\sigma^2_{\epsilon_{1,17}}$	0.81	0.02	0.76	0.85	1.00	78000
$\Phi_1[1]$	0.16	0.01	0.14	0.18	1.00	12000
$\Phi_1[2]$	0.02	0.00	0.02	0.02	1.00	3700
$\Phi_1[3]$	0.03	0.00	0.02	0.03	1.00	13000
$\Phi_2[1]$	0.18	0.04	0.11	0.26	1.00	8900
$\Phi_2[2]$	0.09	0.01	0.07	0.11	1.00	19000
$\Phi_2[3]$	0.18	0.03	0.14	0.23	1.00	3100

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_k, primary loadings $\lambda_{j,k}$, cross-loadings $\lambda^*_{j,k}$, residual variance $\sigma^2_{\epsilon_{1,j}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Appendix E

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
β_0	5.57	0.54	4.59	6.71	1.02	140
β_1	1.79	0.27	1.27	2.32	1.80	7
β_2	0.39	0.28	-0.14	0.97	1.24	15
β_3	0.36	0.39	-0.42	1.05	1.99	6
b_1	0.19	0.07	0.05	0.34	1.03	94
b_2	0.96	0.06	0.80	1.00	1.43	13
b_3	0.78	0.16	0.52	1.00	2.41	5
b_4	0.90	0.10	0.64	1.00	1.28	28
b_5	0.94	0.04	0.84	1.00	1.39	11
b_6	0.87	0.15	0.41	0.99	1.05	77
$\lambda_{2,1,S_1}$	0.18	0.13	-0.01	0.42	3.83	5
$\lambda_{3,1,S_1}$	0.27	0.04	0.18	0.34	1.43	11
$\lambda_{4,1,S_1}$	0.39	0.07	0.26	0.53	1.93	6
$\lambda_{5,1,S_1}$	0.36	0.08	0.23	0.50	2.64	5
$\lambda_{6,1,S_1}$	0.30	0.16	0.06	0.59	3.77	5
$\lambda_{7,1,S_1}$	0.50	0.03	0.44	0.57	1.37	11
$\lambda_{8,1,S_1}$	0.39	0.04	0.31	0.47	1.05	65
$\lambda_{9,1,S_1}$	0.22	0.13	0.04	0.45	3.72	5
$\lambda_{11,2,S_1}$	0.57	0.21	0.26	1.06	2.05	6
$\lambda_{12,2,S_1}$	0.94	0.22	0.50	1.25	2.31	6
$\lambda_{13,2,S_1}$	0.15	0.26	-0.17	0.67	3.28	5
$\lambda_{15,3,S_1}$	0.99	0.10	0.79	1.20	1.43	10
$\lambda_{16,3,S_1}$	0.74	0.13	0.46	0.95	2.26	6

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda_{17,3,S_1}$	0.42	0.15	0.17	0.69	2.48	5
$\lambda_{2,1,S_2}$	0.30	0.17	0.06	0.52	6.55	4
$\lambda_{3,1,S_2}$	0.32	0.04	0.25	0.40	2.90	5
$\lambda_{4,1,S_2}$	0.72	0.06	0.59	0.82	1.71	8
$\lambda_{5,1,S_2}$	0.47	0.03	0.40	0.53	1.16	22
$\lambda_{6,1,S_2}$	0.35	0.15	0.23	0.67	4.95	4
$\lambda_{7,1,S_2}$	0.48	0.04	0.43	0.56	1.92	7
$\lambda_{8,1,S_2}$	0.39	0.03	0.32	0.45	1.39	11
$\lambda_{9,1,S_2}$	0.31	0.13	0.12	0.47	4.30	4
$\lambda_{11,2,S_2}$	0.39	0.12	0.13	0.60	2.26	6
$\lambda_{12,2,S_2}$	1.03	0.16	0.74	1.32	1.55	9
$\lambda_{13,2,S_2}$	0.00	0.00	-0.01	0.01	1.06	53
$\lambda_{15,3,S_2}$	0.92	0.12	0.68	1.14	1.56	9
$\lambda_{16,3,S_2}$	0.52	0.09	0.34	0.70	1.18	19
$\lambda_{17,3,S_2}$	0.34	0.20	0.07	0.73	3.40	5
$\lambda^*_{1,2,S_1}$	1.18	0.77	-0.12	2.10	5.39	4
$\lambda^*_{2,2,S_1}$	0.83	0.66	-0.31	1.66	5.85	4
$\lambda^*_{3,2,S_1}$	0.38	0.34	-0.22	0.82	5.38	4
$\lambda^*_{4,2,S_1}$	1.20	0.54	0.26	1.88	5.39	4
$\lambda^*_{5,2,S_1}$	1.00	0.48	0.12	1.56	5.13	4
$\lambda^*_{6,2,S_1}$	-0.09	0.33	-1.31	0.23	2.08	6
$\lambda^*_{7,2,S_1}$	0.84	0.46	0.03	1.38	4.51	4
$\lambda^*_{8,2,S_1}$	0.57	0.42	-0.12	1.14	5.71	4
$\lambda^*_{9,2,S_1}$	0.89	0.71	-0.19	1.80	6.11	4
$\lambda^*_{10,1,S_1}$	0.06	0.05	-0.02	0.18	1.83	7
$\lambda^*_{11,1,S_1}$	0.15	0.09	-0.00	0.30	3.01	5
$\lambda^*_{12,1,S_1}$	0.33	0.05	0.24	0.42	1.38	11
$\lambda^*_{13,1,S_1}$	-0.09	0.09	-0.23	0.04	3.72	5
$\lambda^*_{14,1,S_1}$	0.40	0.05	0.33	0.51	2.70	5
$\lambda^*_{15,1,S_1}$	0.37	0.05	0.29	0.48	1.63	8
$\lambda^*_{16,1,S_1}$	0.29	0.06	0.19	0.41	2.53	5
$\lambda^*_{17,1,S_1}$	0.17	0.03	0.11	0.23	1.14	23

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{1,3,S_1}$	2.05	0.33	1.56	2.65	4.67	4
$\lambda^*_{2,3,S_1}$	0.31	0.45	-0.27	1.11	5.04	4
$\lambda^*_{3,3,S_1}$	0.52	0.20	0.15	0.86	3.44	5
$\lambda^*_{4,3,S_1}$	0.63	0.28	0.24	1.16	3.05	5
$\lambda^*_{5,3,S_1}$	0.73	0.32	0.30	1.29	3.74	5
$\lambda^*_{6,3,S_1}$	0.67	0.43	0.05	1.40	4.23	4
$\lambda^*_{7,3,S_1}$	1.03	0.23	0.67	1.42	3.81	5
$\lambda^*_{8,3,S_1}$	0.85	0.14	0.55	1.11	1.76	7
$\lambda^*_{9,3,S_1}$	0.44	0.40	-0.05	1.10	4.46	4
$\lambda^*_{10,3,S_1}$	0.36	0.15	0.10	0.67	1.96	6
$\lambda^*_{11,3,S_1}$	0.29	0.24	-0.13	0.68	3.36	5
$\lambda^*_{12,3,S_1}$	0.64	0.16	0.36	0.97	2.03	6
$\lambda^*_{13,3,S_1}$	-0.12	0.20	-0.47	0.18	3.36	5
$\lambda^*_{14,2,S_1}$	0.38	0.24	-0.02	0.70	3.38	5
$\lambda^*_{15,2,S_1}$	0.48	0.34	-0.06	0.94	3.92	5
$\lambda^*_{16,2,S_1}$	0.39	0.22	-0.01	0.73	3.75	5
$\lambda^*_{17,2,S_1}$	0.35	0.25	-0.04	0.71	3.85	5
$\lambda^*_{1,4,S_1}$	-0.04	0.07	-0.19	0.06	2.29	6
$\lambda^*_{2,4,S_1}$	-0.10	0.12	-0.36	0.12	3.63	5
$\lambda^*_{3,4,S_1}$	-0.02	0.09	-0.18	0.14	3.14	5
$\lambda^*_{4,4,S_1}$	-0.03	0.07	-0.17	0.10	2.57	5
$\lambda^*_{5,4,S_1}$	-0.04	0.07	-0.18	0.09	3.38	5
$\lambda^*_{6,4,S_1}$	0.08	0.25	-0.15	0.64	6.21	4
$\lambda^*_{7,4,S_1}$	-0.03	0.10	-0.17	0.18	4.92	4
$\lambda^*_{7,4,S_1}$	0.02	0.09	-0.16	0.22	1.86	7
$\lambda^*_{9,4,S_1}$	-0.06	0.18	-0.48	0.18	3.70	5
$\lambda^*_{10,4,S_1}$	0.05	0.04	-0.02	0.13	1.44	10
$\lambda^*_{11,4,S_1}$	0.02	0.26	-0.50	0.34	5.41	4
$\lambda^*_{12,4,S_1}$	0.02	0.05	-0.07	0.12	1.61	8
$\lambda^*_{13,4,S_1}$	0.03	0.18	-0.30	0.29	4.71	4
$\lambda^*_{14,4,S_1}$	-0.01	0.03	-0.07	0.05	1.47	10
$\lambda^*_{15,4,S_1}$	-0.03	0.07	-0.19	0.06	3.44	5

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{16,4,S_1}$	-0.01	0.07	-0.14	0.12	4.09	4
$\lambda^*_{17,4,S_1}$	-0.02	0.04	-0.10	0.05	2.44	5
$\lambda^*_{1,5,S_1}$	0.02	0.09	-0.12	0.19	3.03	5
$\lambda^*_{2,5,S_1}$	0.10	0.22	-0.14	0.58	6.00	4
$\lambda^*_{3,5,S_1}$	-0.02	0.12	-0.23	0.18	4.01	4
$\lambda^*_{4,5,S_1}$	0.02	0.14	-0.22	0.25	4.89	4
$\lambda^*_{5,5,S_1}$	0.03	0.10	-0.10	0.25	4.60	4
$\lambda^*_{6,5,S_1}$	0.19	0.26	-0.14	0.81	4.13	4
$\lambda^*_{7,5,S_1}$	-0.01	0.07	-0.13	0.12	3.29	5
$\lambda^*_{8,5,S_1}$	0.05	0.12	-0.13	0.30	2.88	5
$\lambda^*_{9,5,S_1}$	0.14	0.16	-0.08	0.48	3.56	5
$\lambda^*_{10,5,S_1}$	-0.00	0.07	-0.14	0.10	2.98	5
$\lambda^*_{11,5,S_1}$	0.02	0.24	-0.38	0.47	4.62	4
$\lambda^*_{12,5,S_1}$	0.01	0.12	-0.22	0.18	4.04	4
$\lambda^*_{13,5,S_1}$	-0.02	0.21	-0.35	0.37	5.07	4
$\lambda^*_{14,5,S_1}$	0.03	0.05	-0.05	0.13	2.65	5
$\lambda^*_{15,5,S_1}$	0.00	0.04	-0.07	0.08	1.65	8
$\lambda^*_{16,5,S_1}$	-0.01	0.03	-0.07	0.03	1.03	110
$\lambda^*_{17,5,S_1}$	0.01	0.04	-0.06	0.11	2.69	5
$\lambda^*_{1,6,S_1}$	0.04	0.07	-0.06	0.20	2.45	5
$\lambda^*_{2,6,S_1}$	-0.02	0.08	-0.15	0.15	2.19	6
$\lambda^*_{3,6,S_1}$	0.06	0.10	-0.15	0.21	4.20	4
$\lambda^*_{4,6,S_1}$	-0.02	0.07	-0.15	0.12	2.83	5
$\lambda^*_{5,6,S_1}$	-0.01	0.04	-0.08	0.07	1.65	8
$\lambda^*_{6,6,S_1}$	0.01	0.04	-0.08	0.10	1.31	14
$\lambda^*_{7,6,S_1}$	0.01	0.05	-0.07	0.12	2.21	6
$\lambda^*_{8,6,S_1}$	0.09	0.22	-0.38	0.35	5.62	4
$\lambda^*_{9,6,S_1}$	0.04	0.17	-0.28	0.35	3.00	5
$\lambda^*_{10,6,S_1}$	-0.04	0.05	-0.17	0.04	2.06	6
$\lambda^*_{11,6,S_1}$	-0.00	0.15	-0.33	0.27	2.72	5
$\lambda^*_{12,6,S_1}$	-0.04	0.09	-0.20	0.15	4.07	4
$\lambda^*_{13,6,S_1}$	-0.02	0.13	-0.31	0.19	2.87	5

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{14,6,S_1}$	0.02	0.05	-0.08	0.10	2.93	5
$\lambda^*_{15,6,S_1}$	-0.03	0.07	-0.15	0.10	3.32	5
$\lambda^*_{16,6,S_1}$	0.00	0.03	-0.05	0.06	1.45	11
$\lambda^*_{17,6,S_1}$	0.00	0.02	-0.03	0.05	1.17	24
$\lambda^*_{1,2,S_2}$	1.40	0.70	0.12	2.17	5.71	4
$\lambda^*_{2,2,S_2}$	0.79	0.59	-0.20	1.50	7.34	4
$\lambda^*_{3,2,S_2}$	0.00	0.01	-0.01	0.02	1.07	40
$\lambda^*_{4,2,S_2}$	1.36	0.53	0.37	2.03	6.06	4
$\lambda^*_{5,2,S_2}$	0.79	0.32	0.21	1.22	4.21	4
$\lambda^*_{6,2,S_2}$	-0.06	0.22	-1.00	0.06	1.48	16
$\lambda^*_{7,2,S_2}$	0.50	0.23	0.06	0.77	4.08	4
$\lambda^*_{8,2,S_2}$	0.64	0.35	0.01	1.11	5.63	4
$\lambda^*_{9,2,S_2}$	0.81	0.52	-0.03	1.47	7.81	4
$\lambda^*_{10,1,S_2}$	0.19	0.09	0.03	0.36	3.04	5
$\lambda^*_{11,1,S_2}$	0.26	0.07	0.15	0.38	3.55	5
$\lambda^*_{12,1,S_2}$	0.52	0.11	0.36	0.71	3.56	5
$\lambda^*_{13,1,S_2}$	0.17	0.02	0.14	0.22	2.90	5
$\lambda^*_{14,1,S_2}$	0.41	0.06	0.32	0.53	3.56	5
$\lambda^*_{15,1,S_2}$	0.41	0.07	0.31	0.56	2.77	5
$\lambda^*_{16,1,S_2}$	0.25	0.05	0.18	0.35	2.01	6
$\lambda^*_{17,1,S_2}$	0.21	0.06	0.11	0.31	2.27	6
$\lambda^*_{1,3,S_2}$	1.75	0.26	1.39	2.28	3.53	5
$\lambda^*_{2,3,S_2}$	0.33	0.47	-0.29	1.06	7.29	4
$\lambda^*_{3,3,S_2}$	-0.01	0.01	-0.04	0.00	1.30	12
$\lambda^*_{4,3,S_2}$	1.19	0.22	0.76	1.58	1.99	6
$\lambda^*_{5,3,S_2}$	0.72	0.14	0.48	1.01	1.95	6
$\lambda^*_{6,3,S_2}$	0.49	0.54	-0.03	1.43	8.15	4
$\lambda^*_{7,3,S_2}$	0.66	0.10	0.48	0.86	1.60	8
$\lambda^*_{8,3,S_2}$	0.72	0.14	0.47	0.98	2.01	6
$\lambda^*_{9,3,S_2}$	0.37	0.34	-0.07	0.90	5.94	4
$\lambda^*_{10,3,S_2}$	0.61	0.23	0.21	0.98	2.97	5
$\lambda^*_{11,3,S_2}$	0.46	0.16	0.16	0.74	2.76	5

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\lambda^*_{12,3,S_2}$	1.03	0.21	0.62	1.38	2.24	6
$\lambda^*_{13,3,S_2}$	-0.01	0.01	-0.02	0.00	1.24	15
$\lambda^*_{14,2,S_2}$	0.42	0.17	0.14	0.76	2.22	6
$\lambda^*_{15,2,S_2}$	0.53	0.15	0.26	0.82	1.57	9
$\lambda^*_{16,2,S_2}$	0.27	0.10	0.06	0.47	1.21	19
$\lambda^*_{17,2,S_2}$	0.48	0.25	-0.00	0.78	4.75	4
$\sigma^2_{\epsilon_{1,1,S_1}}$	0.66	0.06	0.54	0.77	2.73	5
$\sigma^2_{\epsilon_{1,2,S_1}}$	0.74	0.07	0.58	0.84	3.02	5
$\sigma^2_{\epsilon_{1,3,S_1}}$	0.65	0.06	0.55	0.76	3.17	5
$\sigma^2_{\epsilon_{1,4,S_1}}$	0.84	0.04	0.75	0.92	1.20	18
$\sigma^2_{\epsilon_{1,5,S_1}}$	0.83	0.04	0.75	0.91	1.32	12
$\sigma^2_{\epsilon_{1,6,S_1}}$	0.45	0.15	0.30	0.75	6.36	4
$\sigma^2_{\epsilon_{1,7,S_1}}$	0.62	0.03	0.57	0.68	1.11	28
$\sigma^2_{\epsilon_{1,8,S_1}}$	0.48	0.03	0.43	0.54	1.19	18
$\sigma^2_{\epsilon_{1,9,S_1}}$	0.55	0.04	0.46	0.63	1.77	7
$\sigma^2_{\epsilon_{1,10,S_1}}$	0.64	0.10	0.45	0.76	5.28	4
$\sigma^2_{\epsilon_{1,11,S_1}}$	0.44	0.04	0.35	0.51	2.26	6
$\sigma^2_{\epsilon_{1,12,S_1}}$	0.86	0.04	0.78	0.94	1.26	14
$\sigma^2_{\epsilon_{1,13,S_1}}$	0.62	0.04	0.55	0.69	1.53	9
$\sigma^2_{\epsilon_{1,14,S_1}}$	0.84	0.04	0.77	0.92	1.18	19
$\sigma^2_{\epsilon_{1,15,S_1}}$	0.78	0.06	0.67	0.89	2.42	5
$\sigma^2_{\epsilon_{1,16,S_1}}$	0.84	0.04	0.77	0.91	1.09	35
$\sigma^2_{\epsilon_{1,17,S_1}}$	0.91	0.04	0.85	0.99	1.03	91
$\sigma^2_{\epsilon_{1,1,S_2}}$	0.68	0.08	0.56	0.82	3.11	5
$\sigma^2_{\epsilon_{1,2,S_2}}$	0.77	0.11	0.60	0.95	4.39	4
$\sigma^2_{\epsilon_{1,3,S_2}}$	0.00	0.00	0.00	0.00	2.56	5
$\sigma^2_{\epsilon_{1,4,S_2}}$	0.93	0.06	0.83	1.06	1.67	8
$\sigma^2_{\epsilon_{1,5,S_2}}$	0.72	0.05	0.63	0.81	1.93	7
$\sigma^2_{\epsilon_{1,6,S_2}}$	0.17	0.19	0.00	0.50	65.83	4
$\sigma^2_{\epsilon_{1,7,S_2}}$	0.44	0.02	0.40	0.49	1.40	11
$\sigma^2_{\epsilon_{1,8,S_2}}$	0.61	0.03	0.55	0.67	1.09	32

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.

Table 4: Posterior estimates of DLC-SEM with uncorrelated factors (continued)

	M	SD	2.5%	97.5%	$\hat{R}$	ESS
$\sigma^2_{\epsilon_{1,9,S_2}}$	0.53	0.07	0.42	0.64	3.81	5
$\sigma^2_{\epsilon_{1,10,S_2}}$	0.63	0.10	0.43	0.74	4.49	4
$\sigma^2_{\epsilon_{1,11,S_2}}$	0.55	0.03	0.49	0.60	1.35	12
$\sigma^2_{\epsilon_{1,12,S_2}}$	0.97	0.06	0.85	1.09	1.75	7
$\sigma^2_{\epsilon_{1,13,S_2}}$	0.00	0.00	0.00	0.00	3.25	5
$\sigma^2_{\epsilon_{1,14,S_2}}$	0.69	0.05	0.59	0.78	2.23	6
$\sigma^2_{\epsilon_{1,15,S_2}}$	0.92	0.05	0.83	1.01	1.34	12
$\sigma^2_{\epsilon_{1,16,S_2}}$	0.83	0.04	0.76	0.90	1.01	230
$\sigma^2_{\epsilon_{1,17,S_2}}$	0.67	0.04	0.60	0.74	1.47	10
$\Phi_1[1]$	0.01	0.00	0.01	0.01	1.93	7
$\Phi_1[2]$	0.04	0.01	0.02	0.06	2.73	5
$\Phi_1[3]$	0.03	0.00	0.02	0.03	1.30	13
$\Phi_1[4]$	0.31	0.18	0.13	0.77	1.33	12
$\Phi_1[5]$	0.26	0.16	0.10	0.68	1.17	24
$\Phi_1[6]$	0.29	0.13	0.12	0.62	1.06	54
$\Phi_2[1]$	8.52	2.56	4.93	14.62	2.77	5
$\Phi_2[2]$	0.12	0.06	0.06	0.26	3.69	5
$\Phi_2[3]$	1.42	0.36	0.84	2.23	1.38	11
$\Phi_2[4]$	2.88	1.81	0.83	7.38	1.49	9
$\Phi_2[5]$	1.81	1.20	0.49	4.89	1.24	16
$\Phi_2[6]$	4.84	2.79	1.24	11.32	1.19	18

Note. For item j that loads on factor k in state S_l, the following parameters are defined: regression coefficients for transition probabilities β_k, AR coefficients b_{k,S_l}, primary loadings λ_{j,k,S_l}, cross-loadings λ^*_{j,k,S_l}, residual variance $\sigma^2_{\epsilon_{1,j,S_l}}$, covariance matrix for factors Φ_1, covariance matrix for random effects Φ_2. The first primary loading of each factor were fixed to 1.