Latent Markov factor analysis

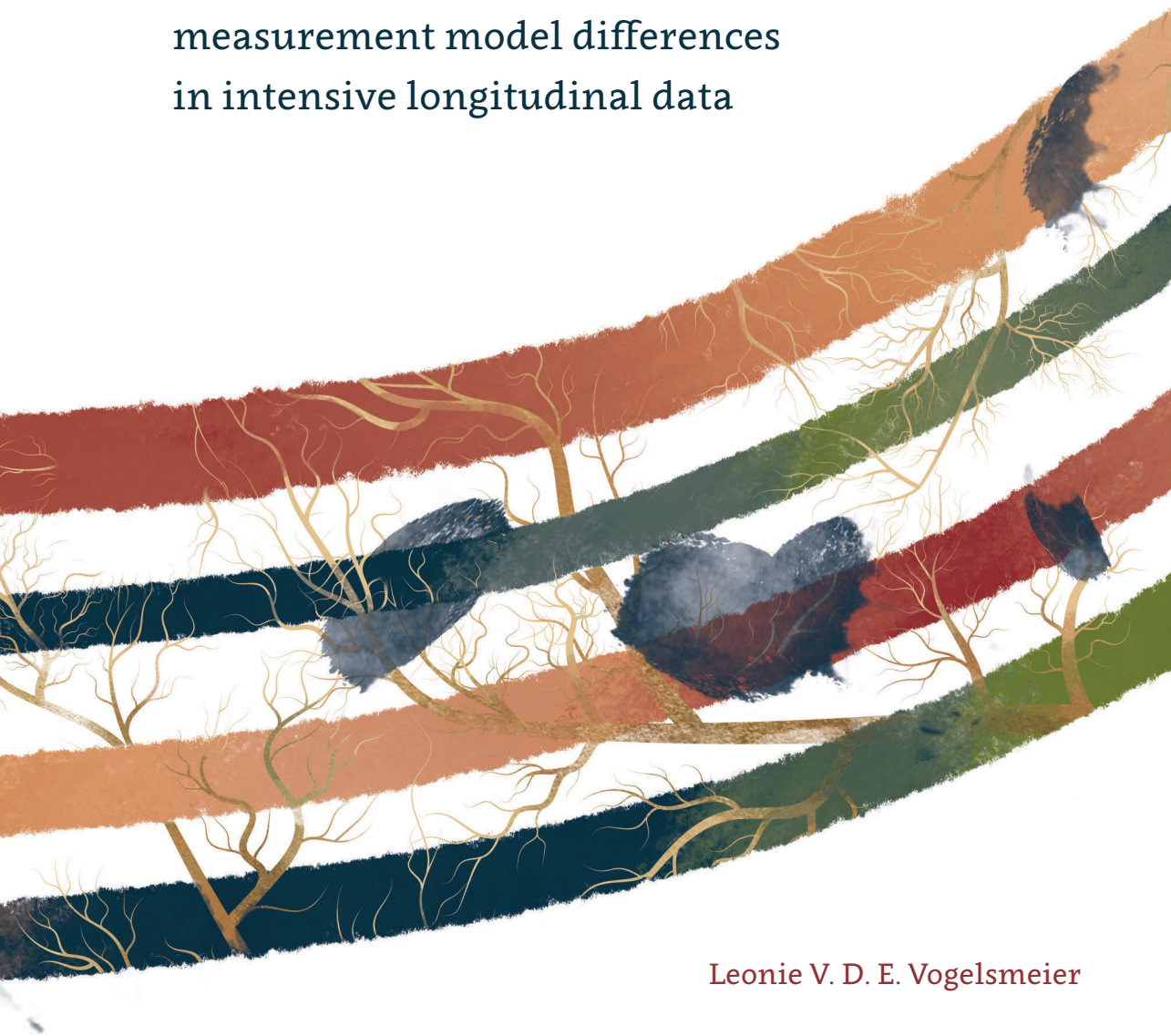
Vogelsmeier, L.V.D.E.

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

LATENT MARKOV FACTOR ANALYSIS

A mixture modeling approach for evaluating within- and between-person measurement model differences in intensive longitudinal data



Leonie V. D. E. Vogelsmeier

Latent Markov Factor Analysis

A mixture modeling approach for evaluating within- and between-person measurement model differences in intensive longitudinal data

Leonie Valentina Dorothea Edith Vogelsmeier

ISBN: 978-94-6416-941-6
Printed by: Ridderprint | www.ridderprint.nl
Cover design: Evelien Jagtman

Latent Markov Factor Analysis

A mixture modeling approach for evaluating within- and between-person measurement model differences in intensive longitudinal data

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W. B. H. J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op

vrijdag 14 januari 2022 om 13.30 uur

door

Leonie Valentina Dorothea Edith Vogelsmeier

geboren te Gummersbach, Duitsland

Promotor:	prof. dr. J. K. Vermunt (Tilburg University)
Copromotor:	dr. K. De Roover (Tilburg University)
Leden promotiecommissie:	prof. dr. E. Ceulemans (KU Leuven) prof. dr. E. L. Hamaker (Utrecht University) prof. dr. J. M. Wicherts (Tilburg University) dr. Z. Bakk (Leiden University)

To my parents,

Gabriele & Eberhard Vogelsmeier

Contents

1

General Introduction

1.1 Measuring Psychological Constructs

Studying psychological constructs, such as personality traits, mental illnesses, or well-being, is part of many fields in psychology (and, more generally, in the social and behavioral sciences; e.g., when studying trust in politics or the government). One of the first things that students in these fields learn is that such constructs are not directly observable but latent (e.g., it is not possible to hold a ruler on a person's head and say that he or she has a well-being level of 10). Therefore, one has to rely on questionnaire items that are supposed to measure certain constructs (Jöreskog & Sörbom, 1979; McNeish & Wolf, 2020). An essential aspect of assessing psychological constructs is to evaluate how well the constructs are measured by the items and, thus, whether the items can be considered valid measures of the constructs. This is often done by means of factor analysis (Lawley & Maxwell, 1962). In the resulting factor model or "measurement model" (MM), the factors correspond to the latent constructs and factor loadings indicate how strongly items are related to the underlying factors (i.e., how well the items measure the factors). Based on the MM, one can derive construct scores (or "factor scores") for each observation in the dataset and use them in subsequent analyses.

To better understand how to interpret a MM, consider the one depicted in Figure 1.1, which corresponds to a questionnaire with the five items "happy", "cheerful", "determined", "enthusiastic", and "excited" to assess positive affect (PA) and the four items "upset", "anxious", "nervous", and "jittery" to assess negative affect (NA). The arrows (representing factor loadings of the items) go from the latent constructs (circles) to the items (squares), which indicates that PA and NA influence the scores on the items. For example, if we observe high scores on positive emotions like "happy" or "cheerful", we assume that this is due to a high score on the PA construct.

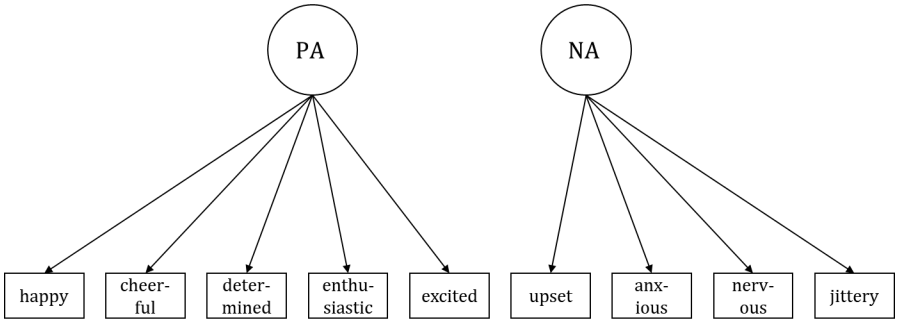


Figure 1.1. Measurement model underlying the measures of the latent constructs positive affect (PA) and negative affect (NA).

1.2 Dynamics in Psychological Constructs

Over the last two decades, studying the dynamic process of psychological constructs (and thus how psychological constructs evolve within persons over time) gained tremendous popularity, which was facilitated by the steep increase in technological possibilities (Hamaker & Wichers, 2017). More specifically, instead of collecting cross-sectional data (consisting of measures for many subjects at a single time-point), researchers collect intensive longitudinal data (ILD; consisting of many repeated measures, usually in everyday situations, for a single subject or for multiple subjects; Ryan, 2020). Such data is easily gathered by means of experience sampling methodology (ESM; Scollon, Kim-Prieto, & Diener, 2003; van Roekel, Keijsers, & Chung, 2019), in which researchers administer their questionnaires to many subjects at the same time and over a longer period of time via smartphone apps at random or event-based time-points. This allows delving into the dynamics of constructs as they take place (Hamaker, 2012; Molenaar & Campbell, 2009; Ryan, 2020). Instead of only answering "who" experiences certain emotions, symptoms, or thoughts, ILD goes beyond this "between-person" approach and also allows to answer "in which contexts" they occur (i.e., taking on a within-person approach; Csikszentmihalyi & Larson, 2014; Scollon et al., 2003). Results from ILD are also clinically relevant, for example for tailoring (depression) treatments to subject-specific dynamics in (depression) symptoms (van der Krieke et al., 2015) or for generating personalized life-style advices (Van Roekel et al., 2017).

To clarify the data structure, consider, for instance, that work- and organizational psychologists ask employees to complete the previously described questionnaire (measuring PA and NA; Figure 1.1) at three random time-points throughout a workday over a course of 30 days. The psychologists' goal is to understand how employees' affective well-being evolves in different contexts (e.g., during teamwork, during lunchbreaks with colleagues, during individual work, just before giving a presentation etc.) and to use the results to create personalized advice for improving general mood, which, in turn, should improve a productive working environment.

1.3 Measurement Invariance When Studying Dynamics in Psychological Constructs

Besides coming with new possibilities, ILD data also come with new psychometric challenges: the measurement quality of the psychological constructs might differ across subjects and situations. On the one hand, subjects might use different response styles at different points in time. For example, some subjects could generally be more inclined to use the extreme values on the measurement scale (Moors, 2003; Morren, Gelissen, & Vermunt, 2011). Others might start applying such an "extreme response style" once they are no longer motivated to repeatedly complete the ESM questionnaire. On the other

hand, the answer to an item might not (only) be affected by the underlying construct, but (also) by the cultural background of a subject (Van De Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015) or the tendency to put items into a different perspective items after a life changing event (Lommen, van de Schoot, & Engelhard, 2014). Differences and/or changes in response styles and substantive item interpretations alter the MM differently across subjects and/or time-points. For example, response styles increase the size of the loadings or come with an additional response style factor (Billiet & McClendon, 2000; Cheung & Rensvold, 2000). Changes in substantive item interpretations can in- or decrease the size of the loadings or cause changes in the configuration of the MM such that an item becomes a measure of a different construct (Oort, Visser, & Sprangers, 2005), which can go as far as changing the meaning of an entire construct (Lommen et al., 2014).

For instance, the work- and organizational psychologists in our ILD example considered the item "excited" a positive emotion. Indeed, this emotion is often used to expresses eagerness and happiness (Heininga & Kuppens, 2020). However, besides the positive meaning, "excited" may also be seen as nervousness or agitation (Galinha, Pereira, & Esteves, 2013; Mackinnon et al., 1999). The item would then be a measure of NA rather than PA, which is depicted in Figure 1.2. The item interpretation might not only differ across employees but also within an employee throughout the day. For instance, during a lunch break, the employee might be excited in a positive way when talking about an upcoming family vacation, whereas the employee may be excited in a negative way prior to giving a presentation. Thus, the employee would switch back and forth between two MMs.

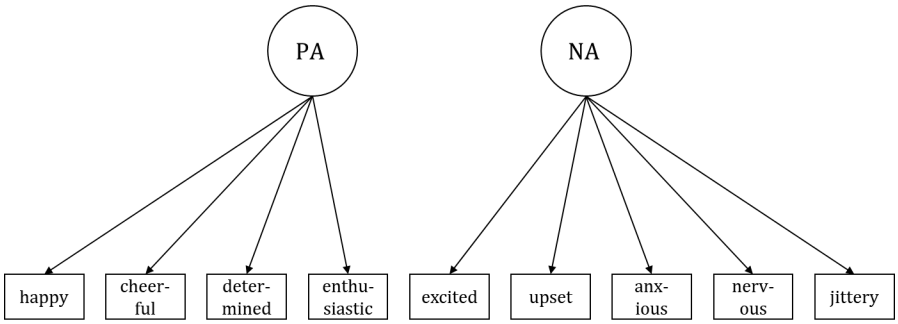


Figure 1.2. Measurement model underlying the measures of the latent constructs positive affect (PA) and negative affect (NA) when being "excited" is interpreted as nervousness or agitation.

If the MM is not the same across subjects and time-points, "measurement invariance" (Meredith, 1993; Meredith & Teresi, 2006) does not hold. However, measurement invariance is a key assumption for validly investigating dynamics in psychological constructs because, if it holds, we know that the same constructs are measured in the exact same way across subjects and time. If researchers fail to detect non-invariance of the MM (or if it is simply ignored), inferences about the dynamics in psychological constructs cannot be trusted (e.g., observing a change in construct means could also be due to changes in the way the construct is measured; Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014; Guenole & Brown, 2014; Van De Schoot et al., 2015).

Consider, for example, that the true NA score for some employees in our ILD example is high prior to giving a presentation (say, their true NA score is a 10) and that the true PA score is low (say, their true PA score is a 3). If the MM in Figure 2.2 is underlying their responses, the employees would give a high rating on the item "excited" because their responses are influenced by their high NA. If the context-dependent interpretation of the item "excited" was ignored, that is, the psychologists would assume that the MM in Figure 2.1 is still underlying the responses of the employees, the psychologists would overestimate the PA of the employees because the high item ratings on "excited" would erroneously be included in their PA scores. In turn, the psychologists would draw wrong conclusions (e.g., that employees who dislike being in the focus should be asked to give a presentation more often).

If differences in the underlying MM would be detected, researchers could safeguard the validity of their conclusions by accounting for non-invariance in their analyses (Byrne, Shavelson, & Muthén, 1989) or by using only observations for which invariance holds. In any case, researchers could learn from it for future studies. For instance, prior to data collection, they could replace or rephrase problematic items (Galinha et al., 2013), or shorten the questionnaire (which can considerably improve the data quality; Eisele et al., 2020). Furthermore, during data collection, they could add additional motivational rewards for participants who are at risk for adopting response styles. Importantly, revealing differences in the MM can also be interesting in its own right. For instance, MM changes might indicate the onset of a mental state, such as mania (Hofmann & Meyer, 2006) or a posttraumatic stress disorder (Lommen et al., 2014). Additionally, one can learn about substantively interesting interindividual differences in how an item functions.

For many researchers, it may not come as a surprise that item interpretations and response styles can be as dynamic as the constructs that the questionnaires purport to measure. Why is it then that measurement invariance often receives little attention in ILD? The answer to this question is twofold. First, researchers often simply employ

established questionnaires (like the Positive and Negative Affect Scale; Watson, Clark, & Tellegen, 1988 or the Beck Depression Inventory; Beck, Rush, Shaw, & Emery, 1979) for which measurement invariance has previously been investigated for cross-sectional data (i.e., for the MM across persons). However, as explained above, the MM can differ across persons and change within a person depending on context-specific influences.

Second, existing methods to detect measurement invariance in ILD are limited because they can only test whether or not invariance is violated either across subjects—thereby assuming invariance across time—or whether it is violated across time—thereby assuming invariance across subjects (Adolf et al., 2014). Furthermore, if measurement invariance is violated, these methods give no insights into sources of non-invariance (i.e., for which subjects/for which time-points the MM varies), making it difficult for researchers to proceed. More specifically, in order to identify the sources of non-invariance, researchers have to conduct many pairwise comparisons of the subject- or time-point-specific MMs because ILD typically contain many observations from more than just a few subjects (e.g., to identify non-invariance across 100 subjects, one would have to make 4950 comparisons). Moreover, the invariance tests are confirmatory in that they assume the pattern of (non-zero) loadings of the MM (i.e., the "configural model") to be known in advance. However, this assumption is easily violated, as was shown in our example where the loading of the item "excited" changed from PA to NA.

Thus, what is actually required is an exploratory approach that detects unknown heterogeneity in the MM across subjects and time. For instance, if the item "excited" indeed has two meanings, then there are two MMs and thus two factor models. This can be conceptualized as persons belonging to one of two latent classes at a certain time-point. Some persons will always answer according to MM 1, others will always answer according to MM 2, and yet others will switch between the two MMs (and thus between the two classes). An efficient tool to detect such unobserved heterogeneity in factor models in cross-sectional data is mixture factor analysis (Lubke & Muthén, 2005; McNicholas, 2016), which classifies subjects according to their underlying factor model into a few latent classes. Each class then has its own MM. However, a methodology that allows subjects to switch between different classes (i.e., MMs) over time is lacking.

1.4 Latent Markov Factor Analysis for Evaluating Measurement (Non-) Invariance

The main objective of this dissertation is to develop and evaluate a method to trace measurement non-invariance in ILD in order to help researchers safeguard valid conclusions about dynamics in psychological constructs and to extend their toolbox for obtaining substantive insights into the dynamics of underlying MMs. To this end, we extend mixture factor analysis to accommodate ILD. More specifically, the new method,

called latent Markov factor analysis (LMFA), models changes in the MM by means of a dynamic latent class or mixture model (known as latent transition or latent Markov model; LMM; Bartolucci, Farcomeni, & Pennoni, 2015; Visser, Raijmakers, & van der Maas, 2009; Zucchini, MacDonald, & Langrock, 2016) and, thereby, clusters observations into a few latent "states" (i.e., dynamic latent classes). Furthermore, factor analysis (FA; Lawley & Maxwell, 1962) within the states reveals the structure of the MM. Note that, to detect all kinds of MM differences—including the number and nature of factors—exploratory FA and not the more restrictive confirmatory FA is employed. As a result, LMFA provides crucial information on how and for which subjects and/or time-points the MM differs.

To clarify how LMFA works, consider the graphical representation in Figure 1.3, which shows (for simplicity) three transitions between the previously introduced MMs for a single employee. LMFA would indicate that the employee is in MM-state 1 at the first ten time-points (i.e., $S_1, \dots, S_{10} = 1$), then changes to MM-state 2 for the next ten time-points ($S_{11}, \dots, S_{20} = 2$), and then changes back to MM-state 1 for the remaining time-points ($S_{21}, \dots, S_{30} = 1$). By looking at the state memberships, the psychologists can see which observations are comparable and, by investigating the state-specific MMs, they can see that the interpretation of the item "excited" changes over time. Based on this information, the psychologists could, for example, decide to remove this item prior to creating construct scores (other possibilities to deal with non-invariance will be described throughout this dissertation).

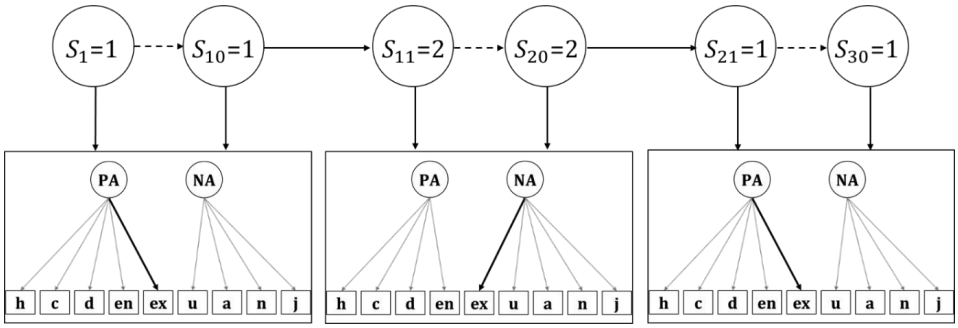


Figure 1.3. Measurement model underlying the measure of the latent constructs positive affect (PA) and negative affect (NA); h = happy, c = cheerful, d = determined, en = enthusiastic, ex = excited, u = upset, a = anxious, n = nervous, j = jittery.

1.5 Outline of the Thesis

In **Chapter 2,** we introduce and extensively describe the new method LMFA and, thereby, lay the foundation for the subsequent chapters. Additionally, we relate LMFA to other methods that could be potentially useful for tracing MM differences or changes in ILD and

explain why these methods are, nevertheless, not sufficient for this purpose. Moreover, we show how the model is estimated by means of a full information maximum likelihood (FIML) estimation in the software Latent GOLD (Vermunt & Magidson, 2016). In an elaborate simulation study, we demonstrate how LMFA performs in recovering parameters in favorable and unfavorable conditions (e.g., with regard to differences between the state-specific MMs and the sample size in terms of the number of subjects and observations). Based on the results, we present recommendations for empirical practice. Furthermore, we address the problem of model selection, which is complicated by LMFA's flexibility to detect all kinds of MM differences, even regarding the number of factors. More specifically, both the number of states and the number of factors per state are unknown in advance and need to be specified. Therefore, in an additional simulation study, we evaluate the Bayesian Information Criterion (BIC; Schwarz, 1978) for finding the best model complexity. Finally, we illustrate the empirical value by means of applying LMFA to a real ESM dataset containing measures of adults suffering from anhedonia. This chapter appeared in *Structural Equation Modeling: A Multidisciplinary Journal*.

In **Chapter 3**, we extend LMFA to accommodate unequally spaced observations by means of a so-called "continuous-time" approach (Böckenholt, 2005; Jackson & Sharples, 2002) because the regular "discrete-time" approach (as employed in Chapter 2) assumes that measurement occasions are equidistant, which is frequently violated in ILD. First, beeps to complete the questionnaires are typically sent out at random (or event-based) time-points. Consequently, the intervals between the measurement occasions differ across subjects and within subjects over time. Second, night intervals are usually longer than intervals throughout the day. Third, when responding is inconvenient (e.g., while giving a presentation) participants skip measurement occasions, which leads to missing data that, in turn, also increase the interval length. When applying LMFA to ILD, it is important to account for differences in intervals because closely spaced observations are more strongly associated than more distant observations (i.e., transitioning to another state is less likely for a 15-minute interval than for a 1-day interval). In a simulation study, we compare the continuous-time approach to the discrete-time approach in the presence of unequal intervals. Moreover, we apply the continuous-time model to diary data of subjects with depressive symptoms and thereby show that LMFA can be applied not only to modern ILD collected by means of ESM but also to other types of longitudinal data (e.g., weekly or end-of-day reports), given that the number of observations is large enough. This chapter was published in *Methodology*.

Chapter 4 is concerned with the inclusion of explanatory variables (or "covariates") in the LMFA model because, usually, researchers do not only want to know if measurement invariance holds and how to proceed with their data analysis if measurement invariance does not hold. They are especially interested to learn what

drives the (between-subject differences in) MM dynamics. For instance, adopting response styles could depend on the experienced stress while completing a questionnaire and item interpretation could depend on the general ability to differentiate between emotions. In order to estimate the LMFA model with covariates, we apply a three-step approach (Di Mari, Oberski, & Vermunt, 2016; Vermunt, 2010), which circumvents the cumbersome model and covariate selection procedure that is inherent to the FIML estimation. The three-step approach splits the estimation into three parts: (1) Evaluating the MMs while disregarding transitions between the MMs and covariate effects on the state memberships. (2) Assigning observations to the most likely MM-state. (3) Using these assignments to estimate the (covariate-specific) transitions between MMs (while keeping the MMs fixed). We explain that the three-step approach is generally a valid alternative to the FIML estimation when correcting for uncertainty in the state assignments of the second step and, in a simulation study, we compare the performance of this three-step approach to the performance of the FIML estimation in the context of LMFA. Finally, by means of applying three-step LMFA to a real ESM dataset containing emotion ratings of young adolescents, we demonstrate how researchers can draw substantive conclusions about MM differences from including both subject- and situation-specific covariates. This chapter was accepted for publication in a similar form in *Multivariate Behavioral Research*.

In **Chapter 5**, we address the challenge of dealing with ordinal data. The original LMFA assumes normally distributed continuous responses. In empirical practice, however, items are often measured with only a few response categories (e.g., with Likert-type scales) and the responses can be heavily skewed. Applying LMFA to such data can lead to inaccurate parameter estimates (Kappenburg -ten Holt, 2014; Rhemtulla, Brosseau-Liard, & Savalei, 2012; Vermunt & Magidson, 2005). Therefore, we present the extension latent Markov latent trait analysis (LMLTA), which adequately deals with such data by replacing the FA model within the states with "latent trait" (or "item response theory") model that directly treats the responses as ordinal (Heinen, 1996). We explain the similarities and differences between the FA model and the latent trait model and illustrate the extension by means of applying LMLTA to ESM data containing observations of young adolescents at risk for developing depression. This chapter appeared in *Evaluation & the Health Professions*.

In **Chapter 6**, we present a tutorial for our R package *lmfa*. We created the R package to present a free open-source alternative to Latent GOLD for performing LMFA. More specifically, we employed the continuous-time variant of LMFA by means of the three-step approach so that researchers can validly analyze their data with unequally-spaced observations and also easily relate explanatory variables to the state memberships. In this chapter's tutorial, we show which steps have to be taken when

performing the analysis with *lmfa* by means of a synthetic ESM example dataset, which is part of the package. Moreover, we provide ideas on how to proceed based on the results of LMFA. In particular, we discuss possibilities to account for different levels of non-invariance in subsequent analyses to investigate (between-person differences in) how within-person dynamics in the latent constructs develop over time. This chapter is submitted for publication in *Behavior Research Methods*.

Finally, in the epilogue in **Chapter 7**, we begin by answering some questions about LMFA that remain open after the first six chapters of the dissertation. Then, we discuss limitations of LMFA and outline possible solutions to solve some of these shortcomings in future research. We conclude by explaining that LMFA is not only a precursor for making decisions about how to proceed with the data analysis but that the LMFA framework also has the potential for providing relevant insights into psychological phenomena, which could possibly be relevant for empirical practice, such as personalized therapy.

2

Latent Markov Factor Analysis

This chapter is published as 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*, 557–575. doi:10.1080/10705511.2018.1554445

Abstract

When time-intensive longitudinal data is used to study daily-life dynamics of psychological constructs (e.g., well-being) within persons over time (e.g., by means of experience sampling methodology), the measurement model (MM)—indicating which constructs are measured by which items—can be affected by time- or situation-specific artefacts (e.g., response styles, altered item interpretation). If not captured, these changes might lead to invalid inferences about the constructs. Existing methodology can only test for a priori hypotheses on MM changes, which are often absent or incomplete. Therefore, we present the exploratory method "latent Markov factor analysis" (LMFA), wherein a latent Markov chain captures MM changes by clustering observations per subject into a few states. Specifically, each state gathers validly comparable observations and state-specific factor analyses reveal what the MMs look like. LMFA performs well in recovering parameters under a wide range of simulated conditions and its empirical value is illustrated with an example.

2.1 Introduction

Time-intensive longitudinal data for studying daily-life dynamics of psychological constructs (such as well-being or positive affect) within persons allows to delve into time- or situation-specific effects (e.g., stress) on the (e.g., emotional) experiences of a large number of subjects (Larson & Csikszentmihalyi, 2014). The go-to research design to collect such data is experience sampling methodology (ESM; Scollon et al., 2003). Participants repeatedly answer questionnaires at randomized or event-based time-points via smartphone apps, for example, eight times a day over a few weeks.

While the technology for collecting ESM data is readily available, the methodology to validly analyze this data is lagging behind. This chapter provides an upgrade of the methodology by presenting a novel method for tracking and diagnosing changes in measurement models (MMs) over time. The MM is the model underlying a participant's answers and indicates which unobservable or latent variables (i.e., psychological constructs) are measured by which items. Traditionally, it is evaluated by factor analysis (FA; Lawley & Maxwell, 1962), where the factors correspond—ideally—to the hypothesized constructs. Factor loadings express the degree to which each of the items measure a factor and thus how strongly an item relates to an underlying factor. In order to meaningfully compare constructs over time, the MM needs to be invariant across measurement occasions (Adolf et al., 2014). However, measurement invariance (MI) does not always hold over time because the MM likely changes over the course of an ESM study. First, in ESM, the measurement quality is undermined by time- or situation-specific artefacts such as response styles (RSs; Moors, 2003; Paulhus, 1991). Indeed, participants fill in their questionnaires repeatedly in various, possibly distracting, situations (e.g., during work) or lose motivation to repeatedly answer questions, which may drive the tendency to, for example, use the extreme response categories only (extreme RS; Moors, 2003; Morren et al., 2011). Second, substantive changes may occur over time in what questionnaire items are measuring. For example, depending on the context or mental state, an item may become more important for the measured construct (i.e., loading increases) or (also) an indicator of another construct (i.e., loads strongly on another factor) (reprioritization or reconceptualization; Oort et al., 2005). Moreover, the nature of the measured constructs might change entirely; e.g., when positive and negative affect factors are replaced by high and low arousal factors (Feldman, 1995). In any case, when ignoring changes in the MM, changes in the scores will be interpreted as changes in the psychological constructs, although they are (partly) caused by RSs or changed item interpretation.

To safeguard validity of their time-intensive longitudinal studies, substantive researchers need an efficient approach to evaluate which MMs are underlying the data

and for which time-points they apply, so they can gain insight into which artefacts and substantive changes are at play and when. Researchers can take these insights into account when analyzing the data, when setting up future projects or to derive new substantive findings from the MM changes. To meet this need, we present latent Markov factor analysis (LMFA)[1], which combines two building blocks to model MM changes within subjects over time: (1) *Latent Markov modeling* (LMM; Bartolucci, Farcomeni, & Pennoni, 2014; Collins & Lanza, 2010) clusters time-points into states according to the MMs and (2) factor analysis (FA; Lawley & Maxwell, 1962) evaluates which MM applies for each state. Note that LMFA can be applied for single cases, when enough observations are available for that one subject.

Within the states of LMFA, exploratory factor analysis (EFA) rather than confirmatory factor analysis (CFA) is used. In CFA, users have to specify which items are measuring which factors based on a priori hypotheses. This implies that certain item-factor relations are assumed to be absent and the corresponding factor loadings are set to zero. Thus, for a large part, CFA already imposes a certain MM and thus limits the changes in the MM that can be found. In contrast, EFA estimates all factor loadings and, thus, explores all kinds of (unknown) MM changes, including changes in cross-loadings (i.e., items loading on more than one factor) or even in the nature and number of factors (e.g., an additional RS factor). However, if desired, CFA can be used within the states.

An existing method to evaluate whether MI holds over time is longitudinal structural equation modeling (LSEM; Little, Preacher, Selig, & Card, 2007). However, this method merely tests whether MI across time-points holds for all individuals simultaneously, without directly providing insight in for which measurement occasions invariance is violated and what the alternative MMs look like. In contrast to LMFA, LSEM provides no clues for understanding or dealing with the non-invariance. Also, it applies CFA and thus already assumes a certain factor structure, and is therefore too restrictive to detect many MM differences. A few methods exist that combine FA with LMM and thus could potentially be useful for identifying violations of MI over time (Asparouhov, Hamaker, & Muthén, 2017; Song, Xia, & Zhu, 2017; Xia, Tang, & Gou, 2016).[2] However,

[1] LMFA builds upon Mixture simultaneous factor analysis (MSFA; De Roover, Vermunt, Timmerman, & Ceulemans, 2017), which captures differences in the factor model between groups. Whereas MSFA typically models the data of subjects nested within groups, LMFA specifically deals with observations nested within subjects and it allows subjects to switch between different MMs over time.

[2] Note that the overview of existing methods focusses on FA-based methods and thus overlooks switching principal component analysis (SPCA; De Roover, Timmerman, Van Diest, Onghena, & Ceulemans, 2014), which is a deterministic method similar to LMFA that can be used to take the first steps towards detecting MM changes over time, yet only for single-subject data. However, SPCA uses component instead of factor analysis. Although components and factors are similar (Ogasawara, 2000; Velicer & Jackson, 1990; Velicer, Peacock, & Jackson, 1982), components do not correspond to latent variables (Borsboom, Mellenbergh, & van Heerden, 2003) and are thus less ideal for evaluating (changes in) MMs.

these methods also apply CFA, making them too restrictive to detect all kinds of MM differences. In contrast, factor-analyzed hidden Markov modeling (FAHMM; Rosti & Gales, 2002) is similar to LMFA because it combines EFA with LMM, but was developed merely for accommodating LMM estimation when conditional independence is violated among many variables, using the state-specific FA to reduce the number of parameters of the state-specific covariance matrices rather than being the point of interest (Kang & Thakor, 2012; Rosti & Gales, 2002). Also, FAHMM cannot analyze multiple subjects simultaneously. Thus, LMFA may be conceived as a multisubject extension of FAHMM, tailored to tackle measurement non-invariance in ESM data.

The remainder of this chapter is organized as follows: Section 2.2 describes the multisubject longitudinal data structure, an empirical example, and the LMFA model specifications and estimation. Section 2.4 presents a simulation study, evaluating the goodness-of-recovery of states and state-specific MMs under several conditions as well as model selection. Section 2.5 illustrates LMFA with an application. Section 2.6 concludes with some points of discussion and directions for future research.

2.2 Methods

2.2.1 Data Structure and Motivating Example

Like in ESM, we assume repeated measures data where observations are nested in subjects. For each measurement occasion, data on multiple continuous variables are available. The observed scores are indicated by y_{ijt}, where $i = 1, \dots, I$ refers to subjects, $j = 1, \dots, J$ to items, and $t = 1, \dots, T$ to time-points, where the latter may differ across subjects (i.e., T_i) but we mostly omit the index i for simplicity of notation. The $J \times 1$ vector $\mathbf{y}_{it} = (y_{i1t}, y_{i2t}, \dots, y_{iJt})'$ contains the multivariate responses for subject i at time-point t and the $T \times J$ dataset $\mathbf{Y}_i = (\mathbf{y}_{i1}', \mathbf{y}_{i2}', \dots, \mathbf{y}_{iT}')'$ contains data for subject i for all time-points T.

To clarify the data structure and illustrate the problem of measurement non-invariance, consider the ESM data of the 'No Fun No Glory' study described in more detail by Van Roekel et al. (2017). In brief, the data contained repeated emotion measures of 69 young adults with persistent anhedonia, which is the diminished pleasure in response to previously enjoyable experiences and one of the core symptoms of depression (American Psychiatric Association, 2013; Treadway & Zald, 2011). Over a course of about three months, every evening, the participants rated on a Visual Analogue Scale, ranging from 0 ("Not at all") to 100 ("Very much"), how much they had felt each of 18 emotions (listed in

Table 2.3, which is further described in Section 2.5) in the past six hours.[3] The number of repeated measures ranged from 86 to 132 (M = 106.86, SD = 8.21) and resulted in 7373 total observations of which 557 were missing.[4] After the first month, the participants randomly received (1) no intervention (n = 22), (2) a personalized lifestyle advice (PLA) (n = 23), or (3) a PLA and a tandem skydive (PLA & SkyD) (n = 24) to potentially reduce anhedonia. After the second month, all participants chose one of the interventions, regardless of their first one (no: n = 3; PLA: n = 17; PLA & SkyD: n = 49). In their original study, Van Roekel et al. (2017) investigated whether the interventions decreased anhedonia, thereby assuming the two underlying factors "positive affect" ("PA") and negative affect ("NA"). However, if the MM changes over the course of participation (e.g., due to the interventions) conclusions about changes in PA and NA may be invalid. In Section 2.5, LMFA is used to trace potential MM changes in this data.

2.3 Latent Markov Factor Analysis

In this section, we introduce LMM (2.3.1) before describing LMFA in more detail (2.3.2).

2.3.1 Latent Markov Modeling

The LMM (also a hidden Markov or latent transition model; Bartolucci et al., 2014; Collins & Lanza, 2010) captures unobserved heterogeneity or changes over time by means of latent states. In contrast to standard latent class models (Hagenaars & McCutcheon, 2002; Lazarsfeld & Henry, 1968), which identify subgroups or so-called latent classes within a population (e.g., high or low risk for depression), a LMM allows respondents to transition between latent states over time and, thus, to switch between subgroups (e.g., from a high risk to a low risk subgroup). Thus, the states may be conceived as dynamic latent classes. Specifically, the LMM is a probabilistic model where the probability of being in a certain state at time-point t depends only on the state of the previous time-point $t-1$ (*first-order Markov assumption*). Furthermore, the responses at time-point t depend only on the state at time-point t (*local independence assumption*; Bartolucci, 2006; Vermunt, Langeheine, & Böckenholt, 1999). The joint probability of observations and states for subject i is then:

[3] In total, participants rated their emotions three times a day with fixed 6-hour intervals. In the morning and midday, they rated their "momentary" emotions and in the evening, they rated their emotions "since the last measure". To have comparable and evenly spaced measures, we focused on the evening measures.
[4] Missing data will be directly handled in the model estimation by considering only observed values.

$$p(\mathbf{Y}_i, \mathbf{S}_i) = \overbrace{p(\mathbf{s}_{i1})}^{\text{initial state probabilities}} \prod_{t=2}^{T} \overbrace{p(\mathbf{s}_{it}|\mathbf{s}_{it-1})}^{\text{transition probabilities}} \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_{it}|\mathbf{s}_{it})}^{\text{response probabilities}}, \tag{2.1}$$

where $\mathbf{s}_{it}$ are $K \times 1$ binary variables indicating whether an observation belongs to a state or not and $\mathbf{S}_i = (\mathbf{s}_{i1}, \mathbf{s}_{i2}, \ldots, \mathbf{s}_{iT})$ is the subject-specific state-membership matrix. In the following, $k = 1, \ldots, K$ refers to the states and, if $s_{itk} = 1$, subject i is in state k at time-point t. Equation 2.1 includes three types of parameters: (1) The *initial state probabilities* indicate the probabilities to start in a certain state, $p(s_{i1k} = 1)$, and thus how the subjects are distributed across the states at $t = 1$. They are often denoted as π_k, with $\sum_{k=1}^{K} \pi_k = 1$, and are gathered in a $K \times 1$ vector $\boldsymbol{\pi}$. (2) The *transition probabilities* indicate the probabilities of being in a certain state at time-point t conditional on the state at $t-1$, $p(s_{itk}|s_{it-1,l})$, where $l = 1, \ldots, K$. These may be denoted as a_{lk}, with $\sum_{k=1}^{K} a_{lk} = 1$, and are collected in a $K \times K$ transition probability matrix $\mathbf{A}$. The transition probabilities are often assumed to be *homogeneous* (i.e., invariant) across time (and subjects). The resulting sequence of states is called a *latent Markov chain* (LMC). (3) The *response probabilities* indicate the probability of a certain item response given the state at time-point t, $p(\mathbf{y}_{it}|\mathbf{s}_{it})$, which correspond to the multivariate normal density for continuous responses.

2.3.2 Latent Markov Factor Analysis

In LMFA, a LMM is used to capture the changes in MMs over time and FA (Lawley & Maxwell, 1962) is applied per state to model the state-specific MMs. The latter is given by:

$$\mathbf{y}_{it} = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \mathbf{f}_{it} + \mathbf{e}_{it}, \tag{2.2}$$

where $\boldsymbol{\Lambda}_k$ is a state-specific $J \times F^k$ loading matrix; $\mathbf{f}_{it}$ is a subject-specific $F^k \times 1$ vector of factor scores at time-point t (where F^k is the state-specific number of factors); $\boldsymbol{\nu}_k$ is a state-specific $J \times 1$ intercept vector; and $\mathbf{e}_{it}$ is a subject-specific $J \times 1$ vector of residuals at time-point t. The distributional assumptions are: $\mathbf{f}_{it} \sim MVN(0; \boldsymbol{\Psi}_k)$ and factor scores are thus centered around zero and $\mathbf{e}_{it} \sim MVN(0; \mathbf{D}_k)$, where $\mathbf{D}_k$ contains the unique variances d_{kj} on the diagonal and zeros on the off-diagonal. To partially identify the model, factor variances in $\boldsymbol{\Psi}_k$ are restricted to one and the remaining rotational freedom is dealt with by means of criteria to optimize the simple structure or between-state agreement of the factor loadings, such as varimax (Kaiser, 1958), oblimin (Clarkson & Jennrich, 1988) or generalized Procrustes (Kiers, 1997).

From Equation 2.2, it is clear that the states may differ in terms of their intercepts $\boldsymbol{\nu}_k$, loadings $\boldsymbol{\Lambda}_k$, unique variances $\mathbf{D}_k$, and/or factor covariances $\boldsymbol{\Psi}_k$. This implies that LMFA allows to explore all levels of measurement non-invariance at once. This is: (1) *configural invariance* (invariant number of factors and pattern of zero loadings), (2) *weak factorial invariance* (invariant non-zero factor loadings), (3) *strong factorial invariance* (invariant item intercepts), and (4) *strict factorial invariance* (invariant unique variances). Conveniently, in any case, the strictest level of invariance applies within each state (for more details, see Little et al., 2007; Meredith, 1993; Meredith & Teresi, 2006; Schaie, Maitland, Willis, & Intrieri, 1998). Figure 2.1 illustrates how LMFA captures the different levels of non-invariance over time based on an example of what might happen in the empirical data by comparing the state-1 MM respectively to the state-2 and state-3 MMs, with dashed lines representing parameter changes.

The depicted loadings can be thought of as standardized rotated loadings higher than, for example, .4 in absolute value (Stevens, 1992). We start by comparing the state 1-MM to the state-2 MM. Here, configural invariance is violated because a third factor (high arousal; HA) appears, implying that the state-1 items measuring either PA or NA with loadings λ_{141}, λ_{151}, and λ_{162} measure another construct (i.e., HA) in state 2 (now with loadings λ_{242}, λ_{252}, and λ_{262}). This also changes the meaning of the other factors into low arousal PA (LA-PA) and low arousal NA (LA-NA)). Next, we compare the state-1 MM with the state-3 MM. Firstly, weak factorial invariance is violated here because λ_{111} differs from λ_{311} and thus, the items measure PA and NA differently. Secondly, strong factorial invariance is violated because ν_{12} differs from ν_{22}. Note that, when weak invariance appears to hold, properly assessing strong invariance would require re-estimating the model with invariant factor loadings across the states and non-zero state-specific factor means. Finally, strict factorial invariance is violated because e_{11} differs from e_{31}. Usually, strong factorial invariance is said to be sufficient for comparing latent constructs over time, that is, differences in factor means then correspond to actual changes in the latent variables.

It is important to note that the subjects do not have to go through all the states nor do they have to go through the states in the same order. Relatedly, LMFA does not assume homogeneous transition probabilities across subjects but allows for subject-specific $\mathbf{A}_i$ matrices, implying that some transition probabilities may be zero for a certain subject if that subject does not go through a particular state. This is because subjects likely differ in how stable they respond to questionnaires (e.g., some people might switch more between contexts than others or may be more sensitive to contextual influence or distractions). The transition process $\mathbf{A}_i$ is assumed to be time-homogeneous for each subject, although this is an assumption that might be relaxed in the future.

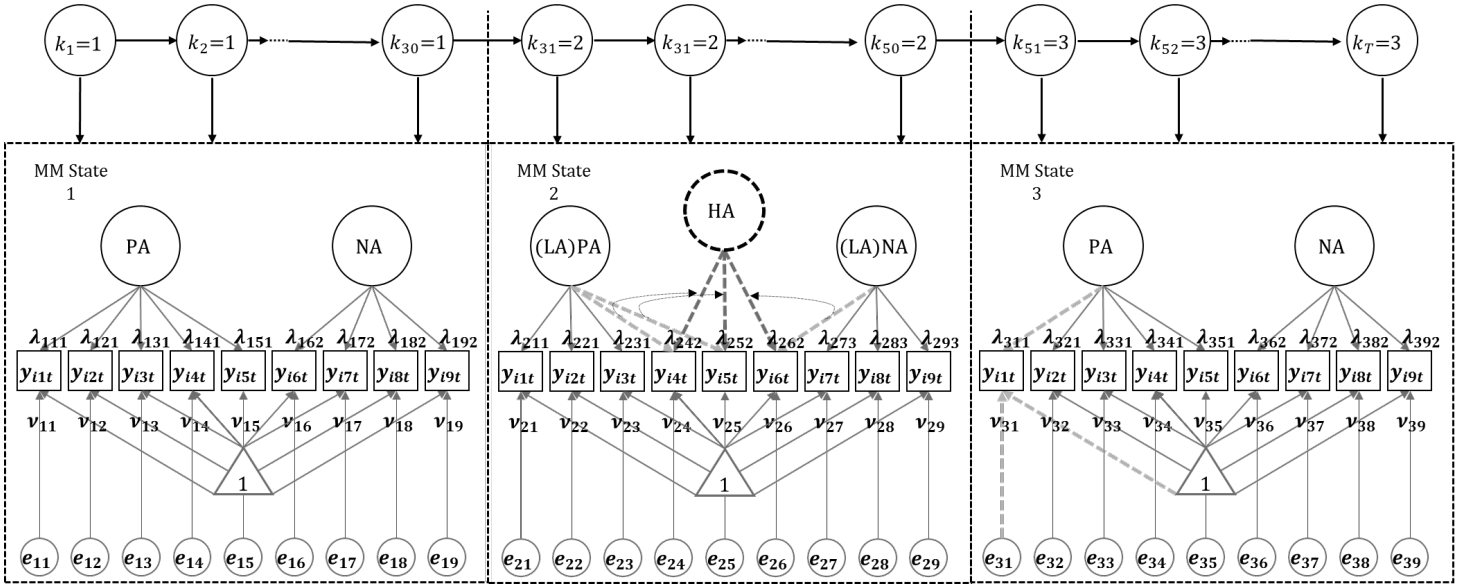


Figure 2.1. Graphical illustration of a subject-specific LMC from a LMFA model, where the latent states per measurement occasion k_t ($t = 1, \ldots, T$) indicate the measurement model underlying a respondent's observed item scores. Note that to give a clear example, only the standardized factor loadings with an absolute value larger than 0.4 are depicted. Also note that the factor scores (e.g., on PA and NA) are not depicted in the figure since they are not part of the MM, but that they may or may not change within a state over time but average zero.

To conclude, in LMFA, the states indicate for which time-points the data are validly comparable (strict MI applies within each state) and by comparing the state-specific MM parameters one may even evaluate which level of invariance holds for which (pairs of) states and which specific MM parameters are non-invariant.

2.3.3 Model Estimation

To estimate the LMFA model we aim to find the model parameters $\boldsymbol{\theta}$ (i.e., the initial state probabilities $\boldsymbol{\pi}$, the transition probabilities $\mathbf{A}_i$, the intercepts $\boldsymbol{\nu}_k$, and the factor-analyzed covariance matrices $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k + \mathbf{D}_k$) that maximize the loglikelihood function log$L$. The log$L$ is derived from Equation 2.1 by summing over all possible state sequences, taking the logarithm, and considering all the subjects at once:

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{i=1}^{I} \log\left(\sum_{\mathbf{s}_{i1}} \dots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}) \prod_{t=2}^{T} p(\mathbf{s}_{it}|\mathbf{s}_{it-1}) \prod_{t=1}^{T} p\left(\mathbf{y}_{it}|\mathbf{s}_{it}\right)\right). \qquad (2.3)$$

Note that the model captures the dependencies only between observations that can be explained by the states but not the autocorrelations of factors within the states. Because the log$L$ is complicated by the latent states, non-linear optimization algorithms are necessary to find the maximum likelihood (ML) solution (e.g., De Roover et al., 2017; Myung, 2003). LMFA can be estimated by means of LG syntax (Vermunt & Magidson, 2016; Appendix B).[5] Specifically, the ML estimation is performed by an expectation maximization (EM; Dempster, Laird, & Rubin, 1977) procedure described in Appendix A. Note that this procedure assumes the number of states K and factors within the states F^k to be known. The most appropriate K and F^k is determined by comparing competing models in terms of their fit-complexity balance. To this end, the Bayesian information criterion (BIC; Schwarz, 1978) can be applied, which proved to be effective for both FA (Lopes & West, 2004) and LMM (Bartolucci, Farcomeni, et al., 2015). Moreover, it may happen that the estimation converges to a local instead of a global maximum. To decrease the probability of finding a local maximum, LG applies a multistart procedure, in which the initial values are automatically chosen based on the loadings and residual variances obtained from PCA (Jolliffe, 1986) on the entire dataset. For each state, randomness is added to get K different sets of initial parameter values (for more details, see De Roover et al., 2017).

[5] A user-friendly graphical interface in LG including a tutorial will be developed in the future.

2.4 Simulation Study

2.4.1 Problem

To evaluate how well LMFA performs in recovering states and state-specific factor models, we manipulated seven factors that affect state-separation and thus potentially the recovery: (a) number of factors, (b) number of states, (c) between-state difference (consisting of differences in factor loadings and intercepts), (d) unique variance, (e) frequency of transitions, (f) number of subjects, and (g) number of observations per subject and state. For the number of factors (a), we expect the performance to be lower for more factors due to the higher model complexity and the lower level of factor overdetermination (given a fixed number of variables) (MacCallum, Widaman, Preacher, & Hong, 2001; MacCallum, Widaman, Zhang, & Hong, 1999; Preacher & MacCallum, 2002). With respect to the number of states (b), a higher number of states also increases the model complexity and thus, probably, decreases the performance. In case of a Markov model, the increase in model complexity with additional states is suppressed by the level of dependency of the states at consecutive time-points, however. Thus, with respect to (e), we anticipate LMFA to performance worse in case of more frequent state transitions, and thus lower probabilities of staying in a state, because this implies a lower dependence on the state of the previous time-point (Carvalho & Lopes, 2007). With respect to (c), we expect a decrease in performance for more similar factor loadings (De Roover et al., 2017) and/or intercepts across states. Regarding (d), LMFA is expected to perform better with a lower unique variance and thus a higher common variance because this increases the factor overdetermination (Briggs & MacCallum, 2003; Ximenez, 2009; Ximénez, 2006). Factors (f) and (g) pertain to the within-state sample size (i.e., the amount of information) per state in terms of number of subjects and observations per subject and state. We expect a higher performance with increasing information (de Winter, Dodou, & Wieringa, 2009; Steinley & Brusco, 2011). Note that we also tested whether lag-one autocorrelations of factors harm the performance of LMFA, which was not the case (Appendix C). In addition, for selected conditions, we evaluated the BIC in terms of the frequency of correct model selection.

2.4.2 Design and Procedure

We crossed seven factors in a complete factorial design with the following conditions:[6]

a. number of factors per state F^k at two levels: 2*, 4*;
b. number of states K at three levels: 2*, 3, 4*;
c. between-state differences at eight levels:

[6] The "*" marks the subset of conditions that is included in the evaluation of model selection.

medium loading difference & no intercept difference,
low loading difference & no intercept difference,
medium loading difference & low intercept difference*,
low loading difference & low intercept difference*,
no loading difference & low intercept difference,
medium loading difference & medium intercept difference*,
low loading difference & medium intercept difference*,
no loading difference & medium intercept difference;

d. unique variance *e* at two levels: .2 and .4*;
e. frequency of state transitions at three levels: highly frequent, frequent, infrequent*;
f. number of subjects *N* at three levels: 2, 5*, 10;
g. number of observations per subject and state T_{ik} at three levels: 50, 100*, 200.

The number of variables *J* was fixed to 20. The numbers of factors per state F^k was either 2 or 4 (a) and was the same across the states. The two, three or four states (b) differed in factor loadings and intercepts. The degree of the between-state loading difference (c) was either medium, low (i.e., highly similar loadings), or non-existent (i.e., identical loadings across states). Between the state-specific intercepts, there was either no difference, a medium difference or a high difference. The combination of no loading difference and no intercept difference was omitted because this implies no difference in MMs and thus only one state. Note that the degree of the between-state differences was the same for each pair of states.

Regarding the factor loadings $\mathbf{\Lambda}_k$ of the generating model, for all conditions, a binary simple structure matrix was used as a common "base" (see Table 2.1). The loading matrices were representative for the ones commonly found in psychological research (cf., the PA and NA structure assumed by the original researchers of the "No Fun No Glory study"). In these matrices, all variables loaded on one factor only and the variables were equally divided over the factors. In case of two factors, this implied that each factor had ten non-zero loadings, whereas, in case of four factors, each factor consisted of five non-zero loadings. For the "no loading difference" conditions, the simple structure base matrix was used as $\mathbf{\Lambda}_k$ in all the states, implying no change in loadings across the states. For the low and medium loading difference conditions, the base matrix was altered differently for each state to create the state-specific loading matrices. Thus, no state will have a factor loading structure equal to the base matrix in Table 2.1. For each state, regardless of the number of factors, we applied the alteration procedure described below.

Table 2.1. Factor base loading matrix and derived loading matrices for state 1 and 2

	Base Loading Matrix		State1		State 2	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Var.1	1	0	λ_1	λ_2	1	0
Var.2	1	0	1	0	λ_1	λ_2
Var.3	1	0	1	0	1	0
Var.4	1	0	1	0	1	0
Var. 5–10	…	…	…	…	…	…
Var.11	0	1	λ_2	λ_1	0	1
Var.12	0	1	0	1	λ_2	λ_1
Var.13	0	1	0	1	0	1
Var.14	0	1	0	1	0	1
Var. 15–20	…	…	…	…	…	…

Note. For the medium loading difference, $\lambda_1 = 0$ and $\lambda_2 = 1$; for the low loading differences, $\lambda_1 = \sqrt{.5}$ and $\lambda_2 = \sqrt{.5}$. Entries of Var. 5–10 and 15–20 equal those of Var.4 and Var.14, respectively. The 4-factor matrices were created by applying the same λ_1 and λ_2 values to other variables because of fewer loadings per factor.

Whether $F^k = 2$ or $F^k = 4$, the manipulations were only applied to the first two factors. Thus, for $F^k = 4$, the third and fourth factor are identical across states. For the medium loading difference conditions, the state-specific loading matrices were created by shifting one loading from the first factor to the second one and one loading from the second factor to the first one. This implies that the overdetermination of the factors is unaffected. For the low loading difference condition, the state-specific loading matrices were created by adding cross-loadings of $\sqrt{.5}$ for two variables, that is, one for factor 1 and one for factor 2, and lowering the primary loading accordingly to $\sqrt{.5}$. This manipulation preserves both the rowwise and columnwise sum of squares (i.e., the variables' common variance and the variance accounted for by each factor). Variables affected by the loading shifts and added cross-loadings differed across states (see Table 2.1).[7]

To quantify the similarity of the state-specific loadings per condition, a congruence coefficient φ (Tucker, 1951) was computed per factor for each pair of the loading matrices.[8] A φ of one indicates proportionally identical factors (as in the no loading difference conditions). The grand mean φ_{mean} across all state pairs and factors amounted to .80 for the medium loading difference conditions and .94 for the low loading difference conditions, regardless of the numbers of states and factors. Finally, the matrices were

[7] Note that we use EFA and, thus, the zero loadings are not fixed but freely estimated in LMFA.

[8] Tucker's (1951) congruence coefficient between column vectors x and y is defined as: $\varphi_{xy} = \frac{x'y}{\sqrt{x'x}\sqrt{y'y}}$

rowwise rescaled such that the sum of squares of each row equaled $1 - e$, where e was either .40 and .20 (g).

Intercept differences were induced as follows. For all variables in all states, the intercept was initially determined to be 5 and kept as such for the no intercept difference conditions. Two of the intercepts (different ones across the states) were increased from 5 to 5.5 for the low intercept difference conditions and from 5 to 7 for the medium intercept difference conditions.

Regarding the frequency of state transitions (e), we manipulated three levels that we considered to be realistic for ESM data. Note that we allowed for between-subject differences in the transition probabilities by randomly sampling each set of subject-specific probabilities from a uniform distribution within a specified range of probabilities. Specifically, the probabilities to stay in a state and to switch to another state were respectively sampled from $U(.73,.77)$ and $U(.01,[\ .27/(K-1)])$ in the highly frequent condition, from $U(.83,.87)$ and $U(.01,[\ .17/(K-1)])$ in the frequent condition and from $U(.93,.97)$ and $U(.01,[\ .07/(K-1)])$ in the infrequent condition. Then, for each resulting matrix, we rescaled the off-diagonal elements of each row to sum to 1 minus the diagonal element of that row, thus maintaining the probabilities to stay in a state and hence also the frequency of switching. As a results, out of the total number of possible transitions (i.e., across subjects (i.e., $\sum_i^I (T_i - 1)$) and across all datasets), a switch to another state occurred for 25% of the possible occasions in the highly frequent condition, for 15% in the frequent condition and for 5% in the infrequent condition.

Depending on the condition, datasets with the above described characteristics were simulated for 2, 5 or 10 subjects (f). Note that limiting our study to such low subject numbers not only confines the computation time but also challenges the method. We expect performance to improve with additional subjects because this accumulates the amount of data within the states. Furthermore, the number of observations per subject and state, T_{ik}, was either 50, 100 or 200 (g) for $i = 1, \ldots, I$ and $k = 1, \ldots, K$. Thus, the total number of observations T_i per subject depended on (b) and (g). Similarly, the within-state sample size per state (over subjects) depended on (f) and (g).

For each subject, a LMC was generated indicating in which state subject i was at each time-point t. The initial state was randomly sampled from a Bernoulli distribution (for $K = 2$) or multinomial distribution (for $K > 2$) with equal initial state probabilities.[9]

[9] We had no intention to evaluate the recovery of the initial state probabilities because more than a few subjects are required to validly estimate the distribution of initial starting states (Vermunt & Magidson, 2016). By sampling the initial state from a Bernoulli/multinomial distribution, some randomness in the initial states was obtained.

The remaining LMC was generated by sampling a random sequence of states based on the subject-specific transition probability matrix (i.e., depending on (e)). Note that whenever a state was not represented in a sampled LMC—as the small sample sizes occasionally led to a data matrix wherein a certain state was not represented—we rejected it and sampled another one so that parameter estimation was possible for all states.

Given this LMC, a subject-specific dataset was generated according to Equation 2.2 assuming orthogonal factors. Firstly, we sampled a factor score vector $\mathbf{f}_{it} \sim MVN(0; \mathbf{I})$ of length F and a residual vector $\mathbf{e}_{it} \sim MVN(0; \mathbf{D}_k)$ of length J for each of the T_i observations, where the diagonal elements of $\mathbf{D}_k$ are equal to .20 or .40 (g). Subsequently, each vector of observations $\mathbf{y}_{it}$ was created with the loading matrix $\mathbf{\Lambda}_k$ and vector of intercepts $\boldsymbol{\nu}_k$ pertaining to the state that subject i was in at time-point t, according to the subject-specific LMC. Finally, the subject-specific datasets $\mathbf{Y}_i$ were concatenated into one data matrix $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_I')'$ with $\sum_i^I T_i$ rows. Twenty data matrices $\mathbf{Y}$ were generated for each cell of the design. In total, 3 (number of states) × 2 (number of factors) × 8 (between-state difference) × 3 (transition frequency between states) × 3 (number of subjects) × 3 (number of observations per subject and state) × 2 (unique variance) × 20 (replicates) = 51840 simulated data matrices were generated. The data was generated in the open-source program R (R Core Team, 2002) and communicated to LG (Vermunt & Magidson, 2016) for analysis. LG syntaxes (for details and an example, see Appendix B) were used to analyze the data with the correct number of states and factors per state. The average time to estimate a model was 85 seconds on an i5 processor with 8GB RAM. Model selection was evaluated for a subset of the conditions (indicated by "*") and for five replications per condition, that is, for 80 datasets. The datasets were analyzed with LMFA models with the number of states equal to $K-1$, K, and $K+1$ states, and the number of factors within the states equal to $F^k - 1$, F^k, and $F^k + 1$ and allowed to differ between the states, resulting in 19 models when $K = 2$ and 46 when $K = 4$.

2.4.3 Results

2.4.3.1 *Sensitivity to Local Maxima*

The estimation procedure, described in Appendix A, may result in a local maximum solution, that is, the best solution may have a log L value that is smaller than the one of the global ML solution. The multistart procedure (described in 2.3.3) increases the chance to find a global ML solution and, in the simulation study—where the global maximum is unknown due to violations of FA assumptions, sampling fluctuations and residuals—we can compare the best solution of the multistart procedure to an approximation (or proxy) of the global ML solution, which we obtain by providing the model estimation with the true parameter values as starting values. A solution is then a local maximum for sure when its log$L$ value is smaller than the one from the proxy. To exclude mere calculation

precision differences, we only considered negative differences with an absolute value larger than .001 as a local maximum. Accordingly, a local maximum was found for 947 out of 51840 simulated datasets (1.83%), which mainly occurred when K = 4.

2.4.3.2 *Goodness of State Recovery*

To investigate the recovery of the state sequence, the Adjusted Rand Index (ARI; Hubert & Arabie, 1985) was computed. The ARI quantifies the overlap between two partitions and is insensitive to permutations of the state labels. It takes values from around 0 when the overlap is at chance level to 1 when partitions are identical. In general, the recovery of states was moderate to good (Steinley, 2004) with a mean ARI-value of .78 (SD = 0.28).

Except for the number of states, all manipulated factors had a large influence on the ARI (Table 2.2). In line with our expectations, the recovery improved with a lower number of factors (b), a greater between-state difference (c), a lower frequency of change (d), a higher number of subjects (e), a higher number of observations per subject and state (f), and lower unique variances (g). Figure 2.2 shows these effects, yet averaged across the number of factors for conciseness. A higher total within-state sample size was especially important for the state recovery in the high unique variance condition when combined with the low and no loading-difference conditions. In contrast, for a low unique variance and a medium loading difference between the states, the state recovery already stabilized at 400 observations. A lower frequency of transitions also further improved the state recovery, but it is most striking that even the most difficult conditions and lowest within-state sample size led to a perfect recovery when there was a medium difference in intercepts between the states.

2.4.3.3 *Goodness of Loading Recovery*

To examine the goodness of state loading recovery ($GOSL$), we computed Tucker congruence coefficients φ between the true loading matrices and the estimated loading matrices and averaged across factors and states:

$$GOSL = \frac{\sum_{k=1}^{K}\sum_{f=1}^{F}\varphi(\boldsymbol{\Lambda}_f^k, \widehat{\boldsymbol{\Lambda}}_f^k)}{\sum_{k=1}^{K} F^K}. \tag{2.4}$$

To deal with the rotational freedom of the factors per state, we rotated the factors prior to calculating the congruence coefficient.[10] Specifically, Procrustes rotation was used to

[10] Note that rotation is not yet included in LG, which is why we took the estimated loadings and rotated them in the open-source program R (R Core Team, 2002).

rotate the estimated towards the true loading matrices. To account for the permutational freedom of the state labels, the state permutation that maximizes the $GOSL$ was retained. The manipulated conditions hardly affected the loading recovery. The overall mean $GOSL$ was .98 (SD = 0.05), indicating an excellent recovery. There was a positive correlation between the ARI value and the $GOSL$ ($r_s = .45, p < 0.001$). Note that the loading recovery was often good despite a bad state recovery because quite similar (to even identical) loading matrices are mixed up.

Table 2.2. Goodness-of-recovery per type of parameter conditional on the manipulated factors

		Goodness of				
		State Recovery (ARI)	Loading Recovery ($GOSL$)	Transition Matrix Recovery (MAD_{trans})	Intercept Recovery (MAD_{int})	Unique Variance Recovery (MAD_{uniq})
Condition	Manipulated Factors					
States	2	0.76	0.97	0.09	0.12	0.04
	3	0.79	0.98	0.08	0.12	0.04
	4	0.80	0.98	0.06	0.12	0.04
Factors	2	0.81	0.98	0.07	0.11	0.03
	4	0.75	0.97	0.09	0.12	0.05
Between-state-difference (Loading Difference-Intercept Difference)	Medium-No	0.69	0.98	0.07	0.09	0.04
	Low-No	0.47	0.94	0.16	0.26	0.08
	Medium-Low	0.81	0.98	0.06	0.08	0.03
	Low-Low	0.68	0.96	0.10	0.16	0.05
	No-Low	0.64	0.96	0.12	0.21	0.06
	Medium-Medium	0.99	1.00	0.04	0.04	0.02
	Low-Medium	0.99	0.99	0.04	0.05	0.02
	No-Medium	0.99	0.99	0.04	0.04	0.02
Transitions	Highly Frequent	0.71	0.97	0.10	0.13	0.04
	Frequent	0.77	0.98	0.08	0.12	0.04
	Infrequent	0.87	0.98	0.06	0.10	0.04
Subjects	2	0.70	0.95	0.12	0.23	0.07
	5	0.81	0.98	0.07	0.09	0.03
	10	0.84	1.00	0.05	0.04	0.02
Observations per subject and state	50	0.70	0.95	0.13	0.21	0.07
	100	0.80	0.98	0.07	0.09	0.04
	200	0.84	0.99	0.04	0.05	0.02
Unique Variances	.2	0.88	0.99	0.05	0.07	0.02
	.4	0.69	0.96	0.11	0.16	0.06

Note. For the between-state difference condition, the combination of no loading difference and no intercept difference was not included because this would imply that the MM does not differ across states.

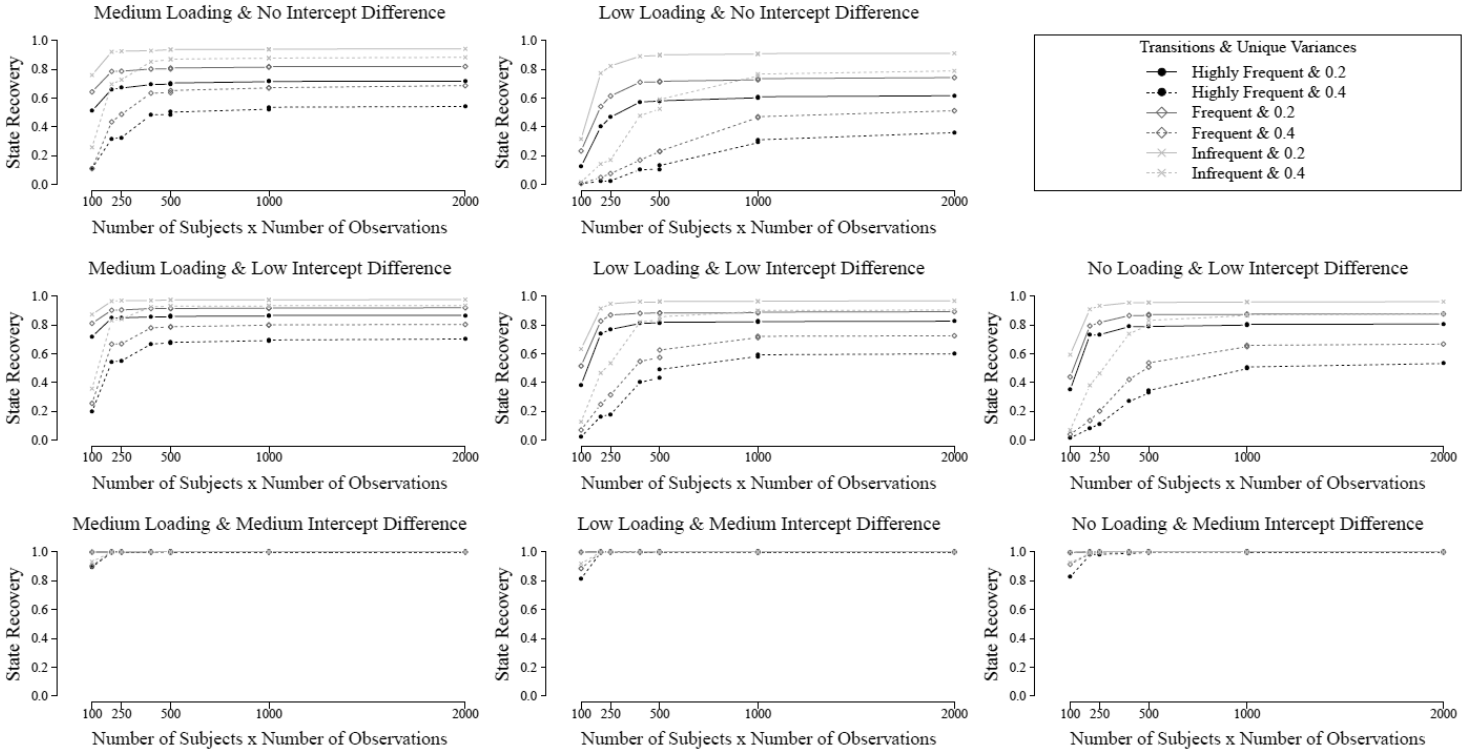


Figure 2.2. Goodness of state recovery (y-axis) under various manipulations of the between-state difference (loading difference (LD) per column and intercept difference (ID) per row), the unique variances and the frequency of transitions (different lines), and the total number of observations per state (number of subjects × number of observations), while averaging across the number of states and factors. Note that the x-axis value 500 and 1000 occur twice due to two possible combinations ($5 \times 100 = 10 \times 50$ and $5 \times 200 = 10 \times 100$, respectively). The dots largely overlap and thus state recovery was not visibly influenced by the combination and thus, it did not matter whether the within-state sample size was increased by additional subjects or additional observations per subject.

2.4.3.4 *Goodness of Transition Matrix Recovery*

To examine the transition matrix recovery, we calculated the mean absolute difference (*MAD)* between the true and estimated matrices (applying the best state permutation obtained from the loading recovery):

$$MAD_{trans} = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{l=1}^{K}|\mathrm{A}_{ikl} - \widehat{\mathrm{A}}_{ikl}|}{IK^2}. \tag{2.5}$$

The transition matrix recovery was good with an average MAD_{trans} of .08 (SD = 0.10). Overall, the effects of the manipulated conditions were rather small (see Table 2.2).

2.4.3.5 *Goodness of Intercept Recovery*

To evaluate the recovery of the state-specific intercepts, we calculated the *MAD* between the true intercepts and the estimated ones.

$$MAD_{int} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{J}|\nu_{kj} - \hat{\nu}_{kj}|}{KJ}. \tag{2.6}$$

The intercept recovery was moderate with an average MAD_{int} of .12 (SD = 0.02). A higher between-state difference of loadings and intercepts (c), more subjects (e), more observations per subject and state (f), and a lower unique variance (g) improved the intercept recovery.

2.4.3.6 *Goodness of Unique Variance Recovery*

To examine the recovery of the state-specific unique variances, we calculated the *MAD* between the true and estimated unique variances.

$$MAD_{uniq} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{J}|d_{kj} - \hat{d}_{kj}|}{KJ}. \tag{2.7}$$

The unique variance recovery was good with an average MAD_{uniq} of .04 (SD = 0.06) and no notable differences across the manipulated conditions. More prominently, the MAD_{uniq} is affected by Heywood cases (Van Driel, 1978), which pertain to "improper" factor solutions with at least one unique variance that is negative or zero. When a Heywood case occurs, LG fixed the unique variance(s) to a very small number. A Heywood case is considered to be diagnostic of underdetermined factors and/or insufficient

sample size (McDonald & Krane, 1979; Rindskopf, 1984a; Van Driel, 1978; Velicer & Fava, 1998). Heywood cases occurred for 5877 of the estimated data matrices (12.19%), where 89% of the Heywood cases indeed occurred in the conditions with the smallest number of observations per subject and state and/or the smallest number of subjects.

2.4.3.7 *Model Selection*

For 74 out of the 80 datasets (93%), the correct model was selected, when considering the converged models only, and for 78 (98%) the correct model was among the three best models. Five of the six incorrect selections occurred for the datasets with four states and four factors and low loading differences as well as low intercept differences. Specifically, one state too few was selected which is explained by the low state separability in these conditions.[11] We conclude that the BIC performs very well with regard to selecting the most appropriate model complexity for LMFA.

2.4.4 Conclusions & Recommendations

To sum up, LMFA is promising for detecting MM changes over time and for exploring what the MM differences look like and for which subjects and which time-points the MMs are comparable. However, the performance of the new method in recovering the correct state sequence and the correct state-specific MMs largely depends on model characteristics (i.e., the number of factors, the MM differences between the states, the unique variances, and the frequency of state transitions), and the within-state sample size. Firstly, larger MM differences between states benefit the recovery of the states. Especially intercept differences increased the separability of the states, to the extent that the states were recovered perfectly even under difficult conditions. Besides intercept differences, less factors, less frequent transitions between the states and lower unique variances improved the recovery. Finally, all else equal, the within-state sample size greatly enhanced the state recovery. In the following we list recommendations for empirical practice:

- When the above-mentioned model characteristics are unknown (or assumed to be unfavorable), aim for 2000 to 4000 observations in total (subjects × observations) to obtain reliable results for 2 to 4 states.
- When favorable model characteristics are assumed—i.e., when between-state differences are expected to be pronounced (e.g., changes in intercept are

[11] Note that, in case there is only one MM underlying the data, model selection would indicate that the one-state model fits best, which was confirmed by a small simulation study with the same design as the one of the model selection described in 3.2.. The correct one-state model was correctly chosen for all 10 models (100 %).

expected), transitions to be infrequent (e.g., measurement occasions are closely spaced) and unique variances to be low (e.g., using reliable measurement instruments)—800 to 1600 observations in total (subjects × observations) are probably enough to obtain reliable results for 2 to 4 states.

- The number of states that can be reliably captured is bound by the total sample size and, when the sample size does not allow for the "true" number of states to be estimated, the obtained results will only convey part of the MM differences present in the data. States that correspond to a few observations only will not be detected.

- Including more subjects in the study might be more feasible than obtaining more measurements from each participants, but be aware that, in practice, subjects do not necessarily switch between the same set of MMs. LMFA allows for this heterogeneity in MMs, since not all subjects need to go through all states. Nevertheless, it is important to keep this potential heterogeneity in mind since it would imply that additional subjects increase the number of MMs and thus the number of states. In that case, the number of observations per subject is essential for the sample size per state.

- The BIC is a suitable criterion to decide on the best number of states and factors. However, when differences between the states are subtle, researchers are advised to consider the three best models and choose one based on interpretability and stability.

2.5 Application

In order to apply LMFA to the empirical dataset introduced in Section 2.2.1, we first selected the number of states and factors by comparing the BIC among LMFA models with one up to three states and one up to three factors per state. Models with four states or factors in a state failed to converge suggesting sample size limitations or model misspecification. The model [3 3 2] (i.e., three states with 3, 3, and 2 factors) was selected as it had the lowest BIC among the converged models and was the most interpretable. To shed light on the MM differences between the three states, we first looked at the state-specific intercepts (Figure 2.3). The intercepts are higher for "positive affect" (PA) items than for "negative affect" (NA) items in all the states. However, the difference between the PA and NA item scores is most visible in State 3 (hereinafter "pleasure state"), intermediate in State 2 (hereinafter "neutral state") and least pronounced in State 1 (hereinafter "displeasure state").

Table 2.3. Standardized oblimin rotated factor loadings, unique variances, and proportions of unique variance of the LMFA model with three states and respectively three, three and two factors for the evening emotion questionnaires

	State 1 (Displeasure)					State2 (Neutral)					State 3 (Pleasure)			
	Factors			Unique V.		Factors			Unique V.		Factors		Unique V.	
Items	PA	Distr.	Drive	V	(prop)	PA	Distr.	Serenity	V.	(prop)	PA	NA	V.	(prop)
interested	**0.66**	0.08	-0.13	178.3	(.48)	**0.70**	0.04	0	166.3	(.52)	**0.85**	0.05	52.6	(.32)
joyful	**0.80**	-0.01	-0.07	118.5	(.30)	**0.86**	0	-0.07	63.4	(.26)	**0.92**	0	25.6	(.16)
determined	0.39	0.01	**-0.48**	187.3	(.43)	**0.81**	0.10	0.16	100.6	(.36)	**0.93**	0.04	29.2	(.18)
calm	**0.43**	**-0.43**	0.10	269.4	(.56)	**0.49**	-0.12	**-0.61**	123.3	(.35)	**0.82**	-0.08	47.2	(.25)
lively	**0.73**	0	-0.13	123.4	(.35)	**0.86**	-0.01	-0.07	65.4	(.26)	**0.90**	-0.01	27.6	(.17)
enthusiastic	**0.81**	0.14	-0.11	129.1	(.29)	**0.87**	0.08	-0.01	91.0	(.27)	**0.93**	0.03	24.0	(.16)
relaxed	**0.62**	-0.34	0.22	207.6	(.48)	**0.54**	-0.14	**-0.67**	65.9	(.21)	**0.85**	-0.07	36.9	(.22)
cheerful	**0.84**	0.08	0.04	133.7	(.35)	**0.84**	-0.05	-0.08	69.3	(.27)	**0.87**	-0.02	39.2	(.23)
content	**0.56**	-0.20	-0.09	179.0	(.51)	**0.84**	0	-0.18	65.3	(.26)	**0.93**	0.01	24.2	(.14)
energetic	**0.59**	0.15	-0.31	183.4	(.41)	**0.78**	0.13	0.03	140.8	(.41)	**0.90**	0.01	37.3	(.21)
upset	-0.18	**0.51**	0.12	290.7	(.58)	0.12	**0.87**	0	13.0	(.27)	0.01	**0.89**	13.2	(.21)
gloomy	-0.33	0.27	0.25	301.8	(.59)	-0.03	**0.77**	-0.03	33.2	(.40)	-0.02	**0.91**	11.5	(.16)
sluggish	0.12	0.13	**0.92**	134.8	(.19)	-0.30	0.26	-0.32	268.5	(.71)	-0.02	**0.85**	22.4	(.27)
anxious	0.07	**0.76**	0.17	183.7	(.39)	0.12	**0.86**	0.03	15.0	(.28)	0.05	**0.91**	14.6	(.19)
bored	-0.09	-0.01	**0.40**	411.1	(.79)	-0.16	0.38	-0.22	163.5	(.74)	-0.01	**0.87**	19.5	(.23)
irritated	-0.14	0.25	0.15	460.3	(.84)	0.05	**0.63**	0.08	92.9	(.61)	-0.01	**0.90**	13.4	(.18)
nervous	0.15	**0.68**	-0.03	326.0	(.58)	0.10	**0.74**	0.11	42.2	(.48)	-0.01	**0.88**	17.4	(.22)
listless	0.18	0.09	**0.58**	304.2	(.48)	-0.30	0.33	-0.22	216.9	(.72)	0	**0.94**	9.0	(.11)

Note. V. = Variance; prop. = proportion; To enhance interpretation, factor loadings were standardized by dividing the loadings by the state-specific standard deviations of the item. The loadings with an absolute value larger than 0.4 are depicted in boldface.

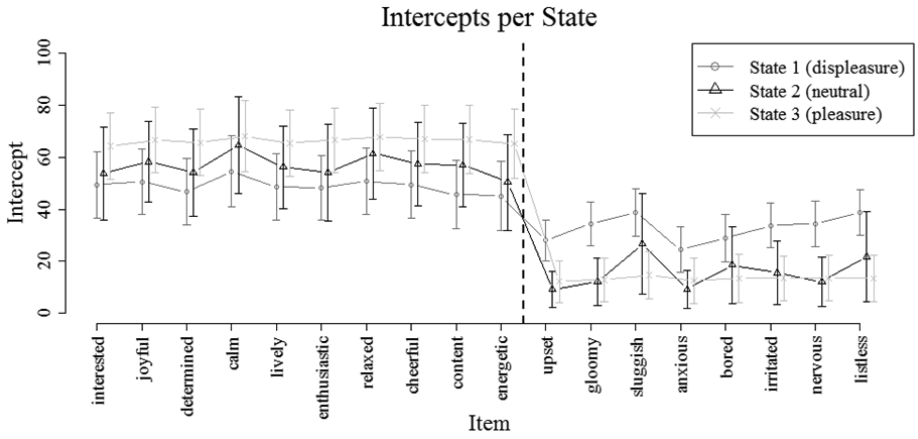


Figure 2.3. Intercepts and standard deviations of the 18 items per state (positive (left) and lower negative emotions (right)).

Second, we investigated the standardized oblimin rotated loadings (Table 2.3). As a notable similarity, we see that the positive items are collected into (i.e., load strongly on) a PA factor in all the states, although the strength of the loadings slightly differs between the states. A striking difference is that the pleasure state has a NA factor, whereas the neutral and displeasure states both have a "distress" factor with high loadings of "upset", "anxious", and "nervous"—although they slightly differ in that "calm" has an additional strong negative loading in the displeasure state, whereas "gloomy" and "sluggish" load on the distress factor in the neutral state only. The neutral and displeasure states are further characterized by a third factor. In the neutral state, the third factor pertains to "serenity" with strong loadings of "calm" and "relaxed". In the displeasure state, it is a bipolar "drive" factor indicating that being "determined" (strong negative loading) is inversely related to feeling "sluggish", "bored", and "listless" (strong positive loadings). This additional drive factor in the displeasure state concurs with theoretical models of anhedonia (Berridge, Robinson, & Aldridge, 2009; Treadway & Zald, 2011), which divide anhedonia in three dimensions: consummatory anhedonia (no longer enjoying pleasurable activities), anticipatory anhedonia (no longer looking forward to pleasurable activities) and motivational anhedonia (no longer experiencing motivation to pursue pleasurable activities). The drive factor confirms that motivation is distinct from general positive affect when individuals are anhedonic. Finally, the state-specific unique variances are listed in Table 2.3. In general, these are highest in the displeasure state, indicating more emotion-specific variability than in the other states. This concurs with research showing that emotional complexity is associated with higher levels of depression (e.g., Grühn, Lumley, Diehl, & Labouvie-Vief, 2013). In sum, LMFA allowed us to find substantively meaningful changes in the MM, both in the number and nature of the

underlying factors. As an important similarity between the states, it is found that PA is captured in each of the three states.[12]

To investigate what potentially triggers the latent states, we explored between-state differences in evening measures on physical discomfort (such as headache) and the occurrence and importance of positive and negative events. We focus on descriptive statistics only since hypothesis testing for MM differences is beyond the scope of the chapter. A question was, for example, "Think about the most unpleasant event you experienced since the last assessment: how unpleasant was this experience?" The scales ranged from 0 ("Not at all") to 100 ("Very much"). Interestingly, when participants were in the displeasure state, they had experienced more unpleasant (M = 48.64, SD = 24.24) events than when being in the neutral (M = 32.54, SD = 19.85) or the pleasure state (M = 29.52, SD = 18.65). Similarly, being in the pleasure state was related to the occurrence of more pleasant events (M = 64.54, SD = 15.48) in comparison to the displeasure (M = 56.02, SD = 20.41) and the neutral state (M = 58.95, SD = 18.03). These findings are in line with the states' labels.

Moreover, we inspected how the states related to the interventions (Table 2.4). Before the intervention (Phase 1), participants were more often in the displeasure or neutral state than in the pleasure state. After the first intervention (Phase 2), the participants in the two intervention groups (i.e., PLA and PLA & SkyD) were more often in the pleasure or neutral state than in the displeasure state. After receiving a second intervention (Phase 3), the distribution across the displeasure and pleasure state stayed about the same or the occurrence of the pleasure state increased. Participants who did not receive an intervention after the first month were distributed equally across the pleasure and displeasure states and were mostly in the neutral state during Phase 2. Notably, in Phase 3, the state membership for these participants—i.e., after receiving their first (self-chosen) intervention—changed in that the pleasure state was now also more frequent than the displeasure state when participants chose PLA & SkyD, while it was the opposite for those who chose PLA, perhaps because the more depressed and anhedonic participants were the ones refraining from a SkyD. Looking at the examples of individual transition plots including the individual transition probabilities (Figure 2.4), it is apparent that participants switched quite often between states, in each phase of the study. This is coherent with previous findings that individuals with anhedonia and depression are often found to experience strong fluctuations in emotional experiences (Heininga, Van Roekel, Ahles, Oldehinkel, & Mezulis, 2017; van Roekel et al., 2015). Some

[12] Note that the NA items were mostly right skewed, warranting caution when drawing substantive conclusions because violations of the normality assumption have yet to be investigated (see Section 2.6). For the purpose of illustrating the aim of LMFA, this is not a problem.

participants switched more often between the states than others, which may pertain to between-subject differences in general stability and experienced events but also to differences in how one reacts to the interventions.

Summarized, the interventions appear to have increased the pleasure state membership and reduced the displeasure state membership, while leaving membership of the neutral state largely unaffected. It is noteworthy that participants receiving no intervention after the first month also slightly moved toward a higher pleasure state membership and a lower displeasure state membership at that point in time. It appears that daily reflections on ones emotions also relieve anhedonia to a certain degree, which was already found in an intervention study using ESM in depressed patients (Kramer et al., 2014). Although these findings are merely exploratory and need to be validated in future research, we demonstrated that LMFA offers valuable insights to applied researchers.

2.6 Discussion

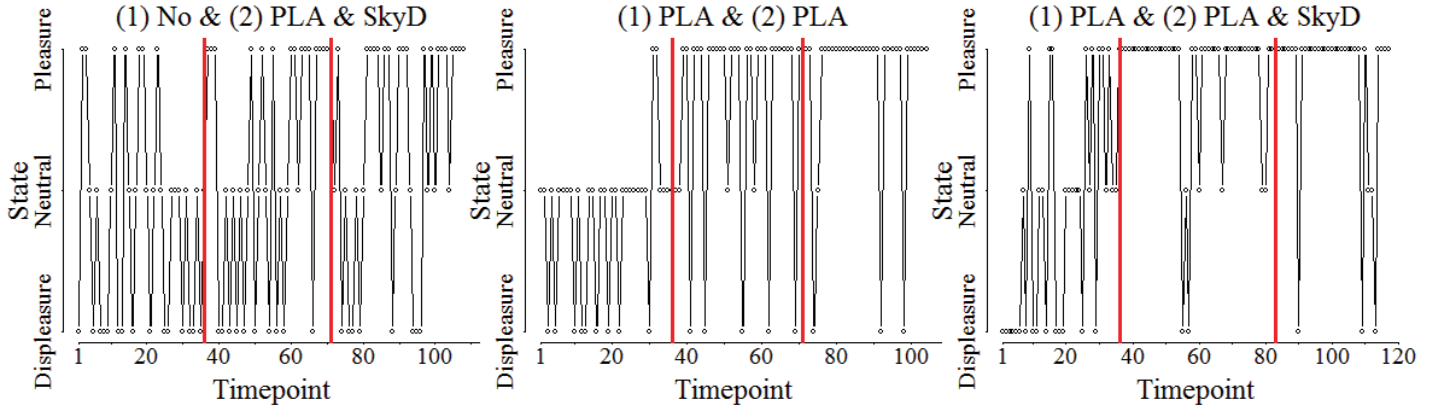
In this chapter, we introduced latent Markov factor analysis (LMFA) for modeling measurement model (MM) changes that are expected to be prevalent in time-intensive longitudinal data like experience sampling data. In this way, LMFA safeguards conclusions about changes in the measured constructs. MM changes may pertain to (potentially interesting) substantive changes or may signal the onset of response styles (RSs). Between-state differences in intercepts and loadings might suggest an extreme RS, whereas differences in intercepts only rather suggest an agreeing RS (Cheung & Rensvold, 2000). When one suspects a RS in a specific state, RS detection and correction (e.g., adding an agreeing RS factor; Billiet & McClendon, 2000; Watson, 1992) can be applied to that specific part of the data, rather than to the entire dataset. Moreover, the subject-specific transition probabilities of LMFA capture, for instance, to what extent each subject is likely to end up and to stay in an extreme RS state. Even when RSs are hard to distinguish, the fact that LMFA pinpoints MM changes—and thus the reliable and comparable parts of the data—is valuable in itself.

In the future, it would be interesting to go beyond the purely exploratory approach applied in this chapter. On the one hand, hypothesis testing to determine which parameters significantly differ between the states might be preferred over visually comparing the state-specific MMs. To this end, LG already provides the researchers with

Table 2.4. Observed state memberships per phase and intervention

	State			State			State		
	1 (Displ.)	**2 (Neutral)**	**3 (Pleasure)**	**1 (Displ.)**	**2 (Neutral)**	**3 (Pleasure)**	**1 (Displ.)**	**2 (Neutral)**	**3 (Pleasure)**
	Int. 1 = No Int. (n = 22, T_B = 790 , T_A = 786)			**Int. 1 = PLA** (n = 23, T_B = 820 , T_A = 832)			**Int. 1 = PLA & SkyD** (n = 24, T_B = 850, T_A = 939)		
Before Int. 1	41.7 %	43.0 %	15.3 %	45.0 %	36.3 %	18.7 %	39.3 %	38.1 %	22.6 %
After Int. 1	27.2 %	45.8 %	27.0 %	27.9 %	30.4 %	41.7 %	27.8 %	40.9 %	31.3 %
	Int. after Int. 1 = No Int.			**Int. after Int. 1 = PLA**			**Int. after Int. 1 = PLA & SkyD**		
	Int. 2 = No Int. (n = 0, T = 0)			**Int. 2 = No Int.** (n = 2, T = 55)			**Int. 2 = No Int.** (n = 1, T = 26)		
	/	/	/	30.9 %	16.4 %	52.7 %	26.9 %	26.9 %	46.2 %
After Int. 2	**Int. 2 = PLA** (n = 3, T = 95)			**Int. 2 = PLA** (n = 3, T = 102)			**Int. 2 = PLA** (n = 11, T = 372)		
	51.6 %	37.9 %	10.5 %	8.8 %	22.6 %	68.6 %	26.1 %	41.4 %	32.5 %
	Int. 2 = PLA & SkyD (n = 19, T = 645)			**Int. 2 = PLA & SkyD** (n = 18, T = 631)			**Int. 2 = PLA & SkyD** (n = 12, T = 430)		
	18.5 %	45.1 %	36.4 %	32.5 %	31.7 %	35.8 %	27.2 %	35.4 %	37.4 %

Note. Int. = Intervention, PLA= positive lifestyle advice, SkyD= skydiving. T_B = T "Before Int. 1", T_A = T "After Int. 1". The state distributions of the observations in the first two phases ("Before Int. 1" and "After Int. 1") are presented per intervention type (3 subgroups). The distributions of the observations in the third phase are presented per combination of the first and second intervention (9 subgroups); e.g., the 22 participants who had no first intervention ("Int. 1 = No Int.") are divided across "Int. 2 = No Int.", "Int 2. = PLA", "Int. 2 = PLA & SkyD".



$$\mathbf{A_1} = \begin{pmatrix} .32 & .51 & .17 \\ .41 & .06 & .53 \\ .22 & .28 & .50 \end{pmatrix} \quad \mathbf{A_2} = \begin{pmatrix} .06 & .49 & .45 \\ .29 & .57 & .15 \\ .15 & .06 & .80 \end{pmatrix} \quad \mathbf{A_3} = \begin{pmatrix} .38 & .21 & .41 \\ .26 & .42 & .32 \\ .08 & .09 & .83 \end{pmatrix}$$

Figure 2.4. Three examples of individual transition plots (including probabilities) of subjects with three different combinations of interventions.

Wald-test statistics when the rotational freedom of the state-specific factors is resolved by a minimal number of restricted loadings (e.g., Geminiani, Ceulemans, & De Roover, 2020). On the other hand, including explanatory variables (i.e., time-constant or time-varying covariates, such as personality traits or social contexts) in the model would allow to evaluate whether they significantly predict the state memberships and the transition probabilities.

Moreover, LMFA assumes normally distributed, continuous variables. Categorical Likert-scale ratings are frequently used in psychology, however. Although these data can often be treated as continuous in case of at least five response categories (Dolan, 1994; Olsson, 1979), the ratings are often skewed, thus violating the assumption of normality. Additionally, even continuous data, such as our application data, might be skewed. Therefore, the robustness of the method to such violations should be examined and, if necessary, possible extensions to deal with non-normality should be considered.

In addition, longitudinal data are often collected in varying time intervals; e.g., when testing the long-term influence of interventions on affect by collecting data in waves. In that case, the transition probabilities can no longer be considered time-homogeneous and continuous-time modeling is necessary (Crayen, Eid, Lischetzke, & Vermunt, 2017). Therefore, in future research, we will develop a continuous-time extension of LMFA.

Moreover, a limitation of the method is the assumption that factor scores are centered around zero and have a variance of one. When factor scores evolve over time but the MM stays the same, changes in factor scores would currently be detected as intercept changes and thus possibly lead to different states according to model selection. In future work, we will investigate necessary LMFA extensions to explicitly model changes in factor means within the states, for example, depending on time or another covariate.

Next to that, we might consider an extension of LMFA using exploratory dynamic factor analysis (EDFA; Browne, 2001; Zhang, 2006) within the states, which models the auto- and cross-lagged correlations of the factors at consecutive time-points but comes with important challenges. Firstly, accurately estimating autocorrelations would require more measurement occasions per subject per state (Ram, Brose, & Molenaar, 2012), which may be undesirable. Secondly, in EDFA, factor rotation is more intricate since the auto- and cross-lagged relations between factors need to be rotated towards specified target matrices (Browne, 2001; Zhang, 2006), again necessitating the a priori hypotheses about (changes in) the MM we want to avoid. The LMM in LMFA already partly captures autocorrelations by the states and uncaptured auto- and cross-lagged correlations will not necessarily introduce bias in the estimates of the state-specific MMs (Baltagi, 2011).

Finally, LMFA is a complex model with many assumptions. Therefore, misspecifications can occur, and tools to locate local misfit are essential. Local fit measures have been developed for related methods (e.g., bivariate residuals measures for multilevel data; Nagelkerke, Oberski, & Vermunt, 2016), but they need to be tailored and extensively evaluated for LMFA.

3

Continuous-Time Latent Markov Factor Analysis for Unequally-Spaced Observations

This chapter is published as Vogelsmeier, L. V. D. E., Vermunt, J. K., Böing-Messing, F., & De Roover, K. (2019). Continuous-time latent Markov factor analysis for exploring measurement model changes across time. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *15*(Suppl 1), 29–42. doi:10.1027/1614-2241/a000176

Abstract

Drawing valid inferences about daily or long-term dynamics of psychological constructs (e.g., depression) requires the measurement model (indicating which constructs are measured by which items) to be invariant within persons over time. However, it might be affected by time- or situation-specific artefacts (e.g., response styles) or substantive changes in item interpretation. To efficiently evaluate longitudinal measurement invariance, and violations thereof, we proposed Latent Markov factor analysis (LMFA), which clusters observations based on their measurement model into separate states, indicating which measures are validly comparable. LMFA is, however, tailored to "discrete-time" data where measurement intervals are equal, which is often not the case in longitudinal data. In this chapter, we extend LMFA to accommodate unequally-spaced intervals. The so-called "continuous-time" (CT) approach considers the measurements as snapshots of continuously evolving processes. A simulation study compares CT-LMFA parameter estimation to its discrete-time counterpart and a depression data application shows the advantages of CT-LMFA.

3.1 Introduction

Longitudinal studies are important to investigate dynamics of latent (i.e., unobservable) psychological constructs (e.g., how depression evolves during or after a therapy). The study design may be, for instance, a traditional daily or weekly diary study or modern experience sampling methodology (ESM; e.g., Scollon et al., 2003), in which subjects may rate questionnaire items say three times a day at randomized time-points over a course of several weeks. Regardless of the design, a measurement model (MM), obtained by factor analysis (FA), indicates to what extent the latent constructs (or "factors") are measured by which items, as indicated by the values of "factor loadings". In order to draw valid inferences about the measured constructs, it is crucial that the MM is invariant (i.e., equal) across time because only then constructs are conceptually similar. However, this longitudinal measurement invariance (MI) is often not tenable because artefacts such as response styles (e.g., an agreeing response style leads to higher loadings; Cheung & Rensvold, 2000), substantive changes in either item interpretation or the number and nature of the measured constructs (e.g., high and low arousal factors replace positive and negative affect factors; Feldman, 1995) may affect the MM differently over time. A confirmatory testing approach is often too restrictive because researchers often have no or incomplete a priori hypotheses about such discrete MM changes. Therefore, Vogelsmeier, Vermunt, van Roekel, and De Roover (2019) proposed latent Markov factor analysis (LMFA), which is an exploratory method that clusters observations of multiple subjects into a few latent states depending on the underlying MM where each state gathers validly comparable observations, as will be described in detail in Section 3.2.2.1.

However, an important aspect of longitudinal data neglected in LMFA so far is that the time lags between two adjacent measurement occasions may vary between and within subjects. For traditional diary studies, the intervals may differ, for instance, because intervals during therapy are shorter (e.g., a day or a week) than follow up intervals after therapy (e.g., six months). Intervals in ESM studies may be unequal because of the "signal-contingent" sampling scheme, which is the most widely used scheme to determine when and how often the participants are questioned (de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). That is, random beeps request the participants to fill in questionnaires with the aim to reduce memory bias and predictability of the measurements. Additionally, night intervals are usually longer than the intervals during the day and, in any study design, participants may skip measurement occasions so that the interval becomes longer.

To accommodate unequally spaced measurement intervals, we extend LMFA in this chapter, following the trend of various modeling approaches to move away from the so-called "discrete-time" (DT) modeling approach that assumes equal intervals and

instead adopt a "continuous-time" (CT) approach that allows for unequal time intervals (TIs). The CT approach fits the idea that we only capture snapshots of the studied process (e.g., because the limitation of observing the entire process) but that processes evolve continually and not only at discrete measurement occasions (Böckenholt, 2005; Crayen et al., 2017; de Haan-Rietdijk et al., 2017; Voelkle & Oud, 2013). Furthermore, in contrast to results from DT studies, where parameters are estimated for a specific interval, results obtained from CT studies are comparable across studies because they are transferable to any interval of interest (de Haan-Rietdijk et al., 2017; Voelkle & Oud, 2013). Moreover, analyzing data containing unequal intervals with DT methods possibly leads to wrong conclusions when not accounting for the exact elapsed time (Driver, Oud, & Voelkle, 2017; Voelkle & Oud, 2013).

The chapter is organized as follows: Section 3.2 describes the data structure, the differences between CT- and DT-LMFA, how the DT approach may approximate CT, and the general model estimation. Section 3.3 presents a simulation study comparing the performance of CT- and DT-LMFA. Section 3.4 illustrated CT-LMFA with an application. Section 3.5 discusses how CT-LMFA safeguards further analyses of factor mean changes when MI cannot be established (e.g., by means of continuous-time structural equation modelling; *ctsem*; Driver et al., 2017) and finally ends with future research plans.

3.2 Method

3.2.1 Data Structure

The repeated measures observations (with multiple continuous variables), nested within subjects are denoted by y_{ijt} (with $i = 1, \dots, I$ referring to subjects, $j = 1, \dots, J$ referring to items, and $t = 1, \dots, T$ to time-points) and are collected in the $J \times 1$ vectors $\mathbf{y}_{it} = (y_{i1t}, y_{i2t}, \dots, y_{iJt})'$, which themselves are collected in the $T \times J$ data matrix $\mathbf{Y}_i = (\mathbf{y}_{i1}', \mathbf{y}_{i2}', \dots, \mathbf{y}_{iT}')'$ for subject i. Note that T may differ across subjects but, for simplicity, we omit the index i in T_i.

3.2.2 LMFA

We first give the building blocks of the regular DT-LMFA (3.2.2.1) and then present CT-LMFA (3.2.2.2).

3.2.2.1 *DT-LMFA*

The first building block of LMFA is a latent Markov model (LMM; Bartolucci et al., 2014; Collins & Lanza, 2010), which is a latent class model that allows subjects to transition between latent classes (referred to as "states"). These transitions are captured by a latent "Markov chain", which follows (a) the "first-order Markov assumption", saying that the

probability of being in state k ($k = 1, \ldots, K$) at time-point t depends only on the previous state at $t - 1$ and (b) the "independence assumption", saying that the responses at time-point t only depend on the state at this time-point. The probability of starting in a state k is given by the initial state $K \times 1$ probability vector $\boldsymbol{\pi}$ with elements $\pi_k = p(s_{1k} = 1)$, where $s_{tk} = 1$ refers to state membership k at time-point t and $\sum_{k=1}^{K} \pi_k = 1$. The probability of being in a state k at time-point t conditional on the state membership l ($l = 1, \ldots, K$) at $t - 1$ is given by the $K \times K$ transition probability matrix $\mathbf{P}$ with elements $p_{lk} = p(s_{tk} = 1 | s_{t-1,l} = 1)$, where the row sums, $\sum_{k=1}^{K} p_{lk}$, are equal to 1. In practice, the transition probabilities depend on the interval length between measurements (e.g., the probabilities to stay in a state are larger if the interval amounts to an hour than when it amounts to a day). Note that typically these probabilities, $\mathbf{P}$, are assumed to be constant over time.

The second building block is the factor analysis (FA; Lawley & Maxwell, 1962) model, which defines the state-specific MMs. The state-specific factor model is

$$\mathbf{y}_{it} = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \mathbf{f}_{it} + \mathbf{e}_{it}, \tag{3.1}$$

with the state-specific $J \times F_k$ loading matrix $\boldsymbol{\Lambda}_k$; the subject-specific $F_k \times 1$ vector of factor scores $\mathbf{f}_{it} \sim MVN(0; \boldsymbol{\Psi}_k)$ at time-point t (where F_k is the state-specific number of factors and $\boldsymbol{\Psi}_k$ the state-specific factor (co-)variances); the state-specific $J \times 1$ intercept vector $\boldsymbol{\nu}_k$; and the subject-specific $J \times 1$ vector of residuals $\mathbf{e}_{it} \sim MVN(0; \mathbf{D}_k)$ at time-point t ,where $\mathbf{D}_k$ contains the unique variances d_{kj} on the diagonal and zeros on the off-diagonal. Note that, for maximum flexibility regarding possible MM differences occurring across persons and time-points, LMFA generally employs an exploratory factor analysis (EFA) approach, that is, without a priori constraints on the factor loadings. If desired, however, confirmatory FA (CFA) could also be used by imposing zero loadings.

From Equation (3.1) it becomes apparent that the state-specific MMs can differ regarding their loadings $\boldsymbol{\Lambda}_k$, intercepts $\boldsymbol{\nu}_k$, unique variances $\mathbf{D}_k$, and factor covariances $\boldsymbol{\Psi}_k$, implying that LMFA explores all levels of measurement non-invariance (described in detail in, e.g., Meredith, 1993): Configural invariance (equal number of factors and zero loading pattern), weak factorial invariance (equal loading values), strong factorial invariance (equal intercepts) and strict factorial invariance (equal unique variances).

To identify the model, factor variances in $\boldsymbol{\Psi}_k$ are restricted to one and rotational freedom is dealt with by means of criteria to optimize simple structure of the factor loadings (e.g., oblimin; Clarkson & Jennrich, 1988), between-state agreement (e.g., generalized Procrustes; Kiers, 1997) or the combination of the two (De Roover &

Vermunt, 2019). The multivariate normal distribution with the state-specific covariance matrices $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k{}' + \mathbf{D}_k$ defines the state-specific response densities $p(\mathbf{y}_{it}|\mathbf{s}_t)$, indicating the likelihood of the J observed item responses at time-point t given the state membership at t.

Summarized, there are three types of probabilities that together make up the joint probability density of subject i's observations and state memberships:

$$p(\mathbf{Y}_i, \mathbf{S}) = \overbrace{p(\mathbf{s}_1)}^{\substack{\text{initial state} \\ \text{probabilities}}} \prod_{t=2}^{T} \overbrace{p(\mathbf{s}_t|\mathbf{s}_{t-1})}^{\substack{\text{transition} \\ \text{probabilities}}} \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_{it}|\mathbf{s}_t)}^{\substack{\text{response} \\ \text{probabilities}}}, \tag{3.2}$$

where $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T)$ is the $K \times T$ state-membership indicator matrix. Here, the columns $\mathbf{s}_t = (s_{t1}, \dots, s_{tK})'$, for $t = 1, \dots, T$, are binary vectors indicating the state memberships at time-point t (e.g., if $K = 3$ and a subject is in state 3 at time-point t, then $\mathbf{s}_t = (0, 0, 1)'$. When applying this model in situations in which measurement intervals are not equal, the encountered transition probabilities will refer to more or less the average interval length in the dataset concerned. For intervals shorter than the average, the transition probabilities yield an overestimation of transitions while for intervals longer than the average, the transition probabilities yield an underestimation.

One solution to account for unequal intervals in the DT approach to a certain extent is to rescale intervals to a finer unit (e.g., 1 hour) and to round the time-points to the nearest unit. So-called "phantom variables" (Driver et al., 2017; Rindskopf, 1984b) containing missing values are inserted for all time-points without observations. Although this is good approximation if the grid is fine enough, for substantive researchers, transforming the dataset is burdensome and choices regarding the interval lengths difficult. Moreover, a high number of iterations of the algorithm described in Section 3.2.3 is required to achieve convergence, causing long computation times (for more information on this, see Appendix E). Therefore, we only consider the CT approach, which is a much more natural alternative to account for the unequal TI.

3.2.2.2 ***CT-LMFA***

The CT approach has been extensively discussed in the literature on Markov models (Cox & Miller, 1965; Kalbfleisch & Lawless, 1985) and latent Markov models (Böckenholt, 2005; Jackson & Sharples, 2002) and overcomes inaccurate estimation by considering the length of time, δ, spent in each of the states. Specifically, transitions from current state l to another state k are here defined by probabilities of transitioning from one state to

another per very small time unit and are called transition intensities or rates q_{lk}. These intensities can be written as:

$$q_{lk} = \lim_{\delta \to 0} \frac{p(s_{tk} = 1 | s_{t-\delta,l} = 1)}{\delta}. \tag{3.3}$$

The $K \times K$ intensity matrix $\mathbf{Q}$ contains the transition intensities q_{lk} for $k \neq l$ as off-diagonal elements and their negative row sums, that is, $-\sum_{k \neq l} q_{lk}$, on the diagonals. For example, for $K = 3$,

$$\mathbf{Q} = \begin{pmatrix} -(q_{12} + q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21} + q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31} + q_{32}) \end{pmatrix}. \tag{3.4}$$

There are three assumptions underlying the CT latent Markov model: (1) the time spent in a state is independent of the time spent in a previous state, (2) the transition intensities q_{lk} are independent of and thus constant across time[13], and (3) the time spent in a state is exponentially distributed (Böckenholt, 2005). The matrix of transition probabilities $\mathbf{P}$ can be computed as the matrix exponential[14] of the intensity matrix $\mathbf{Q}$ times the TI δ (Cox & Miller, 1965):

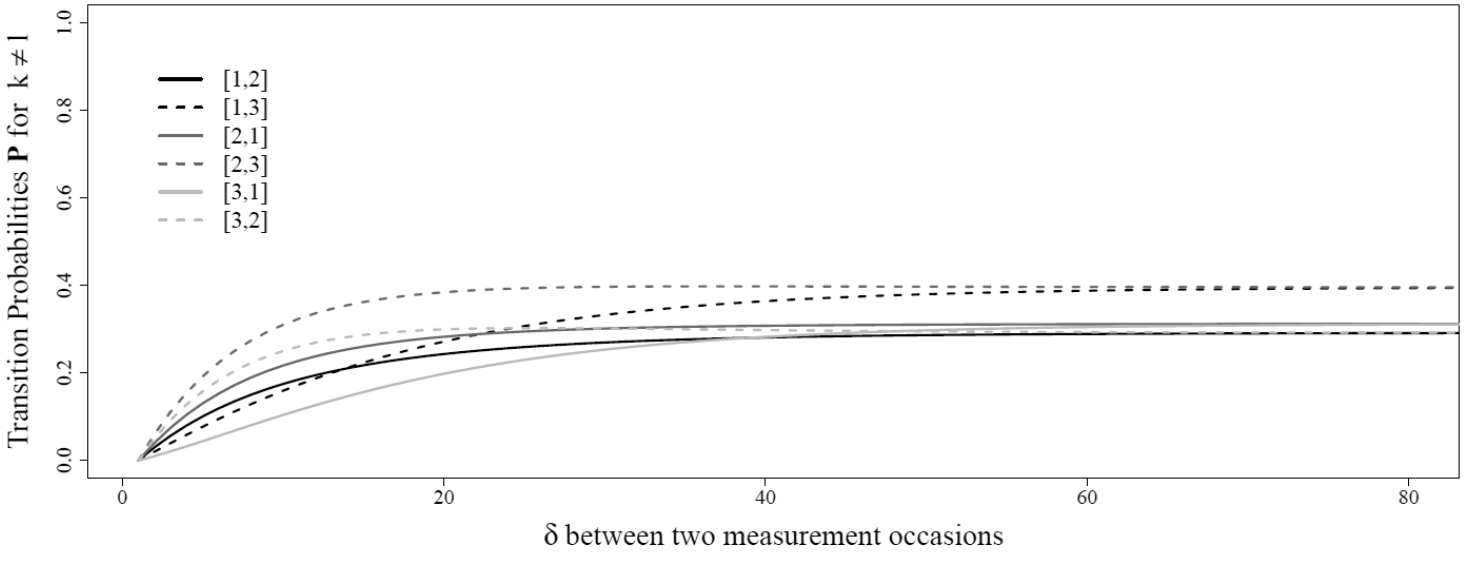
$$\mathbf{P}(\delta) = e^{\mathbf{Q}\delta}. \tag{3.5}$$

Note that the specific structure of $\mathbf{Q}$ (with negative row sums on the diagonal) is a consequence of taking the matrix logarithm of $\mathbf{P}$ with its restriction $\sum_{k=1}^{K} p_{lk} = 1$ (Cox & Miller, 1965). With Equation (3.5), we can compute the transition probabilities for arbitrary TIs, which is, as mentioned in the introduction (Section 3.1), a distinctive advantage of the CT approach. Thus, the probabilities change depending on the interval length between two consecutive observations. How the transition probability matrix $\mathbf{P}$ changes depending on TI δ is shown in Figure 3.1 based on an arbitrary intensity matrix $\mathbf{Q}$.

---

[13] Note that this assumption might be relaxed. For example, one might assume different transition intensities for night and day intervals or that transition intensities change over time. In these cases, one may use covariates or specific model approaches (e.g., a model with a Weibull distribution that models the intensities as a function of time). However, this is beyond the scope of the current chapter.

[14] The matrix exponential $e^{\mathbf{A}}$, where $\mathbf{A}$ can be any square matrix, is equal to $\sum_{a=0}^{\infty} \frac{\mathbf{A}^a}{a!} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{AA}}{2!} + \frac{\mathbf{AAA}}{3!} + \cdots$, where $\mathbf{I}$ is the identity matrix.



Intensities in **Q** = logm(**P** for δ = 1)

	[,1]	[,2]	[,3]
[1,]	−0.052	0.032	0.02
[2,]	0.043	−0.108	0.065
[3,]	0.009	0.054	−0.064

Probabilities in **P** for δ = 1

	[,1]	[,2]	[,3]
[1,]	0.95	0.03	0.02
[2,]	0.04	0.9	0.06
[3,]	0.01	0.05	0.94

Probabilities in **P** for δ = 10

	[,1]	[,2]	[,3]
[1,]	0.64	0.19	0.17
[2,]	0.23	0.45	0.32
[3,]	0.12	0.26	0.62

Probabilities in **P** for δ = 80

	[,1]	[,2]	[,3]
[1,]	0.31	0.29	0.4
[2,]	0.31	0.29	0.4
[3,]	0.31	0.29	0.4

Figure 3.1. Probabilities of transitioning to another state as a function of the time interval δ between two measurement occasions. The transition probabilities increase with δ until they reach a stationary distribution. Three example probability matrices are calculated based on **Q** (left matrix) and $\delta = 1, 10, 80$. Note that "$[l,k]$" indicates the elements in the matrices with l referring to the rows and k to the columns and that the exact **Q** matrix can be obtained by taking the matrix logarithm of **P** for $\delta = 1$.

As a final remark, note that the joint probability density of subject *i*'s observations and state memberships for DT-LMFA in Equation (3.2) also applies to CT-LMFA. The only difference is that the transition probabilities $p(\mathbf{s}_t|\mathbf{s}_{t-1})$ depend on the q_{lk} and the TI δ for subject i at time-point t (with regard to $t-1$) such that $p_{\delta_{ti}}(\mathbf{s}_t|\mathbf{s}_{t-1})$ is a more appropriate notation.

3.2.3 Estimation

Using syntax, Latent GOLD (LG; Vermunt & Magidson, 2016) can be used to find the parameters previously described—collectively referred to as $\boldsymbol{\theta}$—that maximize the log-likelihood function log$L$. Apart from the transition probability formulation in the DT approach, where $p_{\delta_{ti}}(\mathbf{s}_t|\mathbf{s}_{t-1}) = p(\mathbf{s}_t|\mathbf{s}_{t-1})$, the log$L$ formulation is the same for DT-LMFA and CT-LMFA. The log$L$ for both models is given by:

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{i=1}^{I} \log \left(\sum_{\mathbf{s}_{i1}} \dots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}) \prod_{t=2}^{T} p_{\delta_{ti}}(\mathbf{s}_t|\mathbf{s}_{t-1}) \prod_{t=1}^{T} p\left(\mathbf{y}_{it}|\mathbf{s}_{it}\right) \right), \qquad (3.6)$$

which is complicated by the latent states. Therefore, to find the maximum likelihood (ML) solution, LG utilizes the expectation maximization (EM; Dempster et al., 1977) algorithm, more specifically the forward-backward algorithm (Baum, Petrie, Soules, & Weiss, 1970), which is described in detail for DT-LMFA in Vogelsmeier, Vermunt, van Roekel, et al. (2019). Estimation of the CT-LMFA differs in that the maximization-step (M-step) requires using a Fisher algorithm not only for updating the state-specific covariance matrices (Lee & Jennrich, 1979) but also for updating the log transition intensities (Kalbfleisch & Lawless, 1985). A summary is provided in Appendix D. Note that the estimation procedure assumes that we know the number of states *K* and factors within the states F_k. Since these numbers are only known in simulation studies, a model selection procedure is required when working with real data. For LMFA, the Bayesian information criterion (BIC) proved to perform well in terms of selecting the best model complexity (Vogelsmeier, Vermunt, van Roekel, et al., 2019).

3.3 Simulation Study

3.3.1 Problem

We employed an ESM design with unequal TIs—currently the go-to research design to study daily-life dynamics—to evaluate how CT-LMFA and standard DT-LMFA differ in recovering the model parameters. Generally, we expected CT-LMFA to outperform DT-LMFA, although the performance difference might be small (Crayen et al., 2017). We

manipulated three types of conditions that previously were shown to influence MM parameter recovery and state recovery (Vogelsmeier, Vermunt, van Roekel, et al., 2019): (1) factor overdetermination, (2) state similarity and (3) amount of information available for estimation. We expect the differences in MM parameter recovery and state recovery across the two methods to be especially pronounced for (1) a lower factor overdetermination, (2) a lower state similarity, and (3) a lower amount of information because the posterior state probabilities are functions of the observed data and the state memberships at the adjacent time-points (see Appendix D.1). Hence, the estimation benefits from precisely estimated transition probabilities. These precise estimates are likely more important for more "difficult" conditions where the state membership is more difficult to predict based on the observed data.

Based on the simulation study of Vogelsmeier, Vermunt, van Roekel, et al. (2019), the conditions for (1) factor overdetermination were (a) number of factors (where a higher number causes lower factor overdetermination for a fixed number of items; e.g., Preacher & MacCallum, 2002) and (b) unique variances (where lower unique variances increase common variance and therefore also factor overdetermination; e.g., Briggs & MacCallum, 2003). The conditions for (2) state similarity were (c) between-state loading similarity and (d) between-state intercept difference. The conditions for (3) amount of information—with (e) sample size, N, (f) number of days of participation, D, and (g) number of observations per day, T_{day}—were based on a typical ESM design.

Note that T_{day} determines the amount of DT violation (i.e., to what degree the intervals differ from the average day interval) as well as the transition probabilities. A higher T_{day} implies smaller DT violations and fewer transitions to other states at two consecutive observations as will be described in Section 3.3.2. Performance differences regarding the transition parameter recovery are expected to be especially pronounced for a lower T_{day} and thus for higher DT violations and higher transition probabilities to other states, where the latter leads to lower dependence of states at two consecutive time-points, making the estimation more difficult (Vogelsmeier, Vermunt, van Roekel, et al., 2019).

3.3.2 Design and Procedure

We crossed seven factors with the following conditions in a complete factorial design:

a. number of factors per state $F_k = F$ at two levels: 2, 4;
b. unique variance e at two levels: .2, .4;
c. between-state loading difference at two levels: medium loading difference and low loading difference;

d. between-state intercept difference at two levels: no intercept difference, low intercept difference;
e. sample size N at two levels: 35, 75;
f. the number of days D at two levels: 7, 30;
g. the measurements per subject and day T_{day} at three levels: 3, 6, 9;

resulting in $2\ (a) \times 2\ (b) \times 2\ (c) \times 2\ (d) \times 2\ (e) \times 2\ (f) \times 3\ (g) = 192$ conditions. The number of items J was fixed to 20 and the number of states K was fixed to 3.

The loading differences between the states (c) was either medium or low. For both conditions, we started with a common base loading matrix, $\mathbf{\Lambda}_{Base}$, which was a binary simple structure where all items loaded on only one factor and all factors were measured by the same amount of items (i.e., 10 for $F = 2$ and 5 for $F = 4$). To clarify this, consider $\mathbf{\Lambda}_{Base}$ for the example of $F = 2$:

$$\mathbf{\Lambda}_{Base} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \quad (3.7)$$

To induce loading differences between the states, we altered the base matrices differently for each state. Specifically, for the medium between-state loading difference condition, we shifted respectively one loading from the first factor to the second and one from the second to the first for both for $F = 2$ and $F = 4$, so that, for $F = 4$, only the first two factors differed across states. Items for which the loadings were shifted differed across states. This manipulation did not affect the overdeterminaton of the factors, which was therefore the same across states. Thus, for the example of $F = 2$, the loading matrices for the first two (of the three) states were

$$\begin{aligned} \mathbf{\Lambda}_1 &= \begin{pmatrix} \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \\ \mathbf{\Lambda}_2 &= \begin{pmatrix} 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \end{aligned} \quad (3.8)$$

with $\lambda_1 = 0$ and $\lambda_2 = 1$. The low between-state loading difference condition differed from the just described one only in that, instead of shifting loadings, we added one cross-loading of $\sqrt{.5}$ to the first and one to the second factor for different items across states, thereby also lowering the primary loadings to $\sqrt{.5}$. Thus, the entries in $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ in Equation (3.8) were $\lambda_1 = \sqrt{.5}$ and $\lambda_2 = \sqrt{.5}$ for this condition. Finally, we rescaled the loading matrices rowwise so that the sum of squares per row was $1 - e$, where e was either .40 or .20.

To have a measure of between-state loading matrix similarity, we computed the grand mean, φ_{mean}, of Tucker's (1951) congruence coefficient (defined by $\varphi_{xy} = \frac{x'y}{\sqrt{x'x}\sqrt{y'y}}$, where x and y refer to columns of a matrix) across each pair of factors, with $\varphi = 1$ indicating proportionally identical factors. For the medium loading difference condition, φ_{mean} across all states and factors was .8 and for the low loading difference condition .94, regardless of the number of factors.

For creating between state intercept differences (d), we first created a base intercept vector consisting of fixed values of 5:

$$\boldsymbol{\nu}_{Base} = (5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \qquad (3.9)$$

For the no intercept difference condition, we used $\boldsymbol{\nu}_{Base}$ for each state. For the low intercept difference condition, we increased two intercepts to 5.5 for different items across the states. This resulted in the following two intercept vectors for the first and the second state.

$$\begin{aligned} \boldsymbol{\nu}_1 &= (\mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \\ \boldsymbol{\nu}_2 &= (5 \quad 5 \quad \mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \end{aligned} \qquad (3.10)$$

Datasets were generated for either 35 or 75 subjects, N, (e). The number of days, D, for simulated participation was either 7 or 30 (f) and the number of measures per day (h), T_{day}, was 3, 6, or 9. The total number of observations T for one data matrix was therefore, $N \times T_{day} \times D$. Factors (f) and (g) also determined the sampling schedule. The day lasted from 9 am and to 9 pm so that days and nights were on average twelve hours long. Depending on whether T_{day} was 3, 6 or 9, the general intervals between measurement occasions during the day were $\delta_{t_{general}} = 12/(T_{day} - 1)$ and thus 6, 2.4 or 1.5 hours, while the night intervals were not directly affected by T_{day}. To obtain a CT sampling scheme with randomness typical for ESM studies, we allowed for a uniform random deviation around the fixed time-points with a maximum of plus and minus 30 percent of the DT TIs (e.g., for T_{day} = 3, we calculated the product of the general TI and the percentage of violation, 6×0.3, which is 1.8, and therefore, we sample the deviation from the uniform distribution $Unif(-1.8,1.8)$). This explains why the DT violation is bigger for a smaller T_{day}.

Finally, the transition intensities in **Q** were fixed across all conditions, subjects, and time. To determine **Q**, we considered transition probabilities **P** realistic for short TIs

and determined them for the intermediate $T_{day} = 6$ condition and thus for an interval of 2.4 hours. That means, 2.4 hours pertains to one unit and therefore, all other intervals will be scaled to this unit interval. From the chosen probabilities

$$\mathbf{P} = \begin{pmatrix} .950 & .025 & .025 \\ .025 & .950 & .025 \\ .025 & .025 & .950 \end{pmatrix}, \tag{3.11}$$

Q was derived by taking the matrix logarithm:[15]

$$\mathbf{Q} = \begin{pmatrix} -.05 & .03 & .03 \\ .03 & -.05 & .03 \\ .03 & .03 & -.05 \end{pmatrix}. \tag{3.12}$$

Because of the design, the transition probabilities across measurement occasions will be larger for $T_{day} = 3$, where intervals δ_{ti} are longer, and smaller for $T_{day} = 9$, where intervals are shorter.

In the open-source program R (R Core Team, 2020) for each subject, we sampled $T_{day} \times D$ time-points as previously described (see Section 3.3.2). Subsequently, we sampled a random initial state from a multinomial distribution with equal probabilities and, based on the subject-specific TIs, generated a random CT latent Markov chain (LMC) containing state memberships for each subject. According to the LMCs, we generated N data matrices $\mathbf{Y}_i$ with the state-specific factor model of Equation (3.1), assuming orthogonal factors, and concatenated the $\mathbf{Y}_i$'s into one dataset $\mathbf{Y} = \mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_I')'$. In total, 20 replicates of the 192 conditions and thus 3840 datasets were generated.

3.3.3 Results

Performances were evaluated based on 3831 out of 3840 datasets that converged at the first try in both analyses (99.7 % analyses converged in CT-LMFA and all converged in DT-LMFA).[16]

[15] Note that the rows do not sum to zero only because of rounding in this representation.

[16] Note that it may also happen that the estimation results in a locally maximum likelihood (ML) solution, implying that the local ML solution has a smaller log L value than the global ML solution. Note that the latter is unknown but, in the simulation study, an approximation ("proxy") can be obtained by using the population parameters as starting values and comparing the multistart solution to the proxy solution: When $\log L_{multistart} < \log L_{proxi}$ (i.e., by .001 to exclude minor calculation differences), we considered the solution as local maximum. In the converging ML solutions, a local maximum was found for only 0.55 % of the datasets analyzed with CT-LMFA and for 0.47 % of the datasets analyzed with DT-LMFA.

3.3.3.1 *Performance Measures*

First, the state recovery was examined with the Adjusted Rand Index (ARI) between the true and the estimated state MC's. The ARI is insensitive to state label permutations and ranges from around 0 (i.e., overlap is at chance) to 1 (i.e., perfect overlap). Second, to obtain the differences in the goodness of state loading recovery ($GOSL$), we averaged the Tucker congruence coefficient between the true and the estimated loading matrices across factors and states:

$$GOSL = \frac{\sum_{k=1}^{K}\sum_{f=1}^{F}\varphi(\boldsymbol{\Lambda}_k^f, \widehat{\boldsymbol{\Lambda}}_k^f)}{\sum_{k=1}^{K} F_k}. \tag{3.13}$$

We used Procrustes rotation (Kiers, 1997) to rotate state-specific loadings $\widehat{\boldsymbol{\Lambda}}_k^f$ to $\boldsymbol{\Lambda}_k^f$.[17] This solves the label switching of the factors within that state. To account for differences in state labels, we retained the permutation that maximized $\varphi(\boldsymbol{\Lambda}_k^f, \widehat{\boldsymbol{\Lambda}}_k^f)$. Third, for all other parameters (i.e., transition parameters, intercepts, unique variances, and initial state probabilities), we computed the mean absolute difference (MAD) between the true and the estimated parameters.[18] Note that, for the transition and initial state parameters, we considered the state permutation that was found to maximize the loading recovery. Furthermore, the transition parameters are probabilities for DT but intensities for CT. In order to make deviations from the population parameters as comparable as possible, we transformed the intensities from the CT analyses to probabilities for the 1-unit TI of 2.4. Moreover, the "true" parameter in DT-LMFA to evaluate the MAD_{trans} is based on the average population TI.

3.3.3.2 *Goodness of Parameter Recovery*

As can be seen from the "average" results in Table 3.1, CT-LMFA was slightly superior to DT-LMFA regarding the general state and transition probability recovery but still very comparable regarding MM parameter recovery. Moreover, contrary to our expectations, the difference in MM and state recovery across the two analyses were not affected by most of the manipulated conditions, probably because the transition probabilities were overall very well estimated. Only lower levels of T_{day} considerably increased the

[17] We conducted the rotation in R, since factor rotation was just added to LG after the study was conducted.
[18] Note that the MAD_{uniq} may be affected by Heywood cases pertaining to improper factor solutions where at least one unique variance is zero or negative (e.g., Van Driel, 1978). Heywood cases did not occur in any of the analyses and are therefore not further discussed.

performance difference between CT- and DT-LMFA, which was in line with our expectations.

3.3.4 Conclusion and Recommendations

To sum up, there was a striking similarity in recovering parameters under a wide range of conditions across the CT- and DT-LMFA. Nevertheless, it was shown that CT-LMFA leads to the best state recovery and, furthermore, provides researchers with valid transition probabilities for any TI of interest and should therefore be the preferred method. Furthermore, although we demonstrated the robustness of DT-LMFA in recovering the correct MMs for a typical ESM design where the degree of DT violation is rather small, we cannot generalize the findings purporting that DT-LMFA is an adequate substitute for datasets with large DT violations.

3.4 Application

In the following, we apply CT-LMFA to longitudinal data of the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP; Elkin et al., 1989) to evaluate MM changes over time. In brief, the data consisted of repeated depression measures of 122 subjects with a major depression disorder. By means of the 20-item Beck Depression Inventory (BDI; Beck et al., 1979; items listed in Table 3.2), depression was assessed on a 4-point scale before treatment, during treatment (i.e., weekly and additionally after 4, 8 and 12 weeks), at termination, and at follow ups after 6, 12, and 18 months. The total number of observations was 1700 with an average of 14.24 per subject (ranging from 1 to 30). Intervals between the observations varied tremendously from very small (e.g., a day when the weekly and the

Table 3.1. Goodness of recovery per type of parameter and convergence conditional on the manipulated factors

		Goodness of Recovery for											
		States (ARI)		Loadings ($GOSL$)		Transition Parameters (MAD_{trans})		Intercepts (MAD_{int})		Unique Variances (MAD_{uniq})		Initial States ($MAD_{initial}$)	
		Type LMFA											
Condition	Factors	CT	DT	CT	DT	CT	DT	CT	DT	CT	DT	CT	DT
Factors per State F^k	2	.87	.85	1	1	.00	.06	.03	.03	.01	.01	.06	.06
	4	.84	.83	1	1	.00	.06	.03	.03	.01	.01	.06	.06
Between-State Loading Difference	medium	.89	.88	1	1	.00	.06	.02	.03	.01	.01	.06	.06
	low	.82	.81	1	1	.01	.06	.03	.03	.01	.01	.06	.06
Between-State Intercept difference	no	.80	.79	1	1	.01	.06	.03	.03	.01	.01	.06	.06
	low	.90	.90	1	1	.00	.06	.02	.03	.01	.01	.06	.06
Unique Variance e	.2	.91	.90	1	1	.00	.06	.02	.02	.01	.01	.06	.06
	.4	.79	.78	1	1	.01	.06	.03	.03	.02	.02	.06	.06
Sample Size N	35	.85	.84	1	1	.01	.06	.03	.03	.02	.02	.07	.07
	75	.86	.85	1	1	.00	.06	.02	.02	.01	.01	.05	.05
Number of Participation Days D	7	.84	.83	1	1	.01	.06	.03	.03	.02	.02	.06	.06
	30	.86	.85	1	1	.00	.06	.02	.02	.01	.01	.06	.06
Measurements per day T_{day}	3	.75	.74	1	1	.01	.10	.03	.03	.02	.02	.06	.06
	6	.88	.87	1	1	.00	.05	.02	.02	.01	.01	.06	.06
	9	.93	.91	1	1	.00	.03	.02	.02	.01	.01	.06	.06
All Conditions													
Average		.85	.84	1	1	.00	.06	.03	.03	.01	.01	.06	.06
SD		.13	.13	0	0	.01	.03	.01	.01	.01	.01	.03	.03

Note. LMFA = latent Markov factor analysis; CT = continuous-time; DT = discrete-time. The perfect loading recoveries are a consequence of the highly similar loading matrices across the states that have been mixed up.

4-week questionnaire were completed on two consecutive days) to very large (e.g., a year when certain follow ups were skipped).[19]

To begin with the data analysis, model selection with the BIC (comparing converged solutions of one to three states and one to three factors per state) indicated that the best fitting model was a two state model with three factors in the first state and two factors in the second state.[20] Hence, configural invariance is clearly violated. In order to shed light on the state-specific MMs, we investigated the standardized oblimin rotated loadings (Table 3.2). Considering the standardized loadings of higher than .3 in absolute value (e.g., Hair, Anderson, Tatham, & Black, 2014), state 1 is characterized by three factors pertaining to (1) "despair"—with strong loadings of, for example, "mood", "pessimism", and "lack of satisfaction"—, (2) "self-image"—with strong loadings of, for example, "guilty feeling", "self-hate", and "self-accusation"—, and (3) "cognition/body"—with strong loadings of, for example, "irritability", "sleep disturbance", and "fatigability". In state 2, all items beside "sense of punishment", "self-punitive wishes" and "loss of appetite", which all have no variation and thus no loadings at all, mainly load on one factor, therefore pertaining to (1) "depression". Only "indecisiveness" and "work inhibition" have considerable loadings on the other factor as well, which may pertain to (2) "cognition". Moreover, intercepts and unique variances are higher in the first than in the second state.

Next, we look at the estimated transition intensity matrix $\mathbf{Q} = \begin{pmatrix} -.02 & .02 \\ .01 & -.01 \end{pmatrix}$, from which we can calculate $\mathbf{P}$ for any interval of interest, for example, for one week $\mathbf{P}_{week} = \begin{pmatrix} .89 & .11 \\ .08 & .92 \end{pmatrix}$, six month $\mathbf{P}_{0.5year} = \begin{pmatrix} .43 & .57 \\ .42 & .58 \end{pmatrix}$ and a year $\mathbf{P}_{year} = \begin{pmatrix} .43 & .57 \\ .43 & .57 \end{pmatrix}$, showing how transitions become more likely up to a certain point in time. Looking at the estimated initial state probabilities $\boldsymbol{\pi} = (.9 \quad .1)$ and Figure 3.2, which shows the transitions between states over time for six exemplary persons in the sample, it becomes apparent that patients have a high probability of starting in state 1 with the trend of moving towards state 2. Combined with knowing what the MMs look like, we conclude that, over time, patients obtained a more unified concept of depression (high loadings on only one factor), improved assessing their degree of symptoms by means of the BDI (lower unique variances), and perceived less symptoms (lower intercepts) than at the beginning of their therapy. This broadly confirms previous research of Fokkema, Smits, Kelderman, and Cuijpers (2013) who compared the screening and termination MMs of this dataset with CFA and found that the participants obtained a more concrete idea of their depression, perhaps because therapists explain the concept of depression during

[19] Note that some additional information about choices made regarding the data is provided in Appendix G.
[20] The syntax for the final model can be found in Appendix F.

sessions so that patients learn about their illness, which may influence patients' concepts of depression and how they evaluate their symptoms. However, due to the pure exploratory nature of this study, drawing substantive conclusions would require more research, such as a replication study.

3.5 Discussion

In this chapter, we introduced continuous-time- (CT-) latent Markov factor analysis (LMFA)—which models measurement model (MM) changes in time-intensive longitudinal data with unequal measurement intervals—and compared the method to the regular discrete-time- (DT-)LMFA. Although the recovery of states was only slightly superior in CT-LMFA, we demonstrated why the method should be favored: CT-LMFA has a natural match with the assumption that processes evolve at irregular time intervals (TIs) and transition intensities can be transformed to DT transition probabilities for arbitrary TIs. This allows researchers to compare transition probabilities within and across studies, leading to more freedom in interpreting time-intensive longitudinal data.

CT-LMFA is a valuable first data-analysis step because, by pinpointing changes in the MM, it safeguards valid results when further investigating factor mean changes (e.g., by means of ctsem; Driver et al., 2017). For example, the structure of the MM in one state might indicate the presence of a response style. Researchers may then continue with the "reliable" part of the data only (i.e., the measures in the state without the response style) or choose to correct for the response style in the corresponding part of the data. If only, say, two item loadings are invariant across states, researchers may decide to remove these items and to continue with the entire dataset. CT-LMFA may also indicate that there are unobserved groups of subjects that mostly stay in one state. In that case, a mixture CT- structural equation modeling analysis with latent subpopulations could be a suitable next step.

In the future, one would ideally use hypothesis tests to trace significant differences across the states. This will be possible by means of Wald tests once rotational freedom is dealt with in the estimation procedure so that proper standard errors are obtained. To solve the rotation problem for multiple groups simultaneously, De Roover and Vermunt (2019) recently developed a "multigroup factor rotation", which rotates group-specific loadings both to simple structure and between-group agreement. The next step is to tailor this promising method to the states in CT-LMFA and, thereby, to enable hypothesis testing.

Table 3.2. Standardized oblimin rotated factor loadings, intercepts, and unique variances of the CT-LMFA model with two states and respectively three and two factors for the Beck Depression Inventory repeated-measures application data

	State 1					State 2			
	Factors			Int.	Unique V.	Factors		Int.	Unique V.
Items	Despair	Self-image	Cognition/Body			Depression	Cognition		
Mood	**.44**	.22	**.32**	1.20	.30	**.73**	-.03	0.40	.16
Pessimism	**.56**	.29	.08	1.32	.31	**.77**	-.17	0.38	.15
Sense of Failure	**.45**	**.52**	-.03	1.29	.27	**.79**	-.25	0.34	.13
Lack of Satisfaction	**.55**	-.09	**.42**	1.38	.28	**.70**	.29	0.55	.14
Guilty Feeling	.13	**.62**	-.02	1.20	.49	**.49**	-.10	0.23	.18
Sense of Punishment	-.06	**.43**	-.01	0.98	.93	.00	.00	0.00	.00
Self-Hate	.20	**.60**	.17	1.39	.25	**.77**	-.15	0.52	.16
Self-Accusations	.20	**.70**	-.08	1.30	.25	**.80**	-.28	0.50	.14
Self Punitive Wishes	**.41**	.22	.05	0.65	.40	.00	.00	0.00	.00
Crying Spells	.01	.23	**.45**	0.97	.68	**.43**	.10	0.19	.13
Irritability	.02	.14	**.57**	0.97	.31	**.55**	.24	0.53	.21
Social Withdrawal	.23	.04	**.62**	1.14	.30	**.66**	.25	0.46	.20
Indecisiveness	.18	.18	**.48**	1.25	.43	**.64**	**.34**	0.50	.18
Body Image	-.10	**.57**	.18	1.23	.59	**.62**	-.18	0.53	.42
Work Inhibition	**.37**	.14	.28	1.29	.31	**.64**	**.33**	0.49	.15
Sleep Disturbance	.05	-.07	**.58**	1.26	.63	**.42**	.24	0.51	.33
Fatigability	.26	-.01	**.59**	1.32	.32	**.63**	.29	0.61	.18
Loss of Appetite	-.01	-.10	**.45**	0.75	.54	.00	.00	0.00	.00
Somatic Preoccupation	-.15	.10	**.40**	0.56	.38	**.42**	-.04	0.29	.22
Loss of Libido	.04	.00	**.55**	1.08	.75	**.44**	.20	0.49	.43

Note. Int. = Intercepts; V. = Variance; To aid interpretation, we standardized factor loadings by dividing them by the state-specific item standard deviations. Loadings with an absolute value larger than 0.3 are depicted in boldface. In state 1, Cor(Despair, Self-image) = .57, Cor(Despair, Cognition/Body) = -.28, and Cor(Self-image, Cognition/Body) = -.25; In state 2, Cor(Depression, Cognition/Body) = -.77.

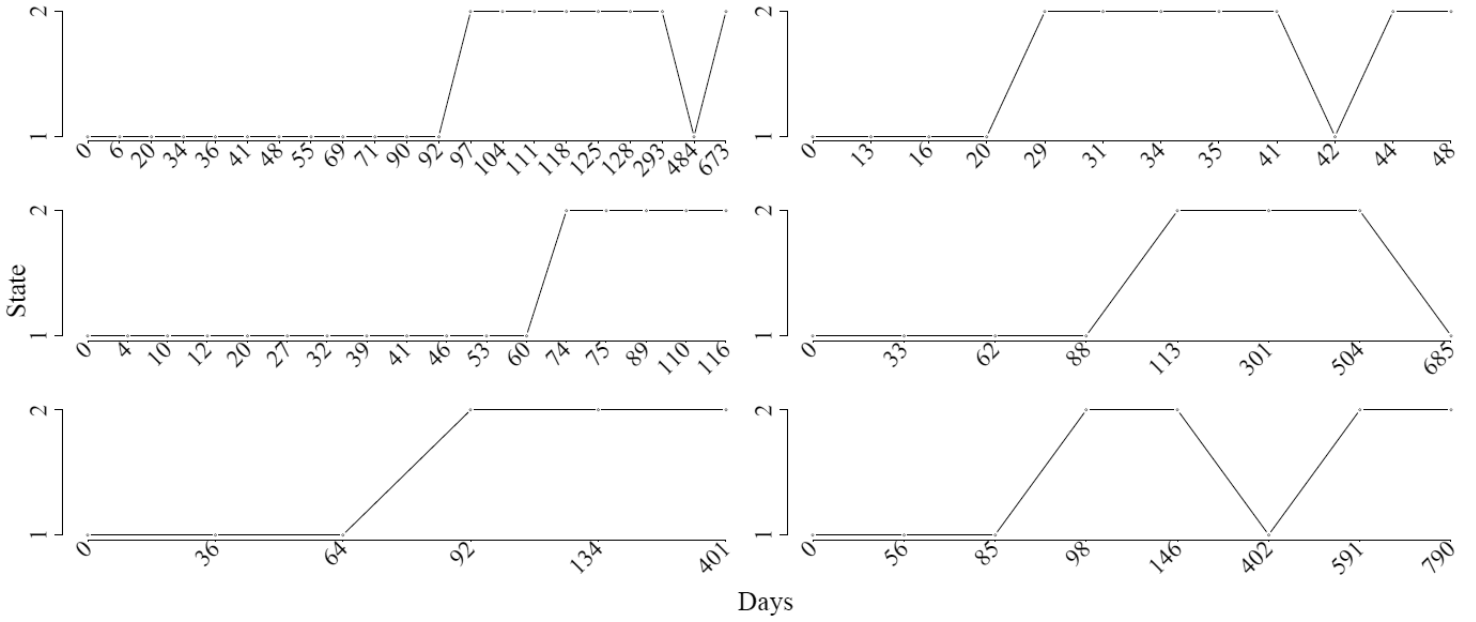


Figure 3.2. Six representative examples of individual transition plots. Note that the scale of the spacing of the x-axis is not in line with the amount of days elapsed but, to enable the illustration, equal spaces are chosen.

4

Three-Step Latent Markov Factor Analysis for Evaluating Covariate Effects

This chapter is accepted for publication in a similar form in Multivariate Behavioral Research as Vogelsmeier, L. V. D. E., Vermunt, J. K., Bülow, A., & De Roover, K. (in press) *Evaluating covariate effects on ESM measurement model changes with latent Markov factor analysis: A three-step approach.*

Abstract

Invariance of the measurement model (MM) between subjects and within subjects over time is a prerequisite for drawing valid inferences when studying dynamics of psychological factors in intensive longitudinal data. To conveniently evaluate this invariance, latent Markov factor analysis (LMFA) was proposed. LMFA combines a latent Markov model with mixture factor analysis: The Markov model captures changes in MMs over time by clustering subjects' observations into a few states and state-specific factor analyses reveal what the MMs look like. However, to estimate the model, Vogelsmeier, Vermunt, van Roekel, et al. (2019) introduced a one-step (full information maximum likelihood; FIML) approach that is counterintuitive for applied researchers and entails cumbersome model selection procedures in the presence of many covariates. In this chapter, we simplify the complex LMFA estimation and facilitate the exploration of covariate effects on state memberships by splitting the estimation in three intuitive steps: (1) obtain states with mixture factor analysis while treating repeated measures as independent, (2) assign observations to the states, and (3) use these states in a discrete- or continuous-time latent Markov model taking into account classification errors. A real data example demonstrates the empirical value.

4.1 Introduction

New methods such as experience sampling methodology (ESM; Scollon et al., 2003) enable the assessment of psychological constructs or "factors" (e.g., depression) in daily life by repeatedly questioning multiple participants via smartphone apps, for example, nine times a day for one week. Such intensive longitudinal studies (say with more than 50 measurement occasions) are often conducted to analyze dynamics in factor means. For instance, researchers investigated how emotional dynamics relate to subjects' mental health (Myin-Germeys et al., 2018) or tailored interventions to subject's daily experience of negative affect (Van Roekel et al., 2017). For drawing valid inferences about the dynamics, it is crucial that the measurement model (MM) is invariant (i.e., constant) between and within persons over time. The MM indicates which factors are measured and how these factors are measured by items, which is expressed by means of "factor loadings". In case of continuous data, the MM is obtained with factor analysis (FA). If the MM is invariant, the factors are conceptually equal across subjects and time-points and therefore comparable. However, the MM might be affected by subject- or time-point-specific response styles or substantive changes in item interpretation. As a result, the MMs might differ between subjects (e.g., the item interpretation might depend on subjects' psychopathology) but the MM might also differ within one subject (e.g., the response style of choosing only the extreme categories might depend on situational motivation to complete the questionnaire). If invariance stays undetected, inferences may be invalid. For example, a mean score change in negative affect might be at least partly due to changes in item interpretations.

To conveniently evaluate (violations of) invariance of intensive longitudinal data for multiple subjects simultaneously, latent Markov factor analysis (LMFA; Vogelsmeier, Vermunt, van Roekel, et al., 2019; Vogelsmeier, Vermunt, Böing-Messing, & De Roover, 2019) was proposed, which combines a discrete- or continuous-time latent Markov model with mixture factor analysis.[21] As will be described in more detail in Section 4.2.3, the Markov model clusters subject- and time-point-specific observations according to their underlying MM into dynamic latent MM classes or "states", which implies that subjects can transition between latent states and thus between MMs over time. State-specific factor analyses reveal which MM applies to each state. Observations that belong to the same state are invariant. Observations that belong to different states are non-invariant, for instance, with regard to the number or nature of the underlying factors or the size of factor loadings. Note that some subjects might stay in one state, which implies

[21] Note that it is also possible to apply LMFA to a single subject if the number of observations was large enough. For guidelines on the required number of observations, see Vogelsmeier, Vermunt, van Roekel, et al. (2019).

within-person invariance (i.e., over time). Other subjects may transition (more or less frequently) between different MMs, which implies within-person non-invariance. Moreover, comparing state memberships across subjects provides information about between-person (non-)invariance.

The aim of assessing non-invariance patterns is usually twofold. On the one hand, detecting non-invariance is important for deciding how to proceed with the data analysis. For example, if the invariance violation is strong, one may decide to conduct factor-mean analyses with observations from one state only. If only a few MM parameters differ across states (i.e., "partial invariance" holds; Byrne et al., 1989), one may decide to investigate dynamics in the factor means but let the corresponding MM parameters differ across states. More specifically, if discrete (i.e., abrupt) changes are of interest, one could continue with LMFA by adding factor means to the model and constraining invariant parameters to be equal across states. The state memberships would then (also) capture discrete changes in factor means (this is further explained in Section 4.2.3.1).[22] If continuous (i.e., smooth) changes are of interest, researchers could opt for a latent growth model (Muthén, 2002) with state-specific parameters.

On the other hand, researchers would typically like to include explanatory variables (in the following referred to as "covariates") that can possibly explain MM changes so they can learn about these substantively interesting aspects of their data. As an example, when studying adolescents' affective well-being in daily life, the situational context (e.g., being with friends versus being with parents) might lead to different MMs in that some items may be more relevant for measuring affect in one over the other context. For instance, "being excited" might be more related to the positive affect factor when being with friends whereas "being content" might be more related to positive affect when being with parents.

Exploring the relations between covariates and state memberships is theoretically possible by adding different (sets of) covariates to the "structural model" (SM). Note that the SM generally refers to the causal relationships between latent variables (and/or exogenous variables) or between latent variables at consecutive measurement occasions. Specifically, in LMFA, the SM refers to the transitions between states and thus between MMs. However, with the currently implemented full information maximum likelihood (FIML) approach, that estimates all parameters (i.e., from the MM and the SM) at the same time, exploring covariate effects is cumbersome. LMFA is an exploratory method, which entails that researchers have to select the best model in terms of number of latent states

[22] This analysis would be comparable to the factor-mean modeling approach that was proposed by Bartolucci and Solis-Trapala (2010).

and number of factors within the states. To this end, one needs to estimate a large number of plausible models and compare them with the Bayesian information criterion (Vogelsmeier, Vermunt, van Roekel, et al., 2019) or an alternative model selection criterion. For example, comparing models with $1-3$ states and $1-3$ factors per state would already result in 19 models that have to be estimated by the researchers. Model selection with covariates is even more cumbersome because the whole model (i.e., the MMs and the SM) would have to be re-estimated for every set of covariates. Especially in exploratory studies, where researcher might want to add or remove covariates until only significant covariates are left, the model selection procedure quickly becomes unfeasible (e.g., if there are only three different sets of covariates for the 19 different model complexities, this would already result in $19 \times 3 = 57$ models).

To avoid the model selection problem with covariates in any latent class analysis (the latent Markov model is a specific variant thereof), researchers sometimes first select the MM (or MMs if they differ across latent classes) without including the covariates to the SM. Once the choice about the complexity of the MM has been made, researchers include the covariates in the SM and re-estimate the whole model (i.e., only $19 + 3 = 21$ models have to be estimated). However, this is problematic because, in the FIML approach, both parts of the model, the MM and the SM, are estimated at the same time so that specifications of the SM may also influence the MM. Thus, including covariates can redefine the states and even impact the optimal number of states or factors (Bakk, Tekle, & Vermunt, 2013; Nylund-Gibson, Grimm, Quirk, & Furlong, 2014).

A better strategy that considerably simplifies the estimation is the so-called "three-step" ("3S") approach, which decomposes the estimation into three manageable pieces. More specifically, the steps for a latent Markov model are as follows: step 1: obtain state-specific MMs by conducting mixture factor analysis on the repeated measures data while disregarding the dependency of these observations; step 2: assign the observations to the states (and thus the MMs) based on posterior state probabilities; step 3: pass the state-specific MMs to a latent Markov model in order to estimate the probabilities to transition between the states (the three steps will be elaborated in Section 4.2.5). Although the MMs are also estimated first without considering the SM with its covariates (step 1), the MMs are kept fixed when adding the covariates to the SM (step 3).

Next to facilitating the inclusion of covariates, the step-wise approach is also more intuitive because it ensures that the states—and thus the formation of the MMs—are free from covariate influences. This is required if latent classes (in our case latent states) should exclusively capture heterogeneity in the MMs (Bakk et al., 2013; Nylund-Gibson et al., 2014). Moreover, the step-wise approach better corresponds with how researchers prefer to approach their analyses (Bakk et al., 2013; Devlieger, Mayer, & Rosseel, 2016;

Vermunt, 2010). That is, they rather see the investigation of the SM (i.e., in our case the transitions between states and the influence of covariates) as a final step that comes after investigating what the MMs look like. Because of the separate steps, the analyses could even be distributed across researchers such that one researcher carries out the first step to obtain the different underlying MMs. A second researcher could take the results and continue with the analyses of the transitions between the MMs. If everything has to be done in one step, it may quickly become overwhelming. Thus, applied researchers are used to and typically prefer such step-wise approaches and perceive simultaneous "one-step approaches" as counterintuitive and more difficult to interpret (Vermunt, 2010). Especially for complex analyses such as LMFA, offering step-wise approaches can therefore help to reach applied researchers and motivate them to apply the new method.

When splitting the estimation of latent class models in general—and latent Markov models in particular—into the estimation of the MM(s) and the SM, estimates of the SM would be biased, however. In order to prevent this bias, the estimation procedure has to take into account the classification error that results from classifying observations into classes or states because classification is never perfect. To this end, Bolck, Croon, and Hagenaars (2004) proposed the "BCH" method in which the classification error is used to reweight the data prior to conducting logistic regressions to predict class membership. Moreover, Vermunt (2010) developed an alternative, more flexible, maximum likelihood correction ("ML" method) in which the estimation of the latent class model in the third step explicitly incorporates the classification error. More recently, the ML approach (or an extension thereof) was applied to the 3S estimation of latent Markov models (e.g., Asparouhov & Muthén, 2014a; Bartolucci, Montanari, & Pandolfi, 2015; Di Mari et al., 2016; Nylund-Gibson et al., 2014) and showed to be a trustworthy alternative to the one-step FIML approach.

The aims of the current study are (1) to tailor the ML correction method to LMFA (in the following referred to as 3S-LMFA) to provide a more accessible alternative to the FIML estimation (in the following referred to as FIML-LMFA) that is more convenient to use (especially with covariates) and easier to interpret for applied researchers and (2) to evaluate whether 3S-LMFA approaches the good performance of FIML-LMFA in terms of state and parameter recovery. Note that both, FIML-LMFA as well as 3S-LMFA, can be estimated by means of Latent GOLD (LG) syntax (Vermunt & Magidson, 2016), which is also used for the current study.

The remainder of this chapter is structured as follows: In Section 4.2, we first describe the data structure, provide a motivating example, outline the general LMFA model, and explain the steps of 3S-LMFA. In Section 4.3, by means of a simulation study, we evaluate the performance of 3S-LMFA and compare it to the performance of FIML-

LMFA. Section 4.4 illustrates the empirical value of 3S-LMFA by means of a real-data application. In Section 4.5, we discuss limitations of 3S-LMFA and directions for future research.

4.2 Method

In the following, we first describe the data structure, the LMFA model, and the FIML estimation before we explain the three steps of the 3S estimation in detail.

4.2.1 Data Structure

We assume ESM data with repeated measures observations (with multiple continuous variables) that are nested within subjects and are denoted by y_{ijt} (where $i = 1, \dots, I$ refers to subjects, $j = 1, \dots, J$ refers to items, and $t = 1, \dots, T$ to time-points). Note that T may differ across subjects but we omit the index i in T_i for simplicity. The y_{ijt} are collected in the $1 \times J$ vectors $\mathbf{y}_{it} = \left(y_{i1t}, \dots, y_{iJt}\right)$ that themselves are collected in the $T \times J$ data matrices $\mathbf{Y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iT})'$.

4.2.2 Motivating Example

In order to motivate the use of LMFA in general and the 3S approach in particular, consider the following ESM data that was collected within the larger ADAPT project (Keijsers, Boele, & Bülow, 2017).[23] Dutch adolescents (N = 27; M_{Age} = 15.8; 67% girls) received five questionnaires a day for 13 consecutive days via the "Ethica Data" mobile app (Ethica Data Services Inc, 2018), resulting in a maximum of 65 potential measurement occasions per participant. In total, the 27 participants completed 1168 questionnaires (compliance rate 67%). During each ESM questionnaire the participants indicated their current affect with the Dutch version of the Positive and Negative Affect Schedule for Children (PANAS-C-NL; Ebesutani et al., 2012; Keijsers, Boele, & Bülow, 2019; Watson et al., 1988), where five items indicated positive affect and another five items indicated negative affect (all items are displayed in Table 4.4). All affect items were measured on a Visual Analog Scale (VAS) from 0 (not at all) to 100 (very much). The visual display of the items in the app can be found in Appendix K.4. Next to the affect questionnaire, adolescents also completed questionnaires to assess time-varying covariates (e.g., participants' current company) at each measurement occasion. Furthermore, before the ESM study, participants completed a baseline questionnaire about time-constant covariates (e.g., on emotion clarity and emotion differentiation capability). A typical next step of substantive or applied researchers would be to investigate changes in positive and negative affect over time. However, if response styles

[23] Materials can be found at https://osf.io/svyau.

or item interpretation differ across time-points and/or subjects, the MM is not invariant within and between subjects and conclusions about dynamics in affect may be invalid. LMFA can be used to trace MM differences between subjects and MM changes over time. More specifically, there are two main research questions that can be answered using LMFA:

1. Which MMs underlie which parts of the data and how do the MMs differ?
2. Are the MMs related to time-varying and/or time-constant covariates?

For answering only the first question, FIML-LMFA can be used. However, if researcher also want to answer the second question, the model selection including covariates would be too cumbersome with FIML-LMFA and 3S-LMFA is indispensable. In Section 4.4, we answer both research questions using 3S-LMFA.

4.2.3 Latent Markov Factor Analysis

LMFA consists of two parts. First, the measurement part concerns the state-specific response variable distributions that, in the case of LMFA, consist of the MMs for the constructs, which are defined by a mixture of factor models. Second, the structural part concerns the discrete latent process that is either defined by a "discrete-time" latent Markov model (Bartolucci et al., 2014; Bartolucci, Farcomeni, et al., 2015; Collins & Lanza, 2010; Zucchini et al., 2016), which assumes equal time intervals, or by a "continuous-time" latent Markov model (Böckenholt, 2005; Jackson & Sharples, 2002), which allows time intervals to differ. Additionally, it is possible to include covariates to the SM. Figure 4.1 depicts the relations between the parameters from the SM and zooms in on the relation between the states from the SM and the state-specific MMs by means of an artificial example. The different parts including the notation will be described next.

4.2.3.1 *Measurement Part*

The measurement part shows how the state memberships define the responses. Thereby, it is important to note that the responses at time-point t, $\mathbf{y}_{it}$, depend only on the latent state k ($k = 1, \dots, K$) at that time-point and the responses are thus independent of the responses at other time-points given that state ("independence assumption"), which is also illustrated in Figure 4.1. In LMFA, the factor model depends on the state membership of subject i at time-point t (denoted by $s_{itk} = 1$) as follows:

$$[\mathbf{y}_{it}|s_{itk} = 1] = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \, \mathbf{f}_{it} + \mathbf{e}_{it}. \tag{4.1}$$

In this equation, $\boldsymbol{\Lambda}_k$ is the state-specific $J \times F_k$ loading matrix (where F_k is the state-specific number of factors); $\mathbf{f}_{it} \sim MVN(\mathbf{0}, \boldsymbol{\Psi}_k)$ is the subject-specific $F_k \times 1$ vector of

factor scores at time-point t (where $\mathbf{\Psi}_k$ is the state-specific factor covariance matrix); $\boldsymbol{\nu}_k$ is the state-specific $J \times 1$ intercept vector; and $\mathbf{e}_{it} \sim MVN(\mathbf{0}, \mathbf{D}_k)$ the subject-specific $J \times 1$ vector of residuals at time-point t (with $\mathbf{D}_k$ containing the unique variances d_{kj} on the diagonal and zeros on the off-diagonal). Thus, the state-specific response densities, $p(\mathbf{y}_{it}|s_{itk} = 1)$, are defined by state-specific multivariate normal distributions with means $\boldsymbol{\nu}_k$ and covariance matrices $\mathbf{\Sigma}_k = \mathbf{\Lambda}_k\mathbf{\Lambda}_k' + \mathbf{D}_k$. To obtain the state-specific factor models, LMFA employs exploratory factor analysis within the states in order to retain maximal flexibility regarding the differences in MMs that can be traced. In contrast to confirmatory factor analysis, exploratory factor analysis puts no a priori constraints on the factor loadings. However, if desired, confirmatory factor analysis can also be used.

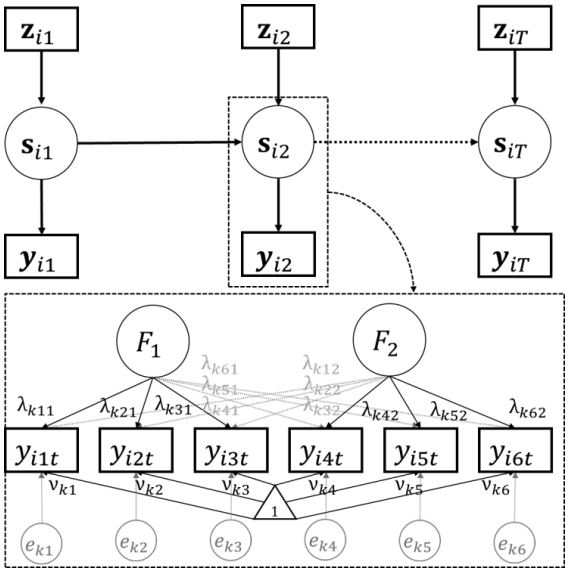


Figure 4.1. Artificial example of the relations between the structural model parameters (top panel) and a zoomed in state-specific measurement model (bottom panel) in the full information maximum likelihood LMFA. Note that the state-specific measurement models may differ regarding all parameters, including the number of factors and the values of the loadings (λ_{kjf}), intercepts (ν_{kj}), and unique errors (e_{kj}).

From Equation (4.1) we can see that the state-specific MMs may differ with regard to their loadings $\mathbf{\Lambda}_k$, intercepts $\boldsymbol{\nu}_k$, unique variances $\mathbf{D}_k$, and factor (co)variances $\mathbf{\Psi}_k$, implying that LMFA explores all levels of measurement non-invariance, that is, configural invariance (invariant number of factors and pattern of zero loadings), weak factorial invariance (invariant loading values), strong factorial invariance (invariant intercepts), and strict factorial invariance (invariant unique variances) (for more details see, e.g., Meredith, 1993). For identification purposes, the factor variances are equal to one in all the states and rotational freedom is dealt with by means of criteria to optimize simple

structure and/or the between-state agreement of the factor loadings (e.g., Kiers, 1997; Clarkson & Jennrich, 1988; De Roover & Vermunt, 2019).

It is important to note that restricting the factors to have a mean of zero and a variance of one has the consequence that changes in factor means and variances may be captured as changes in the intercepts and loadings (i.e., if an additional state is selected for such a change). Therefore, when all intercepts that pertain to the same factor are higher or lower in one state compared to the other, it might be a sign that the factor means rather than separate intercepts differ across these states. Similarly, if all loadings of the same factor are likewise larger or smaller (i.e., the scaling is affected), it might be a sign that factor variances rather than the separate loadings differ across states. However, when the number of factors differs across states, it does not make sense to disentangle loading differences from factor-variance differences and, as long as weak invariance is violated, it does not make sense to disentangle intercept differences from factor-mean differences. In contrast, if the loadings and intercepts are (at least partially) invariant, one could go ahead with an adjusted LMFA—that means, with equality restrictions on the MM parameters and including state-specific factor variances and means—and capture discrete changes in factor variances and means over time, as was already mentioned in the introduction (Section 4.1).

Furthermore, LMFA currently assumes that factors have no auto- and cross-lagged correlations at consecutive time-points. By means of a dynamic factor analysis, it would be possible to incorporate such autocorrelations, but factor rotation would be more intricate as auto- and cross-lagged relations have to be rotated towards a priori specified target matrices (Browne, 2001; Zhang, 2006). This would require a priori hypotheses about MM changes that are often not present or incomplete and that are thus undesired in exploratory studies. In addition, one would require more measurement occasions per subject (Ram et al., 2012), which is often unfeasible. In order to investigate whether ignoring autocorrelations in the data would pose problems for LMFA, Vogelsmeier, Vermunt, van Roekel, et al. (2019) conducted a simulation study using the FIML estimation and showed that the state and parameter recovery of the MMs were largely unaffected. Note, however, that ignoring dependency in the data leads to an underestimation of standard errors (SEs) of the MM parameters. This is only relevant when using hypothesis tests to trace significant differences in the MMs across states, which is possible by means of Wald tests using De Roover and Vermunt (2019)'s recently developed "multigroup factor rotation".[24] One would then have to correct for the

[24] The method solves the rotation problem for multiple groups simultaneously by rotating group-specific factor loadings to simple structure and between-group agreement with user-defined weights on these two aspects of the rotation.

dependency in the data (e.g., with the "primary sampling unit" identifier in LG; Vermunt & Magidson, 2016). Otherwise, invariance of parameters would be rejected too easily. However, the hypothesis tests are outside the scope of this chapter.

4.2.3.2 *Structural Part*

A latent Markov model generalizes a latent class model—the statistical method to identify subgroups with a similar set of indicator values—because subjects can transition between classes in a latent Markov model, while subjects remain in the same classes in a latent class model. The classes in a latent Markov model are therefore referred to as "states". For an extensive description of latent Markov models, see, for example, Bartolucci, Farcomeni, et al. (2015) and Zucchini et al. (2016). In brief, transitions between the states are captured by a latent "Markov chain" defined by the probabilities to start in a state k at time-point $t = 1$ ("initial state probabilities") and the probability of being in a state k at time-point $t > 1$ conditional on the occupied state l ($l = 1, \dots, K$) at $t - 1$ ("transition probabilities"). Note that, according to the first-order Markov assumption, the probability of being in a certain state k at time-point t depends only on the state at $t - 1$. The initial state probabilities are given by the $K \times 1$ probability vector $\boldsymbol{\pi}$, which contains the elements $\pi_k = p(s_{1k} = 1)$ with s_{tk} referring to the state membership k at time-point t (e.g., if a subject is in state 1 at time-point 1, then $s_{11} = 1$ and $s_{12} = \cdots = s_{1K} = 0$). These binary variables are in turn collected in the membership vectors $\mathbf{s}_{it} = (s_{it1}, \dots, s_{itK})'$, for $t = 1, \dots, T$, which are in turn collected in the $K \times T$ state membership matrix $\mathbf{S} = (\mathbf{s}_{i1}, \mathbf{s}_{i2}, \dots, \mathbf{s}_{iT})$. The transition probabilities are collected in the $K \times K$ matrix **P,** which contains the elements $p_{lk} = p(s_{tk} = 1 | s_{t-1,l} = 1)$. Note that the rows indicate the state that a person comes from and the columns determine the state where the person transitions to. Hence, the diagonal elements represent the probabilities to stay in a state and the off-diagonal elements the probabilities to transition to another state. Therefore, diagonal values close or equal to 1 indicate stable state memberships and thus within-person invariance. It applies that the sum of the initial state probabilities, $\sum_{k=1}^{K} \pi_k$, and the row sums of the transition probabilities, $\sum_{k=1}^{K} p_{lk}$, equal 1.

In the discrete-time latent Markov model, the time intervals between observations are assumed to be equal. This assumption is often not tenable in empirical data. For instance, the questionnaires in ESM are usually send out at random moments and participants may skip certain measurement occasions, which automatically increases the distance between two subsequent observations. To accommodate such data, a continuous-time latent Markov model can be employed, which allows for differing intervals across time-points and subjects by considering the length of time spent in a state, δ. In the following, we provide a brief summary. The interested reader is referred to Böckenholt (2005) and Jackson and Sharples (2002) for general information about

continuous-time latent Markov model and to Vogelsmeier, Vermunt, Böing-Messing, et al. (2019) for more specific information on continuous-time-LMFA. In brief, transitioning from the origin state l to destination state k is defined by the "intensities" (or rates) q_{lk} (collected in the $K \times K$ intensity matrix $\mathbf{Q}$) that replace the transition probabilities p_{lk} and can be seen as probabilities to transition between states per very small time unit:

$$q_{lk} = \lim_{\delta \to 0} \frac{p(s_{tk} = 1 | s_{t-\delta,l} = 1)}{\delta}, \tag{4.2}$$

for all $k \neq l$ (thus, for the off-diagonal elements in the intensity matrix $\mathbf{Q}$). The diagonal elements are equal to the negative row sums (i.e., $-\sum_{k \neq l} q_{lk}$; Cox & Miller, 1965). The transition probabilities for any interval of interest can be computed by taking the matrix exponential of $\mathbf{Q} \times \delta$. Note that larger time intervals δ increase the probability to transition to a different state. In turn, $\mathbf{Q}$ can be obtained by taking the matrix logarithm of $\mathbf{P}$.[25]

In the following, we expand the structural part by including U subject and possibly time-point specific covariates z_{itu} (collected in the $U \times 1$ vectors $\mathbf{z}_{it}$) such that they affect the initial and transition probabilities.[26] Note that the measurement part is assumed to be not (directly) affected by the covariates, which can also be seen in Figure 4.1. Also note that the parameters of the structural part are typically modelled using a logit model (for initial and transition probabilities) or via a log-linear model (for transition intensities) in order to prevent parameter space restrictions, which is also what LG does. The covariates enter the model through these parameterizations. For the initial state probabilities, we use the parameterization

$$\log \frac{p(s_{i1k} = 1 | \mathbf{z}_{i1})}{p(s_{i11} = 1 | \mathbf{z}_{i1})} = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{z}_{it=1}, \tag{4.3}$$

with $k = 2, \dots, K$ and $k = 1$ as the reference category. The coefficients β_{0k} are the initial state intercepts and $\boldsymbol{\beta}_k' = (\beta_{k,z_{i11}}, \dots, \beta_{k,z_{i1U}})'$ are the initial state slopes, which quantify

[25] Note that the $\mathbf{Q}$ matrix with the particular structure on the off-diagonals follows naturally from taking the matrix logarithm of the $\mathbf{P}$ matrix with its restriction $\sum_{k=1}^{K} p_{lk} = 1$.
[26] Note that the only difference between time-varying and time-constant covariates is that the former may take different values within a subject (i.e., in the dataset, the covariate scores may differ across rows) and the latter has the same value within a subject (i.e., in the dataset, the covariate scores are repeated/identical across rows).

the effect of the covariates on the initial state memberships. In discrete-time-LMFA, the multinomial logistic model for the transition probabilities is

$$\log \frac{p(s_{itk} = 1 | s_{it-1,l} = 1, \mathbf{z}_{it})}{p(s_{itl} = 1 | s_{it-1,l} = 1, \mathbf{z}_{it})} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk} \mathbf{z}_{it} \tag{4.4}$$

with $k \neq l$. Thus, the logit is modeled by comparing the transition from state l to state k with the probability of staying in state l. The coefficients γ_{0lk} are the transition intercepts and $\boldsymbol{\gamma}'_{lk} = (\gamma_{lk,z_{itu}}, \dots, \gamma_{lk,z_{itU}})'$ are the transition slopes, which quantify the effect of the covariates on transitioning to another state. In continuous-time-LMFA, we use a log-linear model for the transition intensities (for $k \neq l$):

$$\log q_{lk} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk} \mathbf{z}_{it}. \tag{4.5}$$

Finally, the joint distribution of observations and states, given the covariates, is

$$\begin{aligned} p(\mathbf{Y}_i, \mathbf{S}_i | \mathbf{Z}_i) &= p\big(\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT,} \mathbf{s}_{i1}, \dots, \mathbf{s}_{iT} | \mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}\big) \\ &= p(\mathbf{s}_{i1} | \mathbf{z}_{i1}) \prod_{t=2}^{T} p_{\delta_{ti}}(\mathbf{s}_{it} | \mathbf{s}_{it-1}, \mathbf{z}_{it}) \prod_{t=1}^{T} p(\mathbf{y}_{it} | \mathbf{s}_{it}). \end{aligned} \tag{4.6}$$

Note that the δ_{ti} in $p_{\delta_{ti}}(\mathbf{s}_{it} | \mathbf{s}_{it-1}, \mathbf{z}_{it})$ refers to the transition probabilities' dependency on the subject- and time-point-specific time interval in continuous-time-LMFA. The term reduces to $p(\mathbf{s}_{it} | \mathbf{s}_{it-1}, \mathbf{z}_{it})$ in discrete-time-LMFA.

4.2.4 FIML Estimation of Latent Markov Factor Analysis (FIML-LMFA)

In order to obtain the maximum likelihood (ML) parameter estimates with the FIML estimation, the following loglikelihood function has to be maximized:

$$\log L_{FIML} = \sum_{i=1}^{I} \log \left(\sum_{\mathbf{s}_{i1}} \dots \sum_{\mathbf{s}_{iT}} p(\mathbf{Y}_i, \mathbf{S}_i | \mathbf{Z}_i) \right), \tag{4.7}$$

with $p(\mathbf{Y}_i, \mathbf{S}_i | \mathbf{Z}_i)$ as given in Equation (4.6). The ML estimates can be obtained by means of the forward-backward algorithm (Baum et al., 1970), which is an efficient version of the expectation maximization (EM; Dempster et al., 1977) algorithm and is also utilized by LG to find the ML solution. Within the maximization-step, a Fisher algorithm is used to

update the state-specific covariance matrices defined by the factor models (Jennrich & Sampson, 1976) and, in case of continuous-time-LMFA, also to update the log-transition intensities. For a summary of the algorithms (including information about the convergence criteria and the utilized multistart procedure), see Vogelsmeier, Vermunt, van Roekel, et al. (2019) for discrete-time-LMFA and Vogelsmeier, Vermunt, Böing-Messing, et al. (2019) for continuous-time-LMFA.

It is important to note that we assume the number of states (K) and factors per state (F_k) to be known when estimating the models. However, in real data, the best model in terms of the number of states and factors has to be evaluated. The Bayesian information criterion (BIC) performs well in selecting the best model in FIML-LMFA although the final decision regarding the optimal model should also take interpretability into account (Vogelsmeier, Vermunt, van Roekel, et al., 2019). When also including covariates, every model under comparison (i.e., with all possible combinations of K and F_k) has to be re-estimated every time a covariate is added or removed from the model because, using FIML estimation, the best model may change depending on the included covariates. For instance, when researchers want to obtain the best subset of $U = 3$ covariate candidates, they would have to estimate $2^U = 8$ times the number of models that is already under comparison. When all models are estimated, one may use the BIC and interpretability to choose the final model. Thus, the model selection quickly becomes overwhelming and even unfeasible when exploring relations between state memberships and many covariates.

4.2.5 Three-Step Estimation of Latent Markov Factor Analysis (3S-LMFA)

In contrast to FIML-LMFA, 3S-LMFA decomposes the maximization problem for estimating the MMs and the SM into smaller parts. First, the state-specific MMs are estimated (step 1). Second, the observations are assigned to the MMs (i.e., classified to the states) and "classification errors" are calculated (step 2). Finally, the SM is estimated using the state-assignments while correcting for the classification errors (step 3). In the following, we explain the three steps in detail.

4.2.5.1 *Step 1: Estimation of the State-Specific Measurement Models*

The first step as illustrated in Figure 4.2 involves estimating the state-specific MMs underlying the data by means of mixture factor analysis (McLachlan & Peel, 2000; McNicholas, 2016). The structural part (including the covariates) can be validly ignored because, in LMFA, the observations at a given time-point t, $\mathbf{y}_{it}$, are assumed to be conditionally independent of the state at time-point $t-1$, $s_{itl} = 1$, and the covariates at time-point t, $\mathbf{z}_{it}$, given the state membership at time-point t, $s_{itk} = 1$ (see Figure 4.1). For the estimation, all repeated observations are treated as "independent" such that

respectively, say, 100 observations for each of 100 subjects results in 10000 independent observations. The model parameters of interest are the state proportions $p(s_{itk} = 1)$ and the state-specific response probabilities $p(\mathbf{y}_{itk}|s_{itk} = 1) = MVN(\mathbf{y}_{itk}|\boldsymbol{\nu}_k, \boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k + \mathbf{D}_k)$. The mixture factor analysis model is therefore

$$p(\mathbf{y}_{it}) = \sum_{k=1}^{K} p(s_{itk} = 1)\, p(\mathbf{y}_{itk}|s_{itk} = 1) \tag{4.8}$$

and the loglikelihood function is

$$\log L_{STEP1} = \sum_{i=1}^{I}\sum_{t=1}^{T} \log p(\mathbf{y}_{it}). \tag{4.9}$$

In LG, the posterior state probabilities and the state-specific factor models are estimated with an EM algorithm with Fisher scoring (Lee & Jennrich, 1979) in the maximization-step.[27]

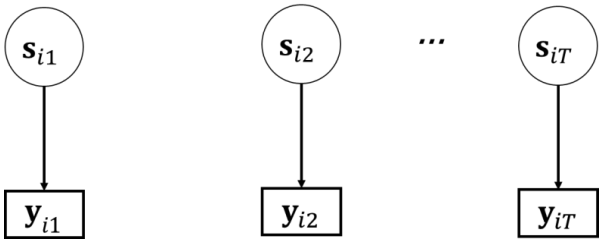


Figure 4.2. Step 1: Estimating the measurement model by performing mixture factor analysis.

As already discussed in the introduction (Section 4.1), in this step, one also selects the optimal number of states K and factors per state F_k without having to be concerned about the covariates. Although the BIC is also a commonly used model selection criterion for mixture factor analysis (McNicholas, 2016), the CHull (Ceulemans & Kiers, 2006) method—which also balances model complexity and fit—proved to outperform the BIC in mixture factor analysis, especially when considering the three best models (Bulteel, Wilderjans, Tuerlinckx, & Ceulemans, 2013). Based on their results, we suggest to use the

[27] Alternatively, one may also use another EM algorithm in the maximization-step (e.g., McNicholas, 2016).

CHull method, potentially combined with the BIC, to select the three best models and compare them in terms of interpretability.

4.2.5.2 *Step 2: Classification of Observations and Calculation of the Classification Error*

Once the state-specific MMs have been estimated, in the second step, we allocate each observation to one of the K states (see Figure 4.3). Therefore, we create a new variable $\mathbf{w}_{it} = (w_{it1}, \dots, w_{itK})'$, that, similar to $\mathbf{s}_{it}$, represents the assignments of the observations to the estimated MMs from step 1. These predicted state memberships are based on the estimated posterior state probabilities $p(s_{itk} = 1|\mathbf{Y}_{it})$ from step 1, which can be expressed using Bayes' theorem as

$$p(s_{itk} = 1|\mathbf{y}_{it}) = \frac{p(s_{itk} = 1)p(\mathbf{y}_{it}|s_{itk} = 1)}{p(\mathbf{y}_{it})} = \frac{p(s_{itk} = 1)p(\mathbf{y}_{it}|s_{itk} = 1)}{\sum_{k'=1}^{K} p(s_{itk'} = 1)p(\mathbf{y}_{it}|s_{itk'} = 1)}. \tag{4.10}$$

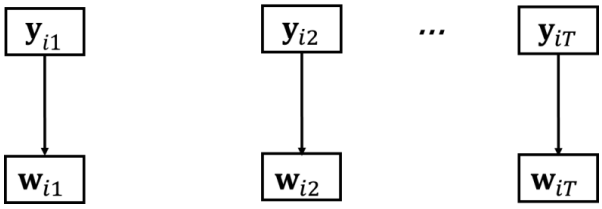


Figure 4.3. Step 2: Assigning states and calculating the classification error.

Thus, all observations $\mathbf{y}_{it}$ belong to each of the K states with a certain probability $p(s_{itk} = 1|\mathbf{y}_{it})$. There are two common rules[28] on how to proceed with these posterior state probabilities with regard to the final state assignments. First, "proportional assignment" assigns a state according to the posterior probabilities such that $p(w_{itk} = 1|\mathbf{y}_{it}) = p(s_{itk} = 1|\mathbf{y}_{it})$, which leads to a "soft" partitioning. Second, "modal assignment" allocates the weight $p(w_{itk} = 1|\mathbf{y}_{it}) = 1$ for the state k with the largest posterior state probability in $\mathbf{s}_{it}$ and a zero weight for all others states. Note that we will focus on modal assignment because proportional assignment is unfeasible with a large

[28] Note that also other assignment rules such as random assignment (Goodman, 2007) can be found in the literature but they are less commonly used for the three-step approaches and are therefore not further discussed.

number of time-points per subject, which would involve separate weights for all K^T possible combinations of states in case of classification uncertainty (Di Mari et al., 2016).

Regardless of the assignment rule, classification error is inherent to any assignment procedure because the largest posterior probability is usually not equal to 1. We have to account for this error because, if not accounted for, the error attenuates relationships between variables. On the one hand, this attenuation will lead to an underestimation of the relation among true states $\mathbf{s}_{it}$ at two consecutive time-points and, thus, an overestimation of the transition probabilities away from a state (Vermunt et al., 1999). On the other hand, estimating the relationship between the estimated memberships $\mathbf{w}_{it}$ and covariates $\mathbf{z}_{it}$ —instead of using the true states $\mathbf{s}_{it}$ —causes underestimation of the covariate effects (Di Mari et al., 2016). Hence, a correction for attenuation of relationships due to classification error is necessary.

In order to calculate the classification error so that we can account for it in step 3, we have to obtain the probability of a certain state assignment $w_{itm} = 1$ conditional on the true state $s_{itk} = 1$, $p(w_{itm} = 1|s_{itk} = 1)$, for all $k, m = 1, \dots, K$. These probabilities are collected in the $K \times K$ "classification error probability matrix". They are computed as

$$p(w_{itm} = 1|s_{itk} = 1) = \frac{\int p(w_{itm} = 1|\boldsymbol{y}_{it})p(\boldsymbol{y}_{it})p(s_{itk} = 1|\boldsymbol{y}_{it})\, d\boldsymbol{y}_{it}}{p(s_{itk} = 1)}. \quad (4.11)$$

For the derivation, see the Appendix H.1. To solve this equation, $p(\mathbf{y}_{it})$ can be validly substituted by its empirical distribution (Di Mari et al., 2016; Vermunt, 2010), resulting in

$$p(w_{itm} = 1|s_{itk} = 1) = \frac{\frac{1}{I \times T}\sum_{i=1}^{I}\sum_{t=1}^{T} p(w_{itm} = 1|\mathbf{y}_{it})p(s_{itk} = 1|\mathbf{y}_{it})}{p(s_{itk} = 1)}. \quad (4.12)$$

Note that another option to solve the integral would be to use Monte Carlo simulation. The larger the probabilities for $m = k$ (corresponding to the diagonal elements of the classification error probability matrix), the better the classification and, thus, the smaller the classification error. Note that classification error is strongly related to separation between the states (i.e., how well the latent states are predicted by $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \dots, \mathbf{Y}_I')$; Bakk et al., 2013; Vermunt, 2010). To qualify the separation in any LC analysis, an entropy-based (pseudo) R-squared measure, $R^2_{entropy}$, is commonly used (Lukočienė, Varriale, & Vermunt, 2010; Vermunt & Magidson, 2016; Wedel & Kamakura, 1998). The $R^2_{entropy}$ value defines the relative improvement of predicting the state membership

when using the observations $\mathbf{y}_{it}$ compared to predicting the state membership without $\mathbf{y}_{it}$. Values range from zero (prediction is no better than chance) to one (perfect prediction). State separation (and hence classification error) depends on various factors. For example, it increases with a lower number of states, higher factor overdetermination (which is higher in case of less factors, more variables, or lower unique variances), and lower between-state similarity (determined by larger differences in the state-specific MMs). The $R^2_{entropy}$ values for the different settings in our simulation study will be reported below in Section 4.3.2.

4.2.5.3 *Step 3: Estimation of the Structural Model*

In the final step, we estimate the SM (i.e., the Markov model with covariates), which is illustrated in Figure 4.4. The key to correct for the classification error obtained in step 2 is to show the relationship between the estimated state memberships conditional on the covariates, $p(\mathbf{W}_i|\mathbf{Z}_i)$, and the true state memberships conditional on the covariates, $p(\mathbf{S}_i|\mathbf{Z}_i)$, where $\mathbf{W}_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots, \mathbf{w}_{iT})$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \dots, \mathbf{z}_{iT})$ and $\mathbf{S}_i = (\mathbf{s}_{i1}, \mathbf{s}_{i2}, \dots, \mathbf{s}_{iT})$ (Di Mari et al., 2016). Therefore, we consider the joint probability $p(\mathbf{W}_i, \mathbf{Y}_i, \mathbf{S}_i, \mathbf{Z}_i)$ and solve for $p(\mathbf{W}_i|\mathbf{Z}_i)$ (see Appendix H.2), which results in

$$p(\mathbf{W}_i|\mathbf{Z}_i) = \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}|\mathbf{z}_{i1}) \prod_{t=2}^{T} p(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it}) \prod_{t=1}^{T} p(\mathbf{w}_{it}|\mathbf{s}_{it}). \tag{4.13}$$

It can be seen that Equation (4.13) resembles the FIML-LMFA model from Equation (4.6), marginalized over $\mathbf{S}_i$ and with different response probabilities. It is in fact a latent Markov model with the state assignments from $\mathbf{W}_i$ as single indicators with K categories replacing the observed item responses $\mathbf{Y}_i$. This demonstrates that $\mathbf{Y}_i$ is no longer needed in step 3 if we have the classification error probabilities $p(\mathbf{w}_{it}|\mathbf{s}_{it})$. The response probabilities are fixed to the classification error probabilities and thus do not have to be estimated. Hence, the focus of the latent Markov model changes. Instead of accounting for unobserved heterogeneity of the MMs (as in FIML-LMFA), the latent Markov model accounts for error in the estimated state assignments $\mathbf{W}_i$.

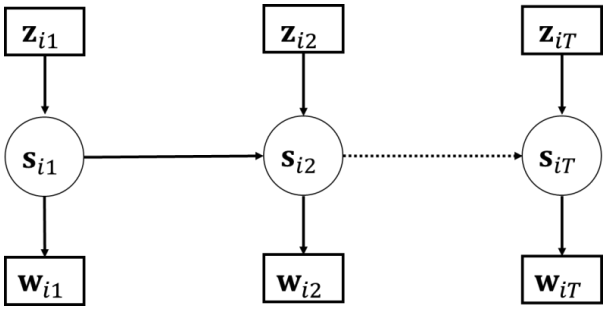


Figure 4.4. Step 3: Estimating the structural model by means of a latent Markov model with single indicators $\mathbf{w}_{it}$.

In order to estimate the SM, the following loglikelihood function has to be maximized:

$$\log L_{STEP3} = \sum_{i=1}^{I} \log\big(p(\mathbf{W}_i|\mathbf{Z}_i)\big). \tag{4.14}$$

The estimation, just as in the regular FIML-LMFA, is done by means of the forward-backward algorithm.[29] However, the classification error probabilities are utilized as fixed response probabilities, such that only the (covariate-specific) transition and initial-state probabilities need to be estimated. Note that the state-assignments $\mathbf{W}_i$ are treated as the manifest (i.e., observed) indicators (that contain error) of the "true" (error-free) latent states $\mathbf{S}_i$, which are inferred through the forward-backward algorithm and used to determine the parameters of the SM. Differences between $\mathbf{W}_i$ and $\mathbf{S}_i$ become less likely for well-separated states with small classification error.

Finally, as already discussed in the introduction (Section 4.1), in the third step, one evaluates which covariates significantly relate to the transition and/or initial state probabilities. Instead of selecting the best subset of covariates by means of an information criterion as in the FIML approach, one may start with a model including all covariate candidates or none of them and use Wald (or likelihood ratio) tests to decide which covariates can be removed from or added to the model one by one (e.g., using forward or backward elimination). Note that, as in any statistical model, there are advantages and disadvantages with regard to such data-driven covariate selection

[29] Note that the third step of 3S-LMFA can be fastened by combining the EM estimation with a Newton-Raphson algorithm which is extensively described in De Roover et al. (2017).

procedures (for a review, see Heinze, Wallisch, & Dunkler, 2018). When in doubt, one may conduct sensitivity analyses comparing the results from different approaches. When having strong a priori hypotheses about covariates, one may also consider a more theory-driven approach.

4.3 Simulation Study

4.3.1 Problem

The aim of the simulation study was to evaluate the performance of 3S-LMFA and to see if it approaches the performance of FIML-LMFA. The specific targeted measures were recovery of the states (i.e., the classification), the MM parameters, and the parameters and SEs of the SM consisting of the Markov model with covariate effects. First, parameter and state recovery have previously been shown to be positively influenced by an increasing amount of information (in terms of sample size) and by higher state-separation (i.e., a better distinction between the states; Bakk et al., 2013; Di Mari et al., 2016; Vermunt, 2010). The more information is available and the more separable the states are, the more accurate the mixture factor analysis can estimate the MM parameters in step 1 and the more accurate the estimation of the SM in step 3.

Second, SEs are possibly slightly underestimated because the error probabilities $p(w_{itm} = 1|s_{itk} = 1)$ are assumed to be known in step 3 although they are actually estimated parameters of the mixture factor analysis in step 1. When the SEs are underestimated, the Wald statistic to test covariate effects would lead to wrong conclusions regarding the statistical significance of covariates. If this underestimation is present, it will likely vanish with large state separation and amount of information (Di Mari et al., 2016; Vermunt, 2010). In the simulation study, we evaluate whether underestimation is present and from what point on state separation and amount of information are sufficient to obtain trustworthy SE values.

Third, the $R^2_{entropy}$ and thus the state separation is higher for FIML-LMFA than for the initial state separation in the first step of 3S-LMFA because the former has additional information from the SM (i.e., the covariates and the states occupied at adjacent time-points) while the latter has information only from the MMs in step 1. Therefore, the recovery of the state memberships is expected to be better for FIML-LMFA. We expect this difference in recovery to decrease when the state memberships are updated in step 3 (i.e., when the SM is also included). However, the degree to which the state membership recovery in 3S-LMFA approaches the recovery in FIML has to be demonstrated in the simulation study.

Note that the evaluation of the model selection procedures in step 1 (i.e., finding the best number of states, K, and factors per state, F_k by means of the BIC and the CHull) and step 3 (i.e., selecting the correct covariates by means of Wald tests, e.g., with backward elimination) is beyond the scope of this chapter and will be used only in the application. As described in Section 4.2.5.1, the BIC and the CHull have already been extensively evaluated for mixture factor analysis. Furthermore, when the simulation study shows that the covariate parameters and their SEs are estimated correctly, we believe that the Wald tests will also correctly identify the significant covariates. However, in Section 4.5, we will discuss the possibility of inaccurate model selection under the violation of the conditional independence assumption.

We manipulated the two key factors: (1) state-separation[30] (this includes (a) between-state loading differences and (b) intercept differences) and (2) amount of information (this includes (c) number of subjects and (d) number of participation days per subject). Note that, for selected conditions, we also investigated whether 3S-LMFA might be more affected by ignoring autocorrelation than FIML-LMFA (see Appendix I.1) and whether varying the strength of the covariate effects and the distribution of the covariates across observations or subjects impacted the estimation procedures differently (see Appendix I.2), which was not the case.

4.3.2 Design and Procedure

The conditions were the following:

- State-separation
  - a. between-state loading differences at two levels: medium loading differences, low loading differences;
  - b. between-state intercept differences at two levels: no intercept differences, low intercept differences;
- Amount of information
  - c. number of Subjects N at four levels: 30, 50, 70, 90;
  - d. number of days D at two levels: 7, 30;

This design resulted in $2 \times 2 \times 4 \times 2 = 32$ conditions. For the population model, we used an ESM setup—with number of subjects, N, days, D, and observations per day, T_{day}—that is often found in practice (e.g., van Roekel et al., 2019; Van Roekel et al., 2017). Furthermore, we used unequal time intervals that are typical for ESM studies and, therefore, employed continuous-time-LMFA. Thereby, the following values were used as constants: number of items $J = 20$, unique variances $e = .2$, number of states $K = 3$,

[30] Note that there are many possibilities to manipulate state-separation as previously stated (e.g., number of factors and states and factor overdetermination). For feasibility of the simulation study, we only chose the two types of between-state differences.

number of factors $F_k = F = 2$, and number of observations per day $T_{day} = 9$. The latter also determined the ESM sampling scheme (comparable to Vogelsmeier, Vermunt, Böing-Messing, et al., 2019): Imposing that a sampling day lasts from 9 am to 9 pm, both day and night intervals were on average 12 hours long. The $T_{day} = 9$ measurement occasions during the day lead to intervals of 1.5 hours if the measurement-occasions were fixed. However, for random variations, we let observations deviate from these fixed time-points by means of a uniform distribution with a maximum of plus and minus 30 percent of the fixed 1.5 hour intervals. Thus, the deviations were drawn from $Unif(-0.3 \times 1.5, 0.3 \times 1.5)$.

To determine the SM, the initial state parameters were chosen to lead to equal probabilities of starting in a state ($\beta_{02} = \beta_{03} = 0$). The transition intercept parameters were specified to be realistic for a short unit interval of 1.5 hours with high probabilities to stay in a state.[31] More specifically, the intercept parameters were $\gamma_{012} = \gamma_{013} = \gamma_{021} = \gamma_{023} = \gamma_{031} = \gamma_{032} = -3.65$ which would correspond to the following transition probability matrix if no covariates were present:

$$\mathbf{P_{day}} = \begin{pmatrix} .950 & .025 & .025 \\ .025 & .950 & .025 \\ .025 & .05 & .950 \end{pmatrix}. \tag{4.15}$$

To alter the transition probabilities, we used one time-varying dichotomous covariate (z_{it1}), which changed in value after 3 days for $D = 7$ or after 15 days for $D = 30$, and one time-constant dichotomous covariate ($z_{it2} = z_{i2}$) that was randomly assigned to the subjects with equal probabilities. Both covariates had values equal to -0.5 or 0.5. A higher value for z_{it1} lowered the probabilities of transitioning to and staying in state 1 and 3 while increasing the probabilities of transitioning to and staying in state 2. For instance, this time-varying covariate could represent an intervention that increased the probability to move to and stay in a "healthy state". The corresponding slope parameters were $\gamma_{12,z_{it1}} = \gamma_{32,z_{it1}} = 1$ and $\gamma_{13,z_{it1}} = \gamma_{21,z_{it1}} = \gamma_{23,z_{it1}} = \gamma_{31,z_{it1}} = -0.5$. Furthermore, a higher value for z_{i2} increased the probability to transition away from the origin state, leading to a less stable Markov chain. For instance, this stable variable could be a trait-like general stability in response behavior that influences all probabilities to transition away from the state at the previous time-point. The corresponding slope parameters were $\gamma_{12,z_{it2}} = \gamma_{13,z_{it2}} = \gamma_{21,z_{it2}} = \gamma_{23,z_{it2}} = \gamma_{31,z_{it2}} = \gamma_{32,z_{it2}} = 0.5$. The four resulting possibilities for the transition probability matrices were

[31] 1.5 hours pertains to one unit and the other intervals are scaled to this unit interval.

$$\mathbf{P}_{z_{it1}=-.5,z_{it2}=-.5} = \begin{pmatrix} .963 & .012 & .025 \\ .025 & .950 & .025 \\ .025 & .012 & .963 \end{pmatrix}, \quad \mathbf{P}_{z_{it1}=-.5,z_{it2}=.5} = \begin{pmatrix} .940 & .019 & .041 \\ .041 & .919 & .041 \\ .041 & .019 & .940 \end{pmatrix},$$
$$\mathbf{P}_{z_{it1}=.5,z_{it2}=-.5} = \begin{pmatrix} .952 & .032 & .015 \\ .015 & .969 & .015 \\ .015 & .032 & .952 \end{pmatrix}, \quad \mathbf{P}_{z_{it1}=.5,z_{it2}=.5} = \begin{pmatrix} .923 & .052 & .025 \\ .025 & .951 & .025 \\ .025 & .052 & .923 \end{pmatrix}. \tag{4.16}$$

Note that the covariate effects appear to be rather small but they increase for larger intervals than the unit interval.

Regarding the state separation, we used the same conditions as in previous simulation studies evaluating LMFA (Vogelsmeier, Vermunt, Böing-Messing, et al., 2019; Vogelsmeier, Vermunt, van Roekel, et al., 2019). More specifically, we generated data with state-specific MMs as defined in Equation (4.1), assuming orthogonal factors (i.e., $\mathbf{f}_{it} \sim MVN(\mathbf{0}, \mathbf{I})$). To induce the between-state loading differences, we started with a common base matrix in both states:

$$\mathbf{\Lambda}_{Base} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}', \tag{4.17}$$

which shows a binary simple structure that is often found in empirical studies (e.g., consider a typical positive vs. negative affect structure that may also underlie the data in the motivating example described in Section 4.2.2). For the medium loading difference condition, respectively one loading was shifted from the first factor to the second and one from the second to the first (for different items across states). Through this manipulation, the overdeterminaton of the factors was not affected and thus equal across states. For example, the first two of three loading matrices were

$$\mathbf{\Lambda}_1 = \begin{pmatrix} \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}'$$
$$\mathbf{\Lambda}_2 = \begin{pmatrix} 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \tag{4.18}$$

with $\lambda_1 = 0$ and $\lambda_2 = 1$. Similarly, for the low between-state loading difference condition one cross-loading of $\sqrt{.5}$ was added to the first and second factor (for different items across states), which also lowered the primary loadings to $\sqrt{.5}$. Specifically, in the example in Equation (4.18), the entries in $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ were $\lambda_1 = \sqrt{.5}$ and $\lambda_2 = \sqrt{.5}$. Finally, row-wise rescaling of the loading matrices leads to a sum of squares of $1 - e$ per row.

The between-state loading matrix similarity was computed by means of the grand mean, φ_{mean}, of Tucker's (1951) congruence coefficient (i.e., $\varphi_{xy} = \frac{x'y}{\sqrt{x'x}\sqrt{y'y}}$, with x and y referring to matrix columns) that was computed for each pair of factors (note that $\varphi = 1$ means proportionally identical factors). The φ_{mean} across all states and factors was respectively .80 and .94 for the medium and low loading difference condition.

For the intercepts, we used the following base vector with fixed values of 5:

$$\boldsymbol{\nu}_{Base} = (5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)', \qquad (4.19)$$

which was used as such in all states for the no intercept difference condition. To induce low intercept differences across states, we altered two intercepts to 5.5 (different items across the states). For example, for the first two states, the vectors were

$$\begin{aligned} \boldsymbol{\nu}_1 &= (\mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)', \\ \boldsymbol{\nu}_2 &= (5 \quad 5 \quad \mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)'. \end{aligned} \qquad (4.20)$$

The combination of the between-state loading difference and intercept difference conditions lead to four different state-separation conditions. To quantify the general state-separation in both analyses based on the population values, we calculated the four $R^2_{entropy}$ values for step 1 of 3S-LMFA where information is only obtained from the MM and for FIML-LMFA where information is retrieved from both the MM and the SM including the two covariates.[32] For FIML-LMFA, starting from the smallest $R^2_{entropy}$, the resulting values amounted to .90 for the low loading difference/no intercept difference condition, to .94 for the medium loading difference/no intercept difference condition, to .96 for the low loading difference/low intercept difference condition, and to .97 for the medium loading difference/low intercept difference condition. For the same conditions in the first step of 3S-LMFA, these values were respectively equal to .52, .65, .76, and .82. Thus, as expected, state-separation is initially lower in 3S-LMFA than in FIML-LMFA, showing the importance for 3S-LMFA to include the information from the SM in step 3.[33]

[32] The population $R^2_{entropy}$ value for a specific choice of population parameters and number of measurement occasions was obtained using Monte Carlo simulation. For this purpose we used the 'Monte Carlo simulation study' option in LG with one random draw of the time intervals and covariate patterns and with the parameters fixed to their population values.

[33] Note that it is always good to check the $R^2_{entropy}$ after step 1 that is automatically provided in LG because for a very small state-separation, say, with a value much lower than 0.5, it might be better to conduct a FIML-LMFA with additional state-separation information from the SM (including covariates). This is

For each condition, we generated 100 datasets in R (R Core Team, 2020) according to the described population models and analyzed them in LG. Note that only one syntax file is required for FIML-LMFA but two files are necessary for 3S-LMFA. First, one syntax file is required to run step 1 and 2. Thereby, the posterior state assignments and the classification error probability matrix are saved and, subsequently, they are loaded in the second syntax file that is required for step 3.

4.3.3 Results

In the following, we evaluate the performance of 3S-LMFA and compare it to the results of FIML-LMFA based on the replications that converged in both steps of the 3S method as well as in the FIML method. Results that did not converge were re-estimated once and were excluded if convergence still failed. After re-estimation, 3180 out of 3200 datasets converged in 3S- and FIML-LMFA (all datasets converged in step 1 and step 3 of 3S-LMFA and 3180 in FIML-LMFA). Non-convergence in FIML was almost exclusively present for the smallest amount of information condition (i.e., $N = 30$ and $D = 7$) and was caused by reaching the maximum number of EM iterations without convergence. Furthermore, we re-estimated the replications of converged results that showed unrealistically large SEs due to boundary values for any of the estimated initial state and transition parameters (i.e., with an SE > 10, such as 100, 400 or 1000) because including such cases would falsify the results. This was only the case for 56 datasets in the third step of 3S-LMFA where re-estimation did not help. As a result, 3124 datasets were included in the performance analyses reported below.[34]

4.3.3.1 *Goodness of State Recovery*

The recovery of the states was assessed by means of the Rand Index (RI) as well as the Adjusted Rand Index (ARI; Hubert & Arabie, 1985). Both indices evaluate the overlap between two sets of elements while being insensitive to permutations of element labels (in our case state labels). The indices in the RI range from 0 (no overlap between any of the pairs) to 1 (perfect overlap) and the ARI takes values from around 0 (overlap is not better than chance) to 1 (perfect overlap). As expected, the state-recovery was rather poor after the first step of 3S-LMFA because of the low $R^2_{entropy}$ values here ($RI = .83$,

because in that case, the actual differences between the states might be even lower than the estimated ones. This would lead to an underestimation of the classification error (Vermunt, 2010). However, such low values are unlikely to be found in practice.

[34] Note that we also investigated whether the solutions converged to local maxima (i.e., that they had smaller log$L$ values than the global maximum likelihood (ML) solution. Although the latter is unknown, we can obtain an approximation ("proxy") in simulation studies by estimating the models with the population parameters as starting values. When $\log L_{\text{multistart}} < \log L_{\text{proxi}}$, the solution is considered a local maximum. This was no issue in FIML-LMFA and the first step of 3S-LMFA and only occurred for 3 datasets in the third step of 3S-LMFA.

$SD_{RI} = .06$, $ARI = .61$, $SD_{ARI} = .14$). However, the overall recovery in 3S-LMFA was excellent (Steinley, 2004; $RI = .94$, $SD_{RI} = .03$, $ARI = .87$, $SD_{ARI} = .06$) and almost as high as in FIML-LMFA ($RI = .97$, $SD_{RI} = .01$, $ARI = .94$, $SD_{ARI} = .03$). Moreover, only the state-separation influenced the state recovery in that larger separation increased recovery (Table 4.1), indicating that already the minimum sample size of 30 (N) × 7 (D) × 9 (T_{day}) = 1890 was sufficient to estimate $K = 3$ states (and thus about 630 observations per state), which is largely in line with previous results showing that about 500 observations per state are sufficient for similar settings and that a higher amount of information in terms of sample size and observations per subject does not aid recovery once this threshold is reached (Vogelsmeier, Vermunt, van Roekel, et al., 2019).

4.3.3.2 ***Goodness of MM Parameter Recovery***

4.3.3.2.1 *Goodness of Loading Recovery*

We computed a goodness of state loading recovery ($GOSL$) as the average Tucker congruence coefficient between the true and the estimated loading matrices:

$$GOSL = \frac{\sum_{k=1}^{K}\sum_{f=1}^{F}\varphi(\boldsymbol{\Lambda}_k^f, \widehat{\boldsymbol{\Lambda}}_k^f)}{K \times F}, \tag{4.21}$$

where $\boldsymbol{\Lambda}_k^f$ corresponds to the state- and factor-specific loadings. By using Procrustes rotation[35] (Kiers, 1997) in order to rotate the estimated state-specific loading matrices $\widehat{\boldsymbol{\Lambda}}_k$ to the true ones $\boldsymbol{\Lambda}_k$, we solved the label switching of the factor labels within the states. Furthermore, to handle the label switching of the states, we retained the state permutation that maximized the $GOSL$ value. Overall, the loading recovery was very good in 3S-LMFA ($GOSL = 1$; $SD = 0$) and was the same for FIML-LMFA. Note that loading recovery can be good despite a bad state recovery because the loading matrices are very similar across states.

4.3.3.2.2 *Goodness of Intercept Recovery and Unique Variance Recovery*

To examine the recovery of the intercepts and the unique variances, we calculated the mean absolute difference (MAD) between the true and the estimated parameters. The overall intercept recovery in 3S-LMFA was very good (MAD_{int} =0.02; $SD = 0.01$) and did not differ from the recovery in FIML-LMFA. The same applied to the unique variance recovery (MAD_{unique} =0.01; $SD = 0.00$). Moreover, only the amount of information had

[35] Note that the rotation was done in R. Although rotation in LG was already possible for known groups, the issue with switching state labels has to be resolved to provide LG with the correct state-specific target matrices before rotation can be applied to unknown groups such as the states.

a marginal effect on the two types of recovery in that the largest number of subjects ($N = 90$) and a higher number of participation days ($D = 30$) slightly improved the recovery in both analyses (Table 4.1).[36]

4.3.3.3 ***Goodness of SM Parameter Recovery***

4.3.3.3.1 *Goodness of Transition and Initial State Parameter Recovery*

To evaluate the recovery of the transition and initial state parameters, we calculated the average bias and the average Root-Mean-Square-Error (RMSE) for the individual parameters of the four parameter types (i.e., initial state and transition intercept parameters and the two slope parameters for the covariates; Table 4.2). As can be seen, the bias in 3S-LMFA is generally very small (i.e., between -0.02 and 0.01) and in line with FIML-LMFA. However, the RMSE is generally higher in 3S-LMFA (e.g., $RMSE = 0.39$ for the first initial state intercept parameter in 3S-LMFA versus $RMSE = 0.36$ for the same parameter in FIML-LMFA). This is because using the step-wise procedure implies some loss of information. Moreover, Table 4.3 illustrates the effects of the manipulated factors for the four different parameter types, yet, averaged across the individual parameters for the sake of brevity and illustrative purposes. The manipulated factors had an influence on the bias and the RMSE in 3S-LMFA that were similar to the effects on the measures in FIML-LMFA. More specifically, a higher amount of information generally decreased the bias, while a larger state-separation only marginally decreased the bias for some of the individual parameters. Furthermore, a higher state-separation as well as a higher amount of information decreased the RMSE.

[36] The unique variance recovery may be affected by Heywood cases (i.e., improper factor solutions with at least one unique variance being negative or equal to zero, possibly caused by insufficient amount of information or underdetermined factors; Van Driel, 1978). However, this was not the case in any of the analyses.

Table 4.1. Goodness of recovery for the states, loadings, intercepts and unique variances averaged across and conditional on the manipulated factors

		Goodness of Recovery											
		States (ARI)			States RI (RI)			Loadings ($GOSL$)		Intercepts (MAD_{int})		Unique Variances (MAD_{unique})	
		Type of LMFA											
Condition	Factors	3S-1	3S-3	FIML	3S-1	3S-3	FIML	3S	FIML	3S	FIML	3S	FIML
Between-State Loading Difference	low	.56	.85	.93	.80	.93	.97	1	1	.02	.02	.01	.01
	medium	.66	.89	.95	.85	.95	.98	1	1	.02	.02	.01	.01
Between-State Intercept Difference	no	.47	.81	.91	.77	.92	.96	1	1	.02	.02	.01	.01
	low	.74	.92	.97	.89	.97	.98	1	1	.02	.02	.01	.01
Number of Subjects N	30	.61	.87	.94	.83	.94	.97	1	1	.02	.02	.01	.01
	50	.61	.87	.94	.82	.94	.97	1	1	.02	.02	.01	.01
	70	.61	.87	.94	.83	.94	.97	1	1	.02	.02	.01	0
	90	.61	.87	.94	.83	.94	.97	1	1	.01	.01	0	0
Number of	7	.61	.86	.94	.83	.94	.97	1	1	.03	.02	.01	.01
Participation Days D	30	.61	.87	.94	.83	.94	.97	1	1	.01	.01	0	0
All Conditions													
Average		.61	.87	.94	.83	.94	.97	1	1	.02	.02	.01	.01
SD		.14	.06	.03	.06	.03	.01	0	0	.01	.01	0	0

Note. LMFA = latent Markov factor analysis; 3S-1 = three-step step 1; 3S-3 = three-step step 3; FIML = full information maximum likelihood. The perfect loading recoveries result from the loading matrices that are highly similar across the states.

Table 4.2. Parameter bias, RMSE, and SE/SD ratio for all individual parameters averaged across all simulation conditions

Initial state intercept parameters β_{0k}												
	Bias (RMSE)						SE/SD					
β_{0k}	0	0	\	\	\	\	0	0	\	\	\	\
FIML	.01 (.36)	.01 (.36)	\	\	\	\	1.02	1	\	\	\	\
3S	0 (.39)	.01 (.39)	\	\	\	\	1.02	1	\	\	\	\
Transition intercept parameters γ_{0lk}												
	Bias (RMSE)						SE/SD					
γ_{0lk}	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65	-3.65
FIML	-.02 (.18)	-.01 (.18)	-.01 (.17)	-.02 (.16)	-.02 (.18)	-.02 (.18)	1.01	0.98	1	1	0.98	0.99
3S	-.01 (.23)	0 (.22)	-.01 (.21)	-.01 (.21)	-.01 (.22)	-.01 (.24)	1.02	0.98	0.99	0.99	0.99	1
Slope parameters covariate 1 $\gamma_{lk,z_{it1}}$												
	Bias (RMSE)						SE/SD					
$\gamma_{lk,z_{it1}}$	1	-0.5	-0.5	-0.5	-0.5	1	1	-0.5	-0.5	-0.5	-0.5	1
FIML	-.02 (.18)	0 (.17)	0 (.16)	0 (.16)	0 (.17)	-.02 (.18)	1.01	0.99	0.98	1	0.98	1
3S	-.02 (.23)	.01 (.22)	-.01 (.22)	-.01 (.20)	0 (.21)	-.02 (.25)	1	0.98	0.95	0.98	0.96	1
Slope parameters covariate 2 $\gamma_{lk,z_{it2}}$												
	Bias (RMSE)						SE/SD					
$\gamma_{lk,z_{it2}}$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
FIML	0 (.15)	0 (.16)	0 (.16)	-0.01 (.16)	0 (.16)	0 (.15)	0.97	0.99	0.98	0.99	1.01	0.99
3S	0 (.18)	0 (.19)	0 (.20)	-0.02 (.20)	-.01 (.20)	0 (.17)	0.97	0.97	0.96	0.97	0.97	1

Note. FIML = full information maximum likelihood; 3S = three-step.

Table 4.3. Parameter bias, RMSE, and SE/SD for the four types of parameters averaged across and conditional on the manipulated factors

		Initial state intercept parameters			Transition intercept parameters			Slope parameters covariate 1			Slope parameters covariate 2		
		Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD
Condition	Factor	**FIML-LMFA**											
	Average	.01	.36	1.01	-.02	.17	0.99	0	.17	0.99	0	.16	0.99
Between-State Loading	low	.01	.37	1.01	-.02	.18	1	0	.18	0.99	0	.16	0.98
Difference	medium	.01	.36	1.01	-.01	.17	0.99	-.01	.16	0.99	0	.15	1
Between-State Intercept	no	.01	.37	1.01	-.02	.18	0.99	-.01	.18	0.99	0	.16	0.99
Difference	low	.01	.35	1.02	-.02	.17	1	0	.16	1	0	.15	0.99
Number of Subjects *N*	30	.02	.47	1.03	-.03	.24	1	-.01	.23	0.98	-.01	.21	0.97
	50	.02	.36	1.02	-.02	.17	1.01	-.01	.17	1	0	.16	0.98
	70	0	.31	0.99	-.01	.15	0.98	0	.14	0.98	0	.13	1
	90	0	.27	1.02	-.01	.13	0.98	0	.12	1.01	0	.11	0.99
Number of Participation	7	0	.35	1.03	-.03	.23	0.99	-.01	.22	0.99	0	.20	0.98
Days *D*	30	.02	.37	0.99	0	.10	0.99	0	.10	1	0	.09	0.99
Condition	Factor	**3S-LMFA**											
	Average	.01	.39	1.01	-.01	.22	0.99	-.01	.22	0.98	-.01	.19	0.97
Between-State Loading	low	0	.40	1	-.01	.23	1.01	-.01	.23	0.98	-.01	.20	0.96
Difference	medium	.01	.38	1.01	-.01	.21	0.98	-.01	.21	0.98	0	.18	0.98
Between-State Intercept	no	.01	.41	0.99	0	.24	1	-.01	.24	0.97	-.01	.21	0.96
Difference	low	0	.37	1.03	-.01	0.20	0.98	-.01	.20	0.99	0	.17	0.98
Number of Subjects *N*	30	.03	.52	1.01	-.02	.31	1.02	-.02	.31	0.97	-.01	.27	0.95
	50	-.02	.38	1.02	-.01	.22	1	-.01	.22	0.98	-.01	.19	0.97
	70	.02	.34	0.99	0	.18	0.97	-.01	.18	0.97	-.01	.15	0.99
	90	0	.29	1.01	0	.15	0.98	0	.15	1	0	.13	0.98
Number of Participation	7	.02	.38	1.02	-.02	.29	0.99	-.02	.29	0.97	-.01	.25	0.95
Days *D*	30	-.01	.40	0.99	0	.12	1	0	.11	0.99	0	.10	0.99

Note. LMFA = latent Markov factor analysis; FIML = full information maximum likelihood; 3S = three-step.

4.3.3.3.2 *Goodness of Covariates' SE Recovery*

To examine the SE recovery, we compared the average estimated SE for all 100 replications within a condition with the SD of the parameter estimates across these replications and calculated the SE/SD ratios for the individual parameters for the four parameter types (Table 4.2). The ratios are generally slightly lower than 1 in 3S-LMFA with values ranging from 0.95 to 1.02, indicating that the SEs are slightly underestimated. However, this is similar in FIML-LMFA, yet with values ranging from 0.97 to 1.02. Moreover, the manipulated factors had no clear impact on the recovery in neither of the analyses as the four parameter types were influenced differently by a higher state-separation and higher amount of information.

4.3.3.4 ***Computation Time***

Exploring the computation time of all replications in the two analyses, we found that, with 178.01 seconds, FIML-LMFA took on average more than twice as much time as 3S-LMFA, where the total computation time was 82.42 seconds (45.37 seconds for step 1 and 37.05 seconds for step 3). It should be noted that we used 25 random start sets with an EM tolerance of 1e-005 in FIML-LMFA and step 1 and 3 of 3S-LMFA. However, one set and a criterion of 0.01 is probably enough in the third step of 3S-LMFA because local maxima are very unlikely when the measurement part is fixed. Adjusting the values accordingly makes the computation even faster.

4.3.3.5 ***Conclusion***

Summarized, the parameter and SE recovery in 3S-LMFA approached the recovery in FIML-LMFA, making the 3S procedure a promising fast alternative when the inclusion of covariates is of interest and hence the FIML estimation is likely unfeasible. Although a small information loss in terms of higher RMSE values for the parameters of the SM and a slightly worse state-recovery in 3S-LMFA could be observed, the general parameter recovery in 3S-LMFA was on average as good as in FIML-LMFA and furthermore much faster.

4.4 Application

To illustrate the empirical value of the 3S-LMFA approach we applied it to the ESM data introduced in Section 4.2.2. Note that this application is only meant to illustrate the possibilities of the new methodology, and since the hypotheses were not preregistered, we consider these analyses exploratory.[37] As previously described, we investigated

[37] Note that the NA items were generally right-skewed. Since the consequences of violating the normality assumption have yet to be investigated, one should be particularly cautious with drawing substantial

which MMs underlie which part of the data and how the MMs differ (step 1), and whether the MMs are related to covariates (step 3). From all covariates offered in the dataset, we included only five covariates that we thought were plausible to influence MM differences/changes and were of interest for this application. Because emotional experiences may vary depending on situational influences (Dejonckheere, Mestdagh, et al., 2019), and adolescents spend most of their time with parents and friends (Larson, 1983; van Roekel, Scholte, Engels, Goossens, & Verhagen, 2014), we chose the following three time-varying covariates for the social context: (1) being alone (nominal), (2) being with a friend (nominal), and (3) being with a parent (nominal). From the baseline measurement, we chose the following two time-constant covariates: (1) emotion clarity deficit measured with the Emotion Clarity Questionnaire (ECQ; Flynn & Rudolph, 2010) on a Likert scale from 1 (totally disagree) to 5 (totally agree) (e.g., "I often have a hard time understanding how I feel.") and (2) differentiation of emotional experience assessed via a subscale of the Range and Differentiation of Emotional Experience Scale (RDEES; Kang & Shaver, 2004) on a Likert scale from 1 (totally disagree) to 7 (totally agree) (e.g., "I am aware that each emotion has a completely different meaning."). These baseline questionnaires can be found in Appendix K.2 and K.3.[38]

In step 1, we investigate which MMs underlie the data by performing mixture factor analysis including the model selection procedure. Given the relatively small number of observations ($T_i \times I$ = 1168) and items (J = 10), we only considered models with $1-3$ states and $1-3$ factors per state. The best fitting model according to the CHull method was a two-state model with two factors in the first state and one factor in the second state ("[2 1]"). We provide more information about the selection procedure in Appendix K.1. The state separation was very high ($R^2_{entropy} = 0.98$). About 60 percent of the observations were classified into state 1 and 40 percent into state 2. We will inspect the differences between the MMs step by step, starting with (1) the loadings, followed by (2) the intercepts and (3) proportions of unique variances, which are all given in Table 4.4. First, looking at the standardized (and in state 1 obliquely rotated) loadings, we can see that state 1 is characterized by two independent positive affect (PA) and negative affect (NA) factors that are hardly correlated ($r = 0.07$), indicating that adolescents in this state differentiate positive and negative emotional experiences. In contrast, the dimensions seem to collapse in state 2, which is characterized by a single ("bipolar") dimension "PA versus NA". Moreover, it is noticeable that the item "miserable" has a high loading in state 2 but not in state 1. Finally, the rather low loadings of the negative

conclusions (Vogelsmeier, Vermunt, van Roekel, et al., 2019). This is, however, not a problem for illustrating the purpose of 3S-LMFA.

[38] Note that three subjects in the ESM study did not have any baseline measures for an unspecified reason. For such cases, LG automatically imputes the average scale score.

emotions indicate that their relation to the general score on the latent factor is weaker than is the case for the positive emotions.

Second, the intercepts in state 1 are rather high for the positive emotions and very low for the negative emotions. In state 2, the intercepts for the positive emotions are somewhat lower and the intercepts for the negative emotions somewhat higher. Note that, as explained in Section 4.2.3.1, intercept differences pertaining to all items that are strongly related to a certain factor are probably due to differences in the factor means.

Third, investigating the proportions of unique variances, it appears that two of the positive emotions "proud" and "lively" have something unique that cannot be explained by the PA dimension/end of the scale in neither of the two states. Comparing them with the other positive emotions, one can imagine that their scores at least partly depend on specific encountered events (e.g., "proud" may be elicited by achievements and "lively" is more likely to occur during high-energy activities). Moreover, "miserable" has a large unique variance in state 1 and therefore, also considering the low loading, is hardly related to the other emotions. It could be that the item is not always suited to assess affect in adolescents as it is an emotion that is likely triggered by rather extreme situations that might not have been encountered for adolescents in the ESM study. Finally, the negative emotions in state 2 have higher unique variances than in state 1, indicating that there is less covariance between them and that there is no large covariance with the positive emotions.

There is a theoretical debate about whether positive and negative affect are two independent factors (Watson & Tellegen, 1985) or two bipolar ends of the same factor (Russell, 1980). However, our results suggest that both theoretical perspectives can be true at different points in time within one individual. In the first state, adolescents are capable of differentiating positive and negative emotional experiences ("independent state"). In contrast, the factor structure in the second state may be a result of adolescents' simplistic representation of either having "positive" or "negative" emotions ("bipolar state"). These findings are in line with recent research, which suggests that both theoretical perspectives can be true, dependent on person specific factors (e.g., Dejonckheere, Kalokerinos, Bastian, & Kuppens, 2019) or situation specific factors (e.g., Dejonckheere, Mestdagh, et al., 2019). Regarding the intercept differences, we conclude that being in a rather good mood is related to the independent state and being in a rather unpleasant mood is related to the bipolar state. This is in line with research indicating that the bipolar state is more common in individuals with depression (Dejonckheere et al., 2018) and who are stressed (Dejonckheere, Mestdagh, et al., 2019; Zautra, Berkhof, & Nicolson, 2002).

In order to better understand what triggers the different states, we investigated the influence of the five covariates. First, based on the posterior probabilities of the observations to belong to the state-specific MMs, we obtained the modal state membership and the classification errors (Step 2). Given the high state separation, the classification errors were very small:

$$p(\mathbf{w}_{it}|\mathbf{s}_{it}) = \begin{pmatrix} .9968 & .0032 \\ .0080 & .9920 \end{pmatrix}. \quad (4.22)$$

Therefore, correction for classification error is hardly necessary, which generally cannot be foreseen before conducting the step-1 analysis. The modal state assignments were subsequently used as indicators in order to estimate the Markov model with covariates on the transition probabilities (step 3).[39] By means of stepwise backward selection with the five covariates, we eliminated the least significant covariate at each step until only covariates were left that met the criterion of $\alpha < 0.05$. The final model contained the two time-constant covariates from the baseline measure and the time-varying covariate being with a parent. Note that, due to the low classification errors, the final state memberships (i.e., 60% in state 1 and 40% in state 2) did not change after step 1. In Table 4.5, we present the parameters of the SM (including the Wald-test statistics). To see the covariate effect more easily, we also present the transition probabilities for a two-hour-interval, which was the most frequently encountered interval length in the data. We calculated them respectively for the highest and lowest score on one covariate while setting the value of the other covariates to their average value in the sample (averages are given in the notes of Table 4.5). Note that the effect of being with a parent was so small that we do not further discuss it. Regarding the time-constant covariates (emotion clarity deficit and differentiation of emotional experience), we can see that adolescents with high emotion clarity deficit have a slightly higher probability to stay in or transition to the bipolar state (i.e., are more likely to be in that state) compared to adolescents with a low emotion clarity deficit, who are equally likely to be in either of the states. Moreover, adolescents with a high differentiation of emotional experiences have a slightly higher probability to stay in or transition to the differentiated state than adolescents with a low differentiation of emotional experiences, who are equally likely to be in either of the states.

From the individual transition plots of the adolescents (see Figure 4.5 for six representative examples), we can clearly see between-person differences (that are apparently partly related to clarity and differentiation of emotions). For instance, some

[39] Note that we did not add covariates to the initial state probabilities as the number of subjects was rather small.

adolescents are mainly in the independent state (row 1) and others are mainly in the bipolar state (row 2). However, we can also see some adolescents with frequent transitions between the states (right picture in row 3) and some adolescents with transitions after a certain amount of completed questionnaires (left picture row 3). These transitions indicate that there are likely time-varying within-person variables that influence the transitions but that we are not aware of. Therefore, in the future, it would be interesting if applied researchers would include time-varying covariates in their ESM studies (e.g., stress Dejonckheere, Mestdagh, et al., 2019; Zautra et al., 2002) that could potentially influence within-person changes between a bipolar and an independent representation of one's emotional state. To conclude, LMFA indicated that configural invariance was violated across states and that some subjects transitioned between the two states frequently over time while others were mainly in one of the two states. Therefore, the questionnaire data is not validly comparable across all subjects and time-points.

4.5 Discussion

In this chapter, we tailored Vermunt's (2010) maximum likelihood (ML) three-step (3S) procedure to latent Markov factor analysis (LMFA)—a method to explore measurement model (MM) changes over time—and showed that the resulting 3S estimation of LMFA (3S-LMFA) is a promising alternative to the original full information maximum likelihood (FIML) estimation of LMFA (FIML-LMFA): 3S-LMFA performs almost as good as FIML-LMFA, is more accessible and intuitive for applied researchers, and facilitates estimation when researchers want to explore the influence of different (sets of) covariates on transitions between MMs.

It is important to note that the work in this chapter is one of the first to apply a 3S approach with a continuous-time Markov model to time-intensive longitudinal data, which is data that becomes increasingly popular in different fields with diverse data characteristics. On top of that, the flexible step-wise nature of 3S-LMFA can be used to easily extend the method. Specifically, it is easy to adjust the method to the data and research questions at hand by exchanging the first step (i.e., the mixture factor analysis), which makes it applicable to a wide range of data. For example, one may consider extending item response theory models for longitudinal categorical data. If it is not possible to estimate the first step in Latent GOLD (LG), one can also estimate the first step in a different program and only communicate the results to LG to continue with the second and third step. The same will soon be possible in the open-source program R as we are working on a package that takes estimated state probabilities from any step 1 model (estimated in R or another program) as input, calculates modal state assignments and the classification errors, and links it to an existing package that can estimate single indicator

(continuous-time) Markov models with fixed response probabilities. Although adaptations of the MMs are also possible in FIML-LMFA, it is much more difficult in practice since a specific part of the estimation procedure would have to be adapted (i.e., inside the LG software), which is not possible for applied researchers but only for the software programmer.

A limitation of the current chapter is that we did not examine the performance of 3S- and FIML-LMFA under violation of the conditional independence assumption and assumed the covariates to influence only the parameters in the structural model (SM), that is, the transitions in the Markov model, and not the factors or the observed variables directly. This assumption might be violated (e.g., being with friends might be related to higher positive affect) and might lead to extracting a wrong number of states and inaccurate parameter estimates (Kim & Wang, 2017; Kim, Vermunt, Bakk, Jaki, & Van Horn, 2016; Masyn, 2017; Nylund-Gibson & Masyn, 2016). As in any other mixture model approach with covariates, the problem of model misspecifications is inherent to both the FIML and the 3S estimation and should be extensively studied in the specific context of LMFA. With regard to extracting the correct number of states, it can be expected that 3S-LMFA performs better than FIML-LMFA when the effects of the covariates on the latent state memberships are included and direct effects of these covariates—for example, on the response variables—are falsely omitted. In the first analysis step of 3S-LMFA, the MMs are formed while disregarding the covariates. Therefore, the covariates do not affect the state enumeration. This is different in FIML-LMFA where covariates may affect the state enumeration. Specifically, if the local independence assumption is violated, FIML-LMFA would require too many states to counter the local independence violation and achieve a good model fit (Kim & Wang, 2017; Nylund-Gibson & Masyn, 2016). However, inaccurate covariate estimates could occur with both estimation approaches (Asparouhov & Muthén, 2014a; Kim et al., 2016; Masyn, 2017). Therefore, it is important to develop diagnostic tools to detect misspecification (e.g., by means of residual statistics) and to account for it, possibly by including the respective covariates with direct effects on the response variables in step 1 of the analysis and by using covariate-specific classification-error adjustments in step 3 (Vermunt & Magidson, 2020). However, tailoring these methods to LMFA is beyond the scope of this chapter.

Table 4.4. Step 1 results: standardized (for state 1 oblimin rotated) factor loadings, intercepts, and unique variances for the ADAPT dataset

	State 1 (independent state)				State 2 (bipolar state)		
	Factors				Factor		
	PA	NA	Int	Unique V.	PA vs NA	Int	Unique V.
joyful	**0.93**	-0.01	75.98	0.14	**0.96**	53.47	0.07
cheerful	**0.95**	0.00	75.06	0.09	**0.96**	51.90	0.08
lively	**0.55**	0.06	62.08	0.69	**0.61**	43.86	0.62
happy	**0.87**	0.08	76.50	0.23	**0.89**	51.75	0.20
proud	**0.53**	0.13	61.43	0.69	**0.66**	42.10	0.57
miserable	-0.22	0.11	5.38	0.96	**-0.61**	26.53	0.63
mad	0.01	**0.94**	1.19	0.16	**-0.34**	14.68	0.88
afraid	0.02	**0.94**	1.14	0.17	**-0.43**	17.03	0.81
scared	0.04	**0.93**	1.11	0.13	**-0.38**	11.46	0.85
sad	0.04	**0.92**	1.20	0.19	**-0.56**	18.49	0.69

Notes. Int. = Intercepts; V. = Variance; Factor loadings were standardized by dividing them by the state-specific item standard deviations. We considered the loadings to be considerable when they were larger than 0.3 in absolute value (e.g., Hair et al., 2014). These loadings are depicted in boldface.

Table 4.5. Logit and log intensity parameters for the structural model and additional transition probabilities respectively for the lowest and highest possible score on the three covariates for a two-hour interval

parameter	coef	s.e.	Wald	df	p-value
β_{02}	0.4043	0.3951	1.0470	1	0.31
γ_{012}	-6.3343	1.2503	25.9605	2	< 0.01
γ_{021}	-4.5298	1.2851			
$\gamma_{12,P}$	0.5214	0.2409	6.0807	2	0.048
$\gamma_{21,P}$	0.1533	0.2312			
$\gamma_{12,ECD}$	0.8274	0.2006	42.3336	2	< 0.01
$\gamma_{21,ECD}$	-0.0534	0.1994			
$\gamma_{12,D}$	0.4932	0.2183	8.9022	2	0.012
$\gamma_{21,D}$	0.6761	0.2268			

Transition Probabilities	
low covariate score	high covariate score
$P_0 = \begin{pmatrix} .82 & .18 \\ .28 & .72 \end{pmatrix}$	$P_1 = \begin{pmatrix} .72 & .28 \\ .29 & .71 \end{pmatrix}$
$ECD_1 = \begin{pmatrix} .93 & .07 \\ .33 & .67 \end{pmatrix}$	$ECD_5 = \begin{pmatrix} .19 & .81 \\ .12 & .88 \end{pmatrix}$
$D_1 = \begin{pmatrix} .95 & .05 \\ .04 & .96 \end{pmatrix}$	$D_7 = \begin{pmatrix} .70 & .30 \\ .66 & .34 \end{pmatrix}$

Notes. We present four decimals as this is the default in Latent GOLD. P = being with parent; EDC = emotion clarity deficit; D = differentiation of emotions; the probabilities were calculated by setting the covariate of interest to the lowest or highest score and the other two on their averages; the average scores were 0.33 for P, 2.48 for EDC and 4.39 for D; the overall model was significant with Wald (6) = 68.43; 8.6e-13 and thus outperformed the intercept only model.

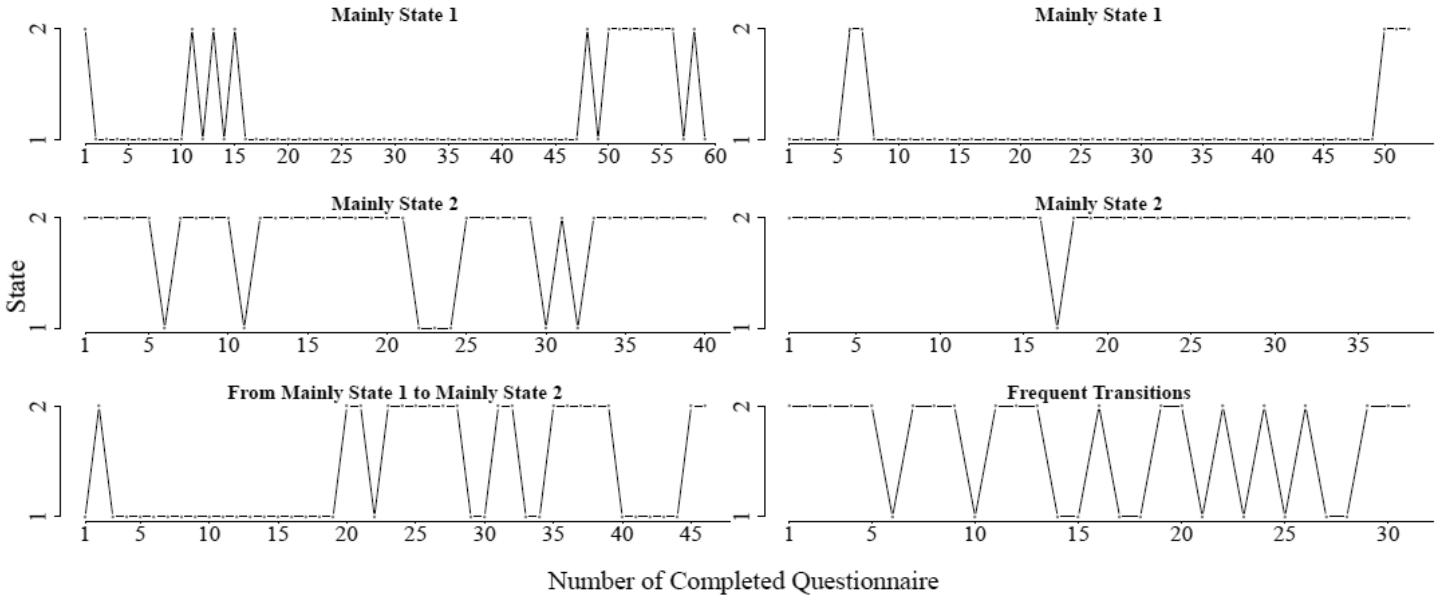


Figure 4.5. Six examples of adolescents' transition plots that are representative for the whole sample. Note that the scale does not consider the time interval between the observations to enable the illustration.

5

Latent Markov Latent Trait Analysis for Categorical Data

This chapter is published as Vogelsmeier, L. V. D. E., Vermunt, J. K., Keijsers, L., & De Roover, K. (2021). Latent Markov latent trait analysis for exploring measurement model changes in intensive longitudinal data. *Evaluation & the Health Professions*, *44*, 61–76. doi:10.1177/0163278720976762

Abstract

Drawing inferences about dynamics of psychological constructs from intensive longitudinal data requires the measurement model (MM)—indicating how items relate to constructs—to be invariant across subjects and time-points. When assessing subjects in their daily life, however, there may be multiple MMs, for instance, because subjects differ in their item interpretation or because the response style of (some) subjects changes over time. The recently proposed "latent Markov factor analysis" (LMFA) evaluates (violations of) measurement invariance by classifying observations into latent "states" according to the MM underlying these observations such that MMs differ between states but are invariant within one state. However, LMFA is limited to normally distributed continuous data and estimates may be inaccurate when applying the method to ordinal data (e.g., from Likert items) with skewed responses or few response categories. To enable researchers and health professionals with ordinal data to evaluate measurement invariance, we present "latent Markov latent trait analysis" (LMLTA), which builds upon LMFA but treats responses as ordinal. Our application shows differences in MMs of adolescents' affective well-being in different social contexts, highlighting the importance of studying measurement invariance for drawing accurate inferences for psychological science and practice and for further understanding dynamics of psychological constructs.

5.1 Introduction

Intensive longitudinal data (ILD; e.g., Hamaker & Wichers, 2017) allow one to investigate the dynamics over time of latent (i.e., unobservable) psychological constructs. By frequently gathering data (say at more than 50 measurement occasions) of multiple subjects, new insights regarding subject-specific dynamics can be obtained, which have clinical implications. For instance, studies are being conducted on dynamics in emotions and behaviors related to mental health (e.g., Myin-Germeys et al., 2018; Snippe et al., 2016), and ILD can also be used to tailor interventions to the subject's real-time dynamics of negative affect (van Roekel et al, 2017). Such data is efficiently gathered by means of experience sampling methodology (ESM; Scollon et al., 2003), in which subjects repeatedly rate questionnaire items over several weeks, say five times a day, at randomized time-points. The recent steep increase in such datasets (e.g., Hamaker & Wichers, 2017; van Roekel et al., 2019) is related to novel technologies to efficiently gather these data with the use of smartphone apps. Hence, there is an urgent need to also develop novel analytical methods.

In order to draw valid inferences about the measured constructs, either for scientific or clinical purposes, it is crucial that the measurement model (MM) is invariant (i.e., constant) across time and subjects (i.e., having within- and between-person invariance). The MM indicates to what extent the latent constructs (or "factors") are measured by which items, as indicated by the "factor loadings". For continuous data, the MM is obtained by factor analysis (FA; Lawley & Maxwell, 1962). If measurement invariance (MI) holds, the constructs are conceptually equal and thus comparable across subjects and over time. Often, MI is not tenable because response styles, substantive changes in item interpretation, or changes in the nature of the measured construct may affect the MM. That is, people may differ from each other in their MMs, for instance, depending on psychopathology, but one subject may also differ over time in its own MM, for instance, depending on the social context in which the questionnaire is filled in. When the non-invariance patterns are undetected or ignored, they cause a potential threat to valid inferences using standard methods for comparing factor means across time and subjects. For instance, changes in subjects' overall emotional well-being may be (partly) due to changes in how subjects interpret the items. Changes in the MM are also important phenomena in their own right. For instance, detecting MM changes is crucial for valid decisions about treatment allocation over time and such changes may even signal the onset of a mental episode. Consider, for example, a psychologist who measures positive affect (PA) and negative affect (NA) in patients with a bipolar disorder. Patients in manic episodes often encounter high arousal PA, such as feeling energetic or excited, together with high arousal NA, such as being irritated or distracted (American Psychiatric

Association, 2013). This might result in a MM with one bipolar "arousal" factor contrasting "low" versus "high" arousal. When patients encounter depressive episodes, PA is generally lower and NA at least somewhat higher (Hamaker, Grasman, & Kamphuis, 2010), which might correspond to a MM with two separate PA and NA affect factors or one bipolar "valence of affect" factor. Assessing MI thus allows for more accurate conclusions, but may also open up novel possibilities of early detection of subtle changes in daily functioning.

In order to assess for whom and when a MM applies, Vogelsmeier, Vermunt, van Roekel, et al. (2019) developed a novel method called latent Markov factor analysis (LMFA) for tracking and diagnosing MM changes for continuous responses in ILD. LMFA combines a latent Markov model (LMM; Bartolucci et al., 2014; Collins & Lanza, 2010) with mixture FA (McLachlan & Peel, 2000; McNicholas, 2016): The LMM clusters subject- and time-point-specific observations into a few dynamic latent classes or "states" according to the MMs underlying these observations and mixture FA evaluates which MM applies for each state. Thus, every state pertains to a different MM and the MM is invariant within one state. Note that not all MMs may apply to each subject. Some subjects may constantly stay in one state while others may transition between different states. By investigating the state memberships, one can see which subjects and measurements are comparable regarding their underlying MM. Investigating the state-specific MMs offers insights into the underlying dynamics and it also helps researchers make decisions about subsequent analyses. For example, when at least "partial" invariance holds across states (i.e., only a few MM parameters differ; Byrne et al., 1989), researchers could study discrete changes in factor means by repeating the LMLTA analysis, restricting invariant MM parameters to be equal across states, and adding factor means to the model.

The new method has raised awareness of possible MM changes in ILD among fundamental and applied researchers who are now eager to evaluate which MM applies to which subject at which time-point (Horstmann & Ziegler, 2020). However, an important limitation of LMFA is the assumption of having normally distributed continuous item responses. This assumption is often violated in ILD. Although continuous items are sometimes used (e.g., participants are asked to give their answer by sliding on the Visual Analog Scale from 0 ("not at all") to 100 ("very much"), many studies use multiple Likert items with 5 to 7 categories for their assessment. Even though it has been shown that items with 5 or more categories might be treated as continuous (Dolan, 1994), it becomes problematic if the item response distributions are heavily skewed (e.g., when most responses have a 0 score, which is quite common with less frequent thoughts, emotions, or behaviors). FA is not robust against strong deviations from normality and, therefore, may yield inaccurately estimated parameters (Kappenburg -ten Holt, 2014; Rhemtulla et al., 2012; Vermunt & Magidson, 2005). Note that the same problem

generally applies to studies that use ordinal items with less than 5 categories, although this is less common in ILD data. If the normal approximation is clearly incorrect, a better alternative is to treat the items as ordinal and to specify the probability of responding in a certain category by means of "item response theory" or "latent trait" (LT) models, where "trait" refers to a latent construct in the psychometric literature (Vermunt & Magidson, 2016).

The aim of this chapter is to combine the strength of LT models to adequately deal with ordinal data with the strengths of LMFA to trace complex measurement non-invariance patterns in the data. The novel and much-needed latent Markov latent trait analysis (LMLTA) for ordinal data is obtained by replacing the mixture FA by a mixture multidimensional version of Muraki's (1992) "generalized partial credit model" (GPCM) that treats the responses as ordinal. Section 5.2 describes LMLTA and how it compares to LMFA. Section 5.3 illustrates the empirical value of LMLTA to detect MM changes in ordinal data on adolescents' well-being in different social contexts. Finally, Section 5.4 concludes with some points of discussion and future directions of research.

5.2 Method

5.2.1 Data Structure

In LMLTA, we assume intensive longitudinal observations that are nested within subjects and we assume multiple Likert and, therefore, ordinal items with response categories ranging, for instance, from 1 = "strongly disagree" to 5 = "strongly agree". The latter differs from LMFA, where the items are assumed to be continuous variables. The observations are denoted by y_{ijt} with $i = 1, \dots, I$ referring to subjects, $j = 1, \dots, J$ referring to items, and $t = 1, \dots, T$ referring to time-points. Furthermore, $g = 1, \dots, G$ refers to the item categories and the number of categories G is assumed to be constant across items. Finally, the number of time-points T typically differs across subjects but, for simplicity, we mostly omit the index i in T_i. The observations are collected in the $1 \times J$ vectors $\mathbf{y}_{it} = (y_{i1t}, \dots, y_{iJt})$ that are collected in the $T \times J$ subject-specific data matrices $\mathbf{Y}_i = (\mathbf{y}_{i1}', \dots, \mathbf{y}_{iT}')'$. The data matrices are concatenated in the dataset $\mathbf{Y} = (\mathbf{Y}_1', \dots, \mathbf{Y}_I')'$ with $\sum_{i=1}^{I} T_i$ rows.

5.2.2 Latent Markov Latent Trait Analysis

In LMLTA, just as in LMFA, a LMM specifies transitions between discrete latent states (e.g., manic and depressive state) characterized by state-specific MMs (e.g., state 1 contains one arousal factor and state 2 two affect factors). A LMM is basically a latent class model (Lazarsfeld & Henry, 1968) and thus a method to find unobserved classes of observations with comparable response patterns. A LMM allows subjects to transition

between latent classes over time, which is why the classes are called "states". To get more insight into what possibly predicts state memberships, one may explore the relation between the state memberships and time-varying or time-constant explanatory variables or "covariates". For instance, sleep quality and disruptions in the daily routine may increase the probability to transition to a manic state (Hamaker et al., 2010). The state-specific MMs are latent variable models that indicate which latent constructs are measured by which items and to what extent. The choice for the type of latent variable model directly follows from the assumed item response distribution: An LT model for ordinal data is used in LMLTA and a FA model for continuous data is used in LMFA.

The parameters in LMLTA can be estimated with the same approaches as in LMFA, using Latent GOLD (LG; Vermunt & Magidson, 2016) syntax. The first approach is a one-step full information maximum likelihood (FIML) estimation (Vogelsmeier, Vermunt, van Roekel, et al., 2019) and the second approach is a three-step (3S) procedure that splits the estimation of the LMM and the state-specific MMs (Vogelsmeier, Vermunt, Bülow, & De Roover, 2020). The latter approach has advantages, especially regarding model selection with covariates. In the following, we first describe the LMM and then introduce the particular LT model applied in this chapter and compare it to the FA model in LMFA. Thereafter, we discuss the two possible estimation procedures and the advantages of the 3S estimation.

5.2.2.1 *Latent Markov Model*

The LMM is a probabilistic model with two assumptions (e.g., Bartolucci et al., 2014; Collins & Lanza, 2010): (1) The probability of being in state k (with $k = 1, \dots, K$) at time-point t depends only on the state membership at the previous time-point $t - 1$ and not on any other state memberships (first-order Markov assumption) and (2) the responses $\mathbf{y}_{it}$ at time-point t depend only on the state membership at this time-point (local independence assumption). The sequence of states is called a latent Markov chain (LMC). Figure 5.1a illustrates a LMC for a single subject: The $K \times 1$ vectors $\mathbf{s}_{it} = (s_{it1}, \dots, s_{itK})'$ contain the binary indicators s_{itk} that are equal to 1 for state k and equal to 0 for all other states. They determine the state membership at time-point t. The $U \times 1$ vectors $\mathbf{z}_{it} = (z_{it1}, \dots, z_{itU})'$ contain the covariate values z_{itu}, with $u = 1, \dots, U$ referring to the subject- and possibly time-point-specific covariates influencing the state memberships. In Figure 5.1a, state 1 (e.g., the manic state) applies to time-points 1–29 and 55–56, while state 2 (e.g., the depressive state) applies to time-points 30–54.

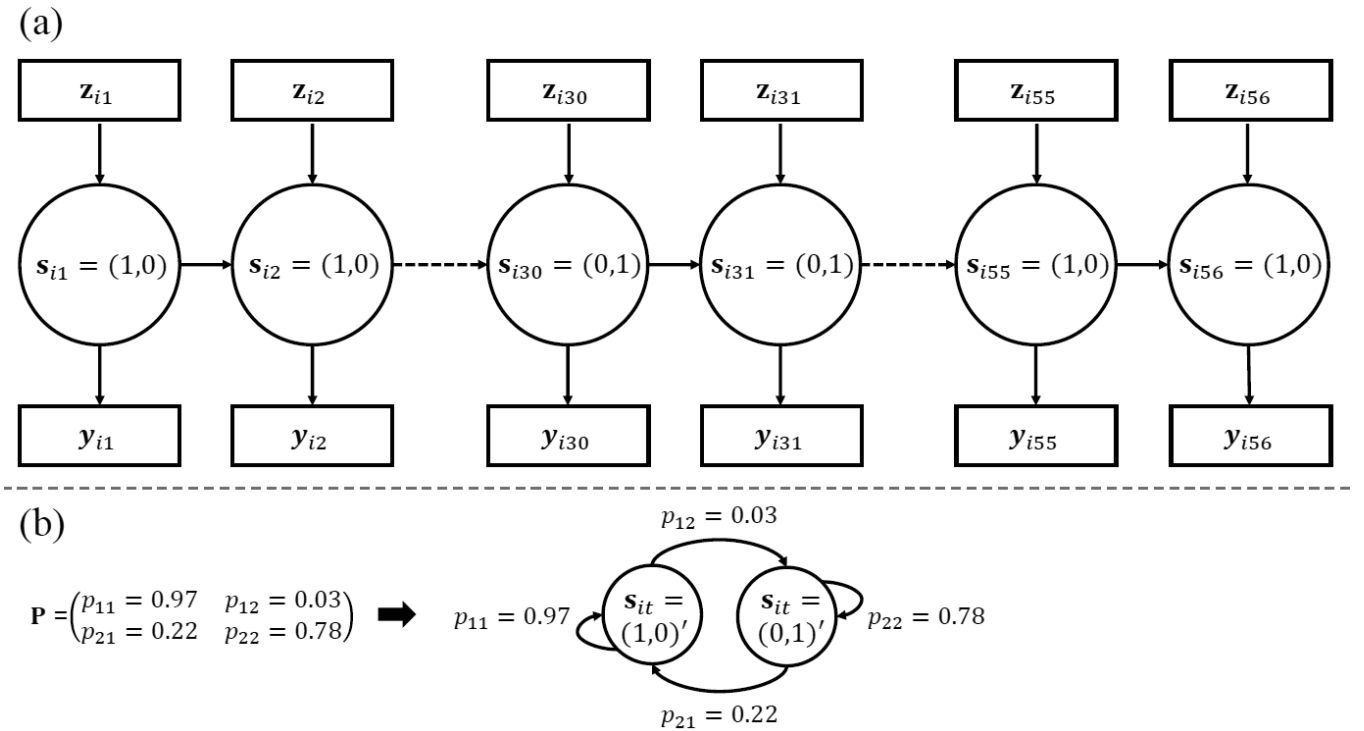


Figure 5.1. Part a) is a graphical illustration of a latent Markov chain from the latent Markov latent trait analysis model. The binary vectors $\mathbf{s}_{it} = (s_{it1} = 1, s_{it2} = 0)' = (1,0)'$ indicate the state memberships at different time-points t, implying that the subject is in state $k = 1$ at time-points 1–29 and 55–56 and in state $k = 2$ at time-points 30–54, implying transitions from state 1 to state 2 at time-point 30 and from state 2 to state 1 at time-point 55. Note that the responses $\mathbf{y}_{it}$ are determined by state-specific latent trait measurement models. Furthermore, the covariates $\mathbf{z}_{it}$ may influence the state memberships $\mathbf{s}_{it}$. Part b) shows a possible transition probability matrix $\mathbf{P}$ for the two states and its corresponding transition diagram that shows how to read the matrix. The diagonal elements correspond to the probabilities to stay in a state and the off-diagonal elements correspond to the transitions away from a state.

A LMM is characterized by the "initial state", "transition", and "response" probabilities. Together, the parameters form the joint distribution of the observations and states. This is:

$$p(\mathbf{Y}_i, \mathbf{S}_i|\mathbf{Z}_i) = p\big(\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT}, \mathbf{s}_{i1}, \dots, \mathbf{s}_{iT}|\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}\big) = \overbrace{p(\mathbf{s}_{i1}|\mathbf{z}_{i1})}^{\substack{initial\ state \\ probabilities}} \prod_{t=2}^{T} \overbrace{p_{\delta_{ti}}(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it})}^{\substack{transition \\ probabilities}} \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_{it}|\mathbf{s}_{it})}^{\substack{response \\ probabilities}} \tag{5.1}$$

for subject i. The initial state and transition probabilities may depend on subject- and time-point-specific covariates $\mathbf{z}_{it}$ but, in the following, we will omit an index z for simplicity. The initial state probabilities in Equation (5.1) define the probabilities to start in state k at time-point $t = 1$ and are collected in a $K \times 1$ probability vector $\boldsymbol{\pi}$ with elements $\pi_k = p(s_{i1k} = 1|\mathbf{z}_{i1})$ and $\sum_{k=1}^{K} \pi_k = 1$. In LG, the initial state probabilities are modeled via a logit model as this prevents parameter range restrictions and the covariates also enter through this parameterization as:

$$\log \frac{p(s_{i1k} = 1|\mathbf{z}_{i1})}{p(s_{i11} = 1|\mathbf{z}_{i1})} = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{z}_{it=1} \tag{5.2}$$

for $k = 2, \dots, K$ and with $k = 1$ as the reference category. Here, the initial state intercepts are denoted by β_{0k} and the initial state slopes that quantify the effect of the covariates on the initial state memberships are captured by the vectors $\boldsymbol{\beta}_k' = (\beta_{k,z_{i11}}, \dots, \beta_{k,z_{i1U}})'$.

Transition probabilities are the probabilities to be in state k at time-point $t > 1$ conditional on state l ($l = 1, \dots, K$) at $t - 1$. In a discrete-time- (DT-)LMM, intervals between measurements, δ_{ti}, are assumed to be equal. A continuous-time- (CT-)LMM (Böckenholt, 2005; Jackson & Sharples, 2002; Vogelsmeier, Vermunt, Böing-Messing, et al., 2019) allows the intervals to differ across time-points and subjects, which is often more realistic in ESM studies and therefore applied throughout the rest of this chapter. The transition probabilities $p_{\delta_{ti},lk} = p_{\delta_{ti}}(s_{itk} = 1|s_{it-1,l} = 1, \mathbf{z}_{it})$ are collected in the $K \times K$ matrix $\mathbf{P}_{\delta_{ti}}$, where the row sums of $\mathbf{P}_{\delta_{ti}}$, $\sum_{k=1}^{K} p_{\delta_{ti},lk}$, are equal to 1. In a DT-LMM, a multinomial logistic model is used for the transition probabilities:

$$\log \frac{p(s_{itk} = 1|s_{it-1,l} = 1, \mathbf{z}_{it})}{p(s_{itl} = 1|s_{it-1,l} = 1, \mathbf{z}_{it})} = \gamma_{0lk} + \boldsymbol{\gamma}_{lk}' \mathbf{z}_{it} \tag{5.3}$$

with $k \neq l$, γ_{0lk} as transition intercepts, and $\boldsymbol{\gamma}'_{lk} = (\gamma_{lk,z_{i11}}, \ldots, \gamma_{lk,z_{i1U}})'$ as slopes that quantify the covariate effects on transitioning to another state compared to staying in a state. In Figure 5.1b, we show how to read a transition probability matrix. The diagonal elements indicate that the probability of staying in state 1 is higher than of staying in state 2. If state 1 is the manic and state 2 the depressive state, we would conclude that the manic state is more persistent for this person.

In the CT-LMM, the transition probabilities themselves are a function of the interval δ_{ti} and the "transition intensity matrix" $\mathbf{Q}$. The $K \times K$ matrix $\mathbf{Q}$ contains the transition intensities (or rates) q_{lk} that define the transitions from the origin state l to the destination state k per very small time unit. For all off-diagonal elements in the matrix $\mathbf{Q}$ (i.e., $k \neq l$) the intensities are:

$$q_{lk} = \lim_{\delta \to 0} \frac{p(s_{itk} = 1 | s_{it-\delta,l} = 1, \mathbf{z}_{it})}{\delta}. \tag{5.4}$$

The diagonal elements are equal to $-\sum_{k \neq l} q_{lk}$ (Cox & Miller, 1965). The transition probabilities $\mathbf{P}_{\delta_{ti}}$ are obtained by taking the matrix exponential of $\mathbf{Q} \times \delta_{ti}$. This implies that the probability to transition to another state at two consecutive measurement occasions (i.e., $k \neq l$) becomes increasingly more likely for larger intervals. As can be seen from Equation (5.4), one may also regress the transition intensities on covariates $\mathbf{z}_{it}$ to better understand what may cause the transitions to or away from a state. In the CT-LMM, LG uses a log-linear model for the transition intensities and the covariates are included as follows (again for $k \neq l$):

$$\log q_{lk} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk} \mathbf{z}_{it}. \tag{5.5}$$

Hence, covariates to predict any of the parameters (i.e., initial state and transition probabilities or intensities) are included by means of regression, as is usually done in LMMs (e.g., Bartolucci et al., 2014; Vermunt et al., 1999; Visser et al., 2009).

Instead of using only observed covariates in any of the parameters, one may also use a time-constant or time-varying latent categorical variable that classifies subjects according to their transition pattern or initial state probabilities into latent classes (Crayen et al., 2017; Vermunt, Tran, & Magidson, 2008). This "mixture (CT-)LMM" captures the most relevant between-subject differences in the transition process. The number of latent classes can be based on theory and interpretability or selected using information criteria, such as the Bayesian information criterion (BIC, Schwarz, 1978) or

the convex hull (CHull; Ceulemans & Kiers, 2006) method. An example is shown in the application (Section 5.3).

Finally, the response probabilities $p(\mathbf{y}_{it}|s_{itk}=1)$ indicate the probability for a certain response pattern at time-point t, given the state membership at that time-point, $s_{itk}=1$. These response probabilities depend on the state-specific MMs described next.

5.2.2.2 *Measurement Model*

The MMs determine how the responses y_{itj} are defined by the state memberships $s_{itk}=1$. To this end, a latent variable model with state-specific parameters is used in both LMFA and LMLTA. For both methods, it holds that: (1) the responses y_{itj} are indicators of underlying latent factors $\mathbf{f}_{it}$, (2) the factors are considered to be normally distributed continuous variables, (3) the responses y_{ijt} are independent given the latent factors, and (4) covariates are only indirectly related to the observed responses via the latent states. As explained before, LMFA and LMLTA differ in the type of latent variable model that is used. In LMFA, the continuous responses y_{ijt} are defined by state-specific linear FA models with parameters that may differ across the latent states. For a single item j this is given by (e.g., McLachlan & Peel, 2000):

$$E\left(y_{ijt}|\mathbf{f}_{it}, s_{itk}=1\right)=\sum_{r=1}^{R_k}\lambda_{jrk}\,f_{rit}+\,\nu_{jk}, \tag{5.6}$$

where R_k is the state-specific number of factors, $r=1,\dots,R_k$ indicates a state-specific factor, λ_{jrk} is a state-specific loading of item j on factor r, $\mathbf{f}_{it}=(f_{1it},\dots,f_{Rit})'$ are subject- and time-point-specific factor scores with $\mathbf{f}_{it}\sim MVN(0,\mathbf{\Phi}_k)$ (note that possible restrictions of $\mathbf{\Phi}_k$ will be discussed further below), and ν_{jk} indicates a state-specific intercept for item j.

In LMLTA, the ordinal responses y_{ijt} are defined by state-specific LT models. It is important to note that there are several LT models that could be used to model Likert-type data (Andrich, 1978; Muraki, 1992; Samejima, 1969). The GPCM (Muraki, 1992) is a relatively flexible and unrestrictive model (Tijmstra, Bolsinova, & Jeon, 2018) and is therefore considered in this study. More specifically, we use the multidimensional version of the GPCM (e.g., Johnson & Bolt, 2010) and, in order to allow for parameter differences across states, we employ a mixture variant (for previous work on mixture LT models see, e.g., Bolt, Cohen, & Wollack, 2001; Cohen & Bolt, 2005; Rost, 1990; Smit, Kelderman, & van der Flier, 2000). In contrast to the state-specific FA models in LMFA, the state-specific GPCMs used in LMLTA do not consist of a set of linear models but of a

set of adjacent-category (i.e., $(g, g\ +1)$) ordinal logit models. More specifically, using as much as possible the same notation as before, the logarithm of the odds of responding in category $g + 1$ instead of responding in category g for item j, given the factor scores $\mathbf{f}_{it}$ and the state membership $s_{itk} = 1$ for subject i at time-point t, has the following linear form:

$$\log\left(\frac{p\left(y_{ijtg+1} = 1|\mathbf{f}_{it}, s_{itk} = 1\right)}{p\left(y_{ijyg} = 1|\mathbf{f}_{it}, s_{itk} = 1\right)}\right) = \sum_{r=1}^{R_k} \lambda_{jrk}\, f_{rit} + \nu^*_{jgk}, \tag{5.7}$$

for $1 \leq g \leq G - 1$, with $y_{ijt} = g$ indicating that this response to item j is in category g. Again, λ_{jrk} is the state-specific loading of item j on factor r. The ν^*_{jgk} are the $G - 1$ intercepts for each of the adjacent-category log-odds. The logistic model for the probability of response g equals:

$$p\left(y_{ijt} = g|\mathbf{f}_{it}, s_{itk} = 1\right) = \frac{\exp\left(\sum_{r=1}^{R_k} g \times \lambda_{jrk} f_{rit} + \nu_{jgk}\right)}{\sum_{g\prime=1}^{G} \exp\left(\sum_{r=1}^{R_k} g' \times \lambda_{jrk} f_{rit} + \nu_{jg\prime k}\right)}. \tag{5.8}$$

As shown, the loadings are multiplied with the category number and the intercepts are now ν_{jgk}, with $\sum_{g=1}^{G} \nu_{jgk} = 0$. The relation between the two sets of intercepts is that $\nu^*_{jgk} = \nu_{jg+1,k} - \nu_{jgk}$.

When comparing Equation (5.6) and Equation (5.7), the loading parameters for the FA model and the GPCM are clearly conceptually similar. In both cases, they indicate how strongly an item j measures a latent factor f_{rit} in state k (Kankaraš, Vermunt, & Moors, 2011). In contrast, the intercepts are not directly comparable across the two models. In the FA model, there is only one intercept per item and state, ν_{jk}, because the responses are treated continuous. For the ordinal responses in the GPCM, there are $G - 1$ free intercept parameters per state, ν^*_{jgk}.

As in LMFA, the state-specific joint response probabilities for LMLTA at time-point t are obtained by marginalizing over the latent factors. Moreover, the J item responses are assumed to be conditionally independent given the latent factors and the state membership. Therefore, the response probabilities are (e.g., Johnson & Bolt, 2010):

$$p(\mathbf{y}_{it}|s_{itk}=1)=\int\dots\int p(\mathbf{f}_{it};\mathbf{0},\mathbf{\Phi}_k)\prod_{j=1}^{J}p\big(y_{ijt}=g|\mathbf{f}_{it},s_{itk}=1\big)\,d\mathbf{f}_{it} \qquad (5.9)$$

with $p\big(y_{itj}=g|\mathbf{f}_{it},s_{itk}=1\big)$ as in Equation (5.8) and $p(\mathbf{f}_{it};\mathbf{0},\mathbf{\Phi}_k)$ denoting the probability density function of the multivariate normal distribution with a mean vector of zero's and covariance matrices $\mathbf{\Phi}_k$.

To enable the exploration of all kinds of MM changes, including the number and nature of the factors, an exploratory model is used in both methods. In contrast to a confirmatory model—in which certain factor loadings are assumed to be absent and therefore, set to zero—an exploratory model estimates all loadings.[40] However, both models are unidentified without further constraints. To partially identify the models and set a scale to the R_k factors, one may restrict the factor means to zero and the factor (co)variances $\mathbf{\Phi}_k$ to equal an identity matrix, which implies normalized and uncorrelated factors. Alternatively, it is possible to freely estimate the covariance matrix of the factors and instead fix one loading for each of the R_k factors to 1 and one extra loading per estimated correlation to 0 (e.g., for a state with $R_k=2$, two loadings would be fixed to 1 and one loading would be fixed to 0). Remaining rotational freedom in the FA model can be dealt with by means of rotation criteria that optimize the simple structure and/or between-state-agreement of the factor loadings (Clarkson & Jennrich, 1988; De Roover & Vermunt, 2019; Kiers, 1997). The identification of the GPCM is more intricate: Despite the model being identified by the constraints imposed so far, one might obtain strongly related parameter estimates and large standard errors. In order to prevent this so-called "empirical underidentification", R_k-1 (additional) loadings of different items have to be fixed to 0 in each state (Skrondal & Rabe-Hesketh, 2011).[41]

As becomes apparent from Equation (5.6) and Equation (5.7), in either model, the state-specific MMs can differ in terms of the number of factors, the loadings, the intercepts, and the factor covariance matrices. However, there is an important difference between the two methods. In LMFA, states may also differ regarding unique variances, say ψ_{kj}, which is variance that is not accounted for by the latent factors. This is possible because the error term in a FA model is assumed to be normally distributed, that is, $e_{ijt}\sim N(0,\psi_{kj})$. In contrast, in the GPCM, the variance of the error is not a free parameter but fixed to the value of the variance of the standard logistic distribution, $\pi^2/3$, and

[40] If desired, however, a confirmatory model may also be used in both LMFA and LMLTA.
[41] Note that these constraints could also be used to solve rotational freedom in the FA model (Vermunt & Magidson, 2016).

hence, in LMLTA, also equal across states. Note that, in the GPCM, fixing the error variance is necessary to identify the model (Long, 1997).[42] Although it might be possible to account for error variance heterogeneity by tailoring "scale adjustment" methods (Magidson & Vermunt, 2007) to LMLTA, this is beyond the scope of this chapter.

Besides this difference, MI analyses with FA and LT models are similar as their primary concern is to detect parameter differences. However, different words may be used to describe (non-)invariance. When using a LT model, researchers typically specify the lack of invariance, which is called "differential item functioning" (DIF). More specifically, "uniform DIF" is present when only intercepts differ, in our case across latent states, and "non-uniform DIF" is present when loadings differ across states, whether intercepts are equal or not (Bauer, 2017). In contrast, when using a FA model, researchers typically specify which level of invariance has been reached, starting from an invariant number of factors and pattern of zero loadings, followed by invariant loadings, intercepts, and finally unique variances (Meredith, 1993). In the next paragraph, we will describe how to obtain the estimates that are used to investigate the level of invariance in LMLTA.

5.2.2.3 *Maximum Likelihood Estimation*

The parameters in LMLTA are obtained with maximum likelihood (ML) estimation. One may choose between (1) the one-step FIML estimation and (2) the 3S estimation, just as is the case for LMFA. However, estimating the LMLTA model with either approach is computationally more complex than estimating the LMFA model. Therefore, LMLTA is limited regarding the number of factors that can be estimated (i.e., including more than 3 factors is usually unfeasible; see Appendix L.1 for detailed explanations). First, for the FIML estimation (Vogelsmeier, Vermunt, van Roekel, et al., 2019), the following loglikelihood function, derived from the joint distribution in Equation (5.1), has to be maximized:

$$\log L_{FIML} = \sum_{i=1}^{I} \log \left(\sum_{\mathbf{s}_{i1}} \dots \sum_{\mathbf{s}_{iT}} p(\mathbf{Y}_i, \mathbf{S}_i | \mathbf{Z}_i) \right). \tag{5.10}$$

In LG, the ML estimates are obtained with the forward-backward algorithm (Baum et al., 1970), which is an efficient version of the expectation maximization algorithm (Dempster

[42] Note that this is generally a limitation, also in other LT models, and it is often ignored. However, it is important to understand that possible differences in error variances across states will be captured as loading and intercept differences (Long, 1997). For instance, when in one state the error variance is two times larger than in the other state, the loadings and intercepts in that state will be $\sqrt{2}$ times smaller than in the other state.

et al., 1977), tailored to LMMs. Additionally, in the maximization-step, a Fisher algorithm is used to update the log-intensities and a combination of the expectation maximization and the Newton-Raphson algorithm (De Roover et al., 2017) is used to update the state-specific MM parameters.

Second, the 3S estimation (Vogelsmeier et al., 2020) builds upon Vermunt's (2010) ML method and decomposes the estimation into three steps. First, in step 1, the state-specific MMs are obtained with a mixture GPCM while treating repeated measures $\mathbf{y}_{it}$ as independent. This entails that the relations between the latent states $\mathbf{s}_{it}$ at consecutive measurement occasions (i.e., the transitions) and the relations between the state memberships and covariates $\mathbf{z}_{it}$ are disregarded. This is valid because observations at one time-point are only indirectly related to covariates and to observations at other time-points, that is, via the latent states. This can also be seen from the graphical representation in Figure 5.1a.[43] The mixture GPCM is estimated with a combination of the expectation maximization and Newton-Raphson algorithms. Then, in step 2, observations are assigned to the state-specific MMs based on the most likely state membership and the corresponding classification error is calculated. Finally, in step 3, the CT-LMM with covariates is estimated using the state assignments from the previous step as indicators (thus fixing the MMs) while correcting for classification error inherent to the state assignments from step 2. At this point, one may also include a latent class variable to capture differences in transition patterns. The (mixture) CT-LMM model is estimated with a combination of the forward-backward and Newton-Raphson algorithms. Summarized, the steps are:

1. Estimating state-specific MMs (disregarding the dependence of the observations).

2. Assigning observations to the states (depending on the most likely state membership).

3. Estimating the (mixture) CT-LMM with fixed MMs (correcting for step 2's classification error).

The 3S estimation is almost as good as the FIML estimation in terms of parameter estimation. Only the state recovery is slightly worse and the standard errors can be

[43] It is important to note, however, that the standard errors of the parameters would be underestimated without applying a correction because observations are nested and thus dependent within subjects. This is only necessary when relying on hypothesis tests to determine which parameters differ significantly between the states (the possibility to use such tests will be described below). By providing LG with a "primary sampling unit" (PSU) identifier, the estimation takes into account that observations may come from the same sampling unit, that is, the subject (Vermunt & Magidson, 2016).

slightly overestimated (Vogelsmeier et al., 2020).[44] Apart from that, the 3S approach comes with several advantages. First, step-wise procedures are more intuitive for researchers who use complex methods such as LMLTA or LMFA to analyze their data because it is in line with how they prefer to conduct their analyses (Vermunt, 2010). That is, they see the investigation of the different MMs underlying their data as a first step and the investigation of subject's transitions between the MMs over time as well as the exploration of possible covariate effects as a next step.

Second, LMLTA (like LMFA) is an exploratory method, which entails that the best number of states k and factors per states R_k has to be determined. To this end, a large number of (plausible) models has to be estimated and compared by means of loglikelihood-based criteria that consider fit and parsimony. The evaluation of model selection criteria in LMLTA is beyond the scope of this chapter but, based on previous findings for related methods (Bulteel et al., 2013; Vogelsmeier, Vermunt, van Roekel, et al., 2019), we suggest to use the BIC in combination with the CHull and compare the three best models in terms of interpretability. Note that CHull balances fit and parsimony without making distributional assumptions and, thus, may perform better for some empirical datasets. In the FIML estimation, the number of models to be compared grows fast. For example, there are 9 models when comparing models with 1 – 3 states and 1 – 2 factors per state. When adding different (sets of) covariates to the CT-LMM, the 9 models have to be re-estimated for every set of covariates (e.g., $9 \times 5 = 45$ models for five different sets).[45] This problem is circumvented in the 3S estimation because the MMs and the CT-LMM are estimated separately. This implies that the model selection can be conducted in the first step, without being concerned about the covariates. Covariates (and latent classes) for the transition probabilities are added when estimating the CT-LMM.[46] As a result, there would only be $9 + 5 = 14$ models for five sets of covariates. Note that LG provides Wald tests (Agresti, 1990) that can be used to evaluate whether the covariates are significantly related to the transition or initial state parameters and to determine which MM parameters differ between the states. For the latter, one may also use visual inspection.

[44] Note that another limitation concerns the possible violation of the first-order Markov assumption (i.e., that the state membership at time-point t is not only influenced by the state membership at $t-1$ but also, e.g., by the occupied state at $t-2$; see Section 5.2.2). Only the FIML approach could capture such a dependency. However, with regard to other violated assumptions (e.g., covariates having direct effects on indicators), the FIML approach would suffer more from bias than the 3S approach but discussing the consequences is beyond the scope of this chapter (for a description of the problems and solutions, see Vermunt & Magidson, 2020).

[45] Note that the number of models grows even faster when also exploring different numbers of latent classes.

[46] The MMs are kept fixed (thus, are not re-estimated) once the covariates are included to the CT-LMM. Otherwise, the optimal model complexity in terms of factors and states could change (Di Mari et al., 2016).

Third, the FIML estimation takes several hours for each model while the 3S estimation is usually done in less than 30 minutes. This makes the FIML estimation less desirable, or even unfeasible, when researchers want to explore several covariate effects on MM changes. For all these reasons, we employ the 3S estimation in this study (for details, see Appendix L.2).

5.3 Application

5.3.1 Data

The data stem from a larger "Grumpy or Depressed?" study, which aimed to assess whether daily mood profiles (i.e., variability in affect) would predict the risk for depression in adolescents in the long run as recent work has indicated that the short-term dynamics could be linked to long-term psychopathology (e.g., Maciejewski et al., 2019; for a description of the study, see, e.g., de Haan-Rietdijk et al., 2017; Janssen, Elzinga, Verkuil, Hillegers, & Keijsers, 2020; van Roekel et al., 2019). Briefly, during three 7-day measurement bursts or "waves" (with approximately 3-month intervals in between), 250 Dutch adolescents (12 to 16 years old) completed up to eight questionnaires per day at random moments (median interval: 2.25 hours).[47] Out of the 250 adolescents, 164 participated in all three waves, 38 in two of the waves and 48 in one of the waves. In total, the adolescents completed 14,432 questionnaires.

5.3.2 Measures

For each assessment, adolescents indicated the degree to which 12 affect items applied to them (see Table 5.1) using 7-point Likert items (ranging from 1 = "not feeling the emotion" to 7 = "definitely feeling the emotion"). The items covered both PA and NA. The NA items were especially heavily right-skewed. Thus, LMLTA is particularly suited to investigate MM changes. The adolescents also indicated their current social interactions, resulting in the three "social context" covariates "being with friends" ("*fri*"), "being at school/with classmates" ("*cm*"), and "being with family" ("*fam*"), with 0 = "no" and 1 = "yes". At the beginning of every ESM wave (i.e., three times), the adolescents completed the Dutch version of the Children's Depression Inventory (CDI-I; Kovacs, 1992; Timbremont, Braet, & Roelofs, 2008) to screen for (sub)clinical depression ("*dep*"). The 27 items refer to symptoms during the last two weeks scored on three levels representing

[47] Note that the researcher studied affect dynamics at multiple time scales because affect can change within hours, days, and weeks; Houben, Van Den Noortgate, & Kuppens, 2015). This measurement burst design (Nesselroade, 1991) enabled the combination of different time scales (i.e., daily fluctuations in affect and long-term change in depression), while minimizing the burden for the participants. Furthermore, random measurement occasions facilitated capturing the continuously evolving daily dynamics in affect, minimizing effects of anticipated beeps and structural day routines on the assessment of affect (van Roekel et al., 2019).

low severity (0), medium severity (1), and high severity (2); for instance, "I get sad from time to time", "I get sad often", and "I'm always sad." Applying CDI-I cut-off scores (Kovacs, 1992; Timbremont et al., 2008), adolescents with a total score under 12 were categorized as "no depression" (89%) and all others as "(sub-)clinical depression" (11%).

The dataset contains several covariates but, in this study, we focused on the social context and depression as we found these variables particularly interesting to relate to possible MM changes: Emotional experiences may vary depending on the social context. For instance, adolescents may experience elevated positive mood when being among friends, whereas they may be somewhat more irritable and unhappy in the company of their parents, and more demotivated at school (Kendall et al., 2014; Soenens, Deci, & Vansteenkiste, 2017; van Roekel et al., 2013). For some adolescents, mood may be context-independent. Firstly, some adolescents could be in an overall positive mood regardless of the social context (Dietvorst et al., 2021). Secondly, adolescents with a depression and those at risk for developing a depression may be rather stable in their emotions in that they often feel unhappy and irritable in any social context (Dietvorst et al., 2021; Kendall et al., 2014; Silk et al., 2011). Therefore, for some adolescents, we expect a particular state membership to be more likely in one social context than in another, but also that adolescents differ in their state membership stability, for example, based on their depression level.

5.3.3 Description of the Applied Mixture CT-LMLTA Model

We will examine the context-dependency of state memberships by regressing the transition intensities (as defined in Equation (5.5)) on the social context covariates when estimating the CT-LMM (in step 3 of the estimation). To capture potential between-adolescent differences in stability, we will include a latent class variable that automatically classifies the adolescents based on their transition patterns, making the model a mixture CT-LMM as briefly introduced in Section 5.2.2. To see how many different patterns there are, we will compare models with 1–3 classes in terms of their fit by means of the BIC and CHull. Note that adolescents are allowed to transition to another class at the beginning of each wave—because subjects may change in their transition patterns over time (possibly related to their wave-specific depression scores)—such that the latent class variable is, strictly speaking, another state variable modeled via a DT-LMM (note that a DT model makes sense here as the intervals between the waves are approximately the same for all adolescents). To prevent confusion with the MM state, we will just refer to this latent variable as "class", with $c_{idv} = 1$ referring to being in a particular class v (with $v = 1, \dots, V$) in a particular wave d (with $d = 1, 2, 3$). To investigate whether experiencing depression affects the class membership, the initial

class and transition probabilities of the classes will be regressed on depression.[48] Moreover, we will evaluate the relation between the social context and the state memberships and investigate whether these relations depend on the class membership. For $V > 1$ and with $v = 1$ as reference category for the class, the specification of the transition intensities of the states (for $k \neq l$) is:

$$\begin{aligned}\log q_{lk} = \gamma_{0lk} + \sum_{v=2}^{V} \gamma_{lk,v} c_{itv} + \sum_{v=1}^{V} \gamma_{lk,fam,v} (fam_{it} \times c_{itv}) \\ + \sum_{v=1}^{V} \gamma_{lk,cm,v} (cm_{it} \times c_{itv}) + \sum_{v=1}^{V} \gamma_{lk,fri,v} (fri_{it} \times c_{itv}).\end{aligned} \tag{5.11}$$

The specification of the initial class (for $v = 2, \dots, V$) and the transition probabilities for the classes (for $v \neq b$ with $b = 1, \dots, V$) are given by:

$$\begin{aligned}\log \frac{p(c_{i1v} = 1 | dep_{i1})}{p(c_{i11} = 1 | dep_{i1})} = \beta_{0v} + \beta_{v,dep} dep_{id} \quad \text{and} \\ \log \frac{p(c_{idv} = 1 | c_{id-1,b} = 1, dep_{id})}{p(c_{idb} = 1 | c_{id-1,b} = 1, dep_{id})} = \gamma_{0bv} + \gamma_{bv,dep} dep_{id},\end{aligned} \tag{5.12}$$

respectively. Note that this application is meant to illustrate the empirical value of tracing MM changes with LMLTA. No hypotheses were preregistered and all analyses are exploratory so that interesting findings should be validated in future research before drawing any conclusions.

5.3.4 Obtaining and Investigating the Results of the Mixture CT-LMLTA Model

Below, we follow the three consecutive steps of the 3S estimation described in Section 5.2.2.

[48] Note that some adolescents (17 in wave 1, 26 in wave 2, and 18 in wave 3) missed out on the CDI-I questionnaire, but did participate in the ESM study, and therefore had no depression score in a given wave. For adolescents who had at least one score in any wave, we imputed their average total score and calculated the scale scores according to the cut-off values. For the other cases (i.e., 9 in wave 1, 1 in wave 2, and none in wave 3), LG automatically uses the average effect for predicting the initial class and transition probabilities.

Table 5.1. Differences in factor loadings, factor (co-)variances, factor correlations, and item means across the two states

	State 1 loadings λ_{jr1}		State 2 loadings λ_{jr2}		Between-state loading difference statistics						Item means	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$	$r = 1$			$r = 2$				
Item j	PA	NA	HA-PA	LA-PA/NA	Wald	df	p-value	Wald	df	p-value	State 1	State 2
relaxed	**0.63**	0.03	0.17	**-0.71**	5.34	1	0.02	7.22	1	< 0.01	5.72	6.89
content	**0.96**	0.00	0.32	**-1.09**	6.65	1	< 0.01	7.31	1	< 0.01	5.76	6.92
confident	**0.46**	0.02	0.21	**-0.48**	1.53	1	0.22	6.93	1	< 0.01	5.66	6.85
happy	**1.00**	0.00	**1.00**	0.00	/	/	/	/	/	/	5.62	6.81
energetic	**0.51**	0.00	**1.18**	0.30	6.67	1	< 0.01	3.21	1	0.07	5.21	6.41
excited	**0.69**	0.00	**1.23**	0.17	6.27	1	0.01	1.35	1	0.25	5.27	6.60
sad	0.04	**0.74**	0.05	**0.88**	0.18	1	0.67	1.44	1	0.23	1.09	1.03
unhappy	0.00	**1.00**	0.00	**1.00**	/	/	/	/	/	/	1.06	1.02
disappointed	0.08	**1.06**	0.14	**1.18**	0.34	1	0.56	0.30	1	0.58	1.07	1.04
angry	0.14	**0.99**	0.14	**1.08**	0.00	1	1.00	0.11	1	0.74	1.04	1.02
nervous	-0.01	**0.41**	0.10	**0.52**	1.70	1	0.19	0.33	1	0.57	1.24	1.09
irritated	0.00	**0.48**	0.05	**0.48**	0.53	1	0.47	0.00	1	1.00	1.24	1.16
Variances (chol)	3.69	3.53	2.18	0.96	14.02	1	< 0.01	23.94	1	< 0.01	/	/
Cov. (chol) with $q = 1$	/	-2.32	/	-1.5	/	/	/	2.38	1	0.12	/	/
Cor. with $q = 1$	/	-0.55	/	-0.84	/	/	/	/	/	/	/	/

Note. PA = Positive Affect; NA = Negative Affect; HA = High Arousal; LA = Low Arousal; Cov. = covariances; chol = Cholesky decomposed; Cor. = correlation; j refers to items, and r to factors. For identification purposes, we set the underlined loadings of the items "happy" on the first factors ($r = 1$) equal to 1 and on the second factors ($r = 2$) equal to 0 and the underlined loadings of the item "unhappy" on the first factors ($r = 1$) equal to 0 and on the second factors ($r = 2$) equal to 1. For each item and state, the loading with the largest absolute value is printed in boldface.

5.3.4.1 ***Step 1 & 2: Estimating state-specific MMs & assigning observations to the states***

5.3.4.1.1 *Model Selection*

To select the best fitting model, we conducted the mixture GPCM analysis for models with 1 – 3 states and 1 – 2 factors per state (i.e., 9 models[49]). Considering 1 to 2 factors not only preserves computational feasibility but also makes sense for affect questionnaires as PA and NA are often found as primary affect dimensions that may collapse into one bipolar factor if the emotions are strongly negatively related (Dejonckheere et al., 2018; Vogelsmeier et al., 2020). We selected the model with two states and two factors in each state because it was the best according to the BIC and among the two best models according to the CHull (for model selection details, see Appendix L.3; for the syntax of the selected model, see Appendix L.5). Forty-two percent of the observations belonged to MM 1 and 58% to MM 2.

5.3.4.1.2 *Results and Interpretation*

To examine the between-state MM differences, we first looked at the state-specific loadings in Table 5.1. Note that we modeled the covariance matrices in both states. To set the factor scales, we set the loadings of the items “happy” on factor 1 and “unhappy” on factor 2 equal to 1 in both states. To eliminate rotational freedom, we set the remaining loadings of the same items equal to zero. This has led to a well-interpretable simple structure. State 1 is characterized by separate PA and NA factors that correlated negatively ($r = -.55$) among observations in the same state. This means that adolescents distinguish somewhat between PA and NA, but that adolescents who score high on PA tend to score low on NA and vice versa. In contrast, in state 2, the three low arousal PA (LA-PA) emotions collapse with the NA emotions into one bipolar factor whereas the three high arousal PA (HA-PA) emotions make out the second factor. However, the factors have an even larger negative correlation than in state 1 ($r = -.84$). This implies that adolescents in state 2 distinguish more between LA-PA and HA-PA than they do between (LA-)PA and NA. Note that strong negative correlations between PA and NA are common in assessments that take place within small time-periods and in questionnaires that contain items with semantic antonyms such as "happy" and "unhappy" or "sad" (Dejonckheere et al., 2018).[50]

---

[49] The 9 models are [2 2 2], [2 2 1], [2 1 1], [1 1 1], [2 2], [2 1], [1 1], [2], and [1]. The notation means, for instance, that model [2 1 1] has three states with 2, 1, and 1 factors in each state, respectively.

[50] One might wonder if the loading pattern emerged only because of our chosen identification constraints. Therefore, for the same model, we also investigated a solution without correlations between the latent factors, with variances set to 1, and with the loadings of the item “irritated” set to 0 for the first factor in

Next, we investigated the between-state differences in the mean item scores. These scores are directly related to the state- and category-specific intercepts (which are given in Table L.2 in Appendix L), but the item means are easier to interpret. They are calculated as $\sum_{g=1}^{G} g \times p\left(y_{itj} = g | \mathbf{f}_{it} = \mathbf{0}, s_{itk} = 1\right)$ and thus a function of the logistic model for the probability of giving a response g as defined in Equation (5.8) with the factor scores $\mathbf{f}_{it}$ set equal to $\mathbf{0} = (0, 0)'$. As can be seen from Table 5.1, the means of the PA items are higher than the means of the NA items in both states. However, the PA means are lower in state 1 than in state 2. Thus, adolescents who distinguish more between LA-PA and HA-PA report a slightly better mood.

5.3.4.2 ***Step 3: Estimating the Mixture CT-LMM with Fixed MMs***

Since each adolescent may have a different MM at different measurement occasions, we examined adolescents' transitions from one state to another. Additionally, as motivated above, we investigated (1) whether adolescents differed in their state- (and thus MM-) membership by classifying the adolescents based on their transition patterns (i.e., transitions between states from one measurement occasion to the next) into latent classes that could differ across the three waves, (2) whether the wave-specific covariate depression had an influence on this class membership, and (3) whether the time-varying social context covariates (family, classmates, and friends) affected the transitions between the states and whether these effects differ across classes. To this end, we estimated the mixture CT-LMM with the state assignments from step 2 of our analysis as indicators while accounting for the inherent classification errors. Note that the correction was hardly necessary as the classification errors were very small due to a high state separation (with $R^2_{entropy} = .86$)[51], which means that most observations were assigned to a state with a high certainty in step 2 of the analysis.

5.3.4.2.1 *Model Selection*

We first estimated the "full" model as summarized in Equation (5.11) and (5.12) for 1–3 classes (i.e., with all possible covariates as just described). In the 2- and 3-class solutions, the effects of depression on the initial class ($\beta_{v,dep}$) and on the transition probabilities for the classes ($\gamma_{rv,dep}$) were non-significant. Hence, the class membership was unaffected by the level of depression. Furthermore, the effects of being with family ($\gamma_{lk,fam,v}$) and classmates ($\gamma_{lk,cm,v}$) on the transitions between the states significantly differed across

---

both states. The results can be found in Appendix L.4. Again, the solution shows that the three HA-PA emotions in state 2 stand out from the other emotions. Thus, we are confident about this finding.

[51] The $R^2_{entropy}$ value defines how much the state-membership prediction improves when using the observations $\mathbf{y}_{it}$ compared to when the state membership is predicted without them. The values range from zero, where the prediction is no better than chance, to one, where the prediction is perfect.

classes, whereas the effect of being with friends ($\gamma_{lk,fri,v}$) did not significantly differ across classes. However, being with friends in itself had a significant effect on the transitions between the states (i.e., there was an effect but it did not differ across classes). Therefore, we re-estimated the 2- and 3-class models while omitting depression and the conditional effect of being with friends but including a class-independent effect of being with friends (i.e., $\gamma_{lk,fri}$). Comparing all full and "reduced" models, the reduced 3-class model had the best fit according to the BIC and was among the best three models according to the CHull (for model selection details, see Appendix L.6; for the syntax of the full and reduced 3-class models, see Appendix L.5).[52]

5.3.4.2.2 *Results and Interpretation*

Table 5.2 shows the parameters of the final model. First, we looked at the three classes that captured differences in adolescents' between-state transitions. To this end, we computed the probabilities for the median interval (2.25 h) and mean covariate values:[53]

$$\mathbf{P}_{states}^{v=1} = \begin{pmatrix} 0.86 & 0.14 \\ 0.44 & 0.56 \end{pmatrix}, \mathbf{P}_{states}^{v=2} = \begin{pmatrix} 0.58 & 0.42 \\ 0.15 & 0.85 \end{pmatrix}, \mathbf{P}_{states}^{v=3} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad (5.13)$$

Class 1 and 2 each include 25% of the adolescents, whereas 50% were assigned to class 3. As can be seen from the relatively large values in column 1 of $\mathbf{P}_{states}^{v=1}$, adolescents in class 1 had a higher probability to transition to and stay in state 1 (i.e., PA vs. NA), whereas adolescents in class 2 had a higher probability to transition to and stay in state 2 (HA-PA vs. LA-PA/NA), which can be seen from the relatively large values in column 2 of $\mathbf{P}_{states}^{v=2}$. Thus, 25% of the adolescents are mostly in state 1 and 25% are mostly in state 2. In class 3, transitions to another class were highly unlikely since the (rounded) off-diagonal elements are equal to zero in $\mathbf{P}_{states}^{v=3}$, implying that adolescents in this class largely showed within-person invariance. Over the three waves with 3-month intervals, more adolescents transitioned to the stable class 3, as can be seen from the third column of the matrix containing the probabilities to transition between classes from one wave to another:

[52] Note that we also explored whether using the total depression scores instead of the dichotomous cut-off scores would change the results, which was not the case.
[53] As previously described, Figure 5.1 shows how to read a transition probability matrix. In Appendix L.7, we provide R code for calculating the transition probability matrix from the parameter estimates in Table 5.2 for any class, covariate, and time interval of interest.

$$\mathbf{P}_{classes} = \begin{pmatrix} 0.69 & 0.09 & 0.21 \\ 0.12 & 0.59 & 0.29 \\ 0.06 & 0.05 & 0.88 \end{pmatrix}. \quad (5.14)$$

Thus, over the three waves, adolescents developed a more stable assessment of their feelings. Perhaps their repeated answers to the questionnaire helped them to develop emotional awareness.

Considering the most prominent results (i.e., $p < 0.01$) of the social context covariates, we can see that the two class-dependent covariates (being with family and with classmates) had no effect in the stable class 3. In class 1 and 2, being with family decreased the probability of moving to state 1 ($\hat{\gamma}_{l=2,k=1,fam,v=1} = -0.63$; $\hat{\gamma}_{l=2,k=1,fam,v=2} = -1.12$). This implies that the probability to be in state 2 increased. Thus, when being with family (compared to not being with family), adolescents distinguish more between LA-PA and HA-PA and less between (LA-)PA and NA. One can imagine that HA-PA and LA-PA can emerge as separate factors. For example, while watching Netflix with the family, adolescents might feel "content" or "relaxed" but not "excited".

For adolescents in class 1, being with classmates decreased both the probability of moving to state 2 and moving to state 1 ($\hat{\gamma}_{l=1,k=2,cm,v=1} = -2.62$; $\hat{\gamma}_{l=2,k=1,cm,v=1} = -1.30$) such that state memberships became more stable. It is plausible that schools provide a relatively structured and therefore stable environment, which affects adolescents' emotional well-being less strongly than the more volatile experiences of being with family and friends.

In all three classes, being with friends (compared to not being with friends) decreased the probability of moving to state 2 ($\hat{\gamma}_{l=1,k=2,fri} = -0.63$).[54] The same was found for adolescents being with classmates in class 2 ($\hat{\gamma}_{l=1,k=2,cm,v=2} = -0.75$). This implies that, for them, the probability to be in state 1 increased and thus that adolescents tended to distinguish more between PA and NA. One possible explanation is that social support of friends is very important for adolescents (Bokhorst, Sumter, & Westenberg, 2010), so that adolescents who are "unhappy", for instance, because they failed a test, may still feel "content" when they are among their friends (and possibly classmates). Although one would expect to find an elevated mood when adolescents are with their friends (Kendall et al., 2014; van Roekel et al., 2013), the PA in this state is slightly lower

[54] Note that there is only one effect because the relation between being with friends and the state membership was not conditional on the classes in the final model.

than in state 2, perhaps because adolescents visit their friends more often when feeling bad and/or are more likely to discuss negative emotions with friends than with, for instance, family.

Table 5.2. Parameter estimates for the mixture CT-LMM in step 3 of LMLTA

	Parameter	Estimate	SE	z-value	p-value	Wald	df	p-value
DT-LMM for Classes								
Initial Class	$\beta_{0v=2}$	0.19	0.22	0.90	0.37	12.12	2	< 0.01
	$\beta_{0v=3}$	0.60	0.19	3.22	< 0.01			
Transition Intercepts	$\gamma_{0b=1,v=2}$	-2.02	0.54	-3.75	< 0.01	103.6	6	< 0.01
	$\gamma_{0b=1,v=3}$	-1.18	0.33	-3.62	< 0.01			
	$\gamma_{0b=2,v=1}$	-1.62	0.49	-3.34	< 0.01			
	$\gamma_{0b=2,v=3}$	-0.70	0.30	-2.35	0.02			
	$\gamma_{0b=3,v=1}$	-2.61	0.43	-6.04	< 0.01			
	$\gamma_{0b=3,v=2}$	-2.86	0.47	-6.06	< 0.01			
CT-LMM for States								
Initial State	$\beta_{0k=2}$	0.02	0.13	0.17	0.86	0.03	1	0.86
Transition Intercepts	$\gamma_{0l=1,k=2}$	-0.55	0.20	-2.69	< 0.01	23.19	2	< 0.01
	$\gamma_{0l=2,k=1}$	-0.08	0.20	-0.40	0.69			
Effect of Class	$\gamma_{l=1,k=2,v=2}$	0.00	0.25	-0.01	0.99	588.60	4	< 0.01
	$\gamma_{l=1,k=2,v=3}$	-7.21	0.38	-19.16	< 0.01			
	$\gamma_{l=2,k=1,v=2}$	-1.71	0.27	-6.32	< 0.01			
	$\gamma_{l=2,k=1,v=3}$	-8.74	0.60	-14.55	< 0.01			
Effect of Family × Class	$\gamma_{l=1,k=2,fam,v=1}$	-0.48	0.22	-2.17	0.03	40.49	6	< 0.01
	$\gamma_{l=1,k=2,fam,v=2}$	-0.10	0.20	-0.51	0.61			
	$\gamma_{l=1,k=2,fam,v=3}$	-1.11	0.55	-2.02	0.04			
	$\gamma_{l=2,k=1,fam,v=1}$	-0.63	0.22	-2.81	< 0.01			
	$\gamma_{l=2,k=1,fam,v=2}$	-1.12	0.26	-4.22	< 0.01			
	$\gamma_{l=2,k=1,fam,v=3}$	-2.27	1.47	-1.54	0.12			
Effect of Classmates × Class	$\gamma_{l=1,k=2,cm,v=1}$	-2.62	0.39	-6.77	< 0.01	113.30	6	< 0.01
	$\gamma_{l=1,k=2,cm,v=2}$	-0.75	0.25	-3.04	< 0.01			
	$\gamma_{l=1,k=2,cm,v=3}$	-2.70	1.87	-1.45	0.15			
	$\gamma_{l=2,k=1,cm,v=1}$	-1.30	0.26	-4.94	< 0.01			
	$\gamma_{l=2,k=1,cm,v=2}$	0.51	0.25	2.07	0.04			
	$\gamma_{l=2,k=1,cm,v=3}$	-0.96	0.84	-1.14	0.25			
Effect of Friends	$\gamma_{l=1,k=2,fri}$	-0.63	0.16	-3.92	< 0.01	16.96	2	< 0.01
	$\gamma_{l=2,k=1,fri}$	-0.39	0.17	-2.36	0.02			

Note. DT = discrete-time, CT = continuous-time, LMM = Latent Markov Model, Family (fam) refers to being with family, Classmates (cm) refers to being at school/with classmates, Friends (fri) refers to being with friends, v refers to a class in wave d, b refers to a class in wave $d-1$, k refers to a state at time-point t, and l refers to a state at time-point $t-1$. The overall Wald test for the differences in parameters between the classes for Family × Class was Wald $(4) = 18.29$, $p < 0.01$. For Classmates × Class the Wald test was Wald $(4) = 27.86$, $p < 0.01$. The covariate effects on the state transitions can be understood as follows: negative estimates imply that the log intensities and therefore also the transition probabilities decrease (e.g., the estimate $\hat{\gamma}_{l=2,k=1,fam,v=2} = -1.12$ means that the probability of transitioning from state $l = 2$ to state $k = 1$ for a subject in class $v = 2$ is lower when the subject is with family compared to when the subject is not with family). The estimates can also be used to calculate the transition probabilities for any class, covariate value and time interval of interest. An example showing how to calculate the parameters in R is provided in Appendix L.7.

5.3.4.3 *Summary of the LMLTA Findings*

We conclude that (1) two MMs were underlying adolescents' responses: in state 1 (42% of all observations), adolescents distinguished mainly between PA and NA and had a slightly worse mood than in state 2 (58% of all observations), where adolescent distinguished more between LA-PA (e.g., content) and HA-PA (e.g., excited) than they did between (LA-)PA and NA; (2) three state-transition patterns were found, implying that adolescents indeed differed in the stability of their emotional experience: in class 1, adolescents frequently transitioned between the states with a high probability to be in state 1; in class 2 they frequently transitioned but were more likely to be in state 2, and in class 3, they mainly stayed in one of the two states; (3) depression did not influence the class membership and thus the transition pattern; (4) for the unstable classes 1 and 2, being with family increased the probability to be in state 1; (5) for class 1, being with classmates increased the probability of staying in either state; (6) for all classes, being with friends—and for class 2, being with classmates—increased the probability to be in state 1. Our results show that researchers can obtain valuable insights from investigating MM changes and that it is important to consider the possibility that changes in positive or negative affect (e.g., evaluated by means of investigating changes in sum scores) could come from variability in the underlying MMs. Therefore, the novel method LMLTA (or LMFA) can improve the emerging trend of studying emotional dynamics as predictors of future well-being and psychopathology. In the future, it would be interesting to study the MMs and transition patterns in a larger group of adolescents with (different levels of) depression and to include other covariates that may explain differences in transition patterns and state-membership probabilities. For example, stress can cause a simplified representation of emotions (Dejonckheere, Mestdagh, et al., 2019), which can lead to very high correlations between emotions.

5.4 Discussion

In recent years, the awareness of potential measurement model (MM) changes in intensive longitudinal data—and the associated comparability problems—increased among substantive researchers and they are keen to evaluate such changes with new methods like latent Markov factor analysis (LMFA) (Horstmann & Ziegler, 2020). Understanding subject- and context-dependent MMs in more detail may benefit future studies on daily life dynamics and also have clinical implications, for instance, when MMs can be related to the onset of psychopathology. However, up to now, only researchers whose data contained (approximately) normally distributed continuous items could benefit from LMFA, whereas intensive longitudinal data often contain ordinal item responses with few categories or skewed distributions. In this chapter, we combined the strength of LMFA to evaluate MM changes over time with the strength of latent trait (LT)

models accommodate ordinal data in the new latent Markov latent trait analysis (LMLTA).

We showed that LMFA and LMLTA are similar as they both capture discrete changes or differences in subjects' underlying MM and thus in how latent constructs are measured by observed item responses. The difference lies in the type of latent variable model that is used to specify the relations between the latent constructs and observed variables, which directly follows from the assumed distribution of the observed item responses. Whereas the factor analysis (FA) model in LMFA assumes normally distributed continuous item responses, the generalized partial credit model (GPCM) in LMLTA assumes ordinal responses. The GPCM differs from the FA model in that (1) it has one intercept per item category and not one per item, (2) error variances cannot be freely estimated as they need to be fixed for identification, (3) rotation is only possible by means of setting identifying constraints, and (4) the number of constructs that can be included in the model is limited due to the computationally more complex estimation. This implies that, in LMLTA, more parameters have to be estimated, error variances are assumed to be identical across states, and the model specification is less flexible than in LMFA. For these reasons, we believe that LMFA should be the preferred method if the items are approximately normal and are measured with at least five categories (Dolan, 1994). The robustness of LMFA against violations of normality has never been evaluated, however. In the future, it would therefore be important to formulate more concrete guidelines on the basis of a simulation study that is tailored to intensive longitudinal data and that provides information on the robustness of LMFA, for instance, in terms of sample size and number of measurement occasions, degree of skewness, and number of item response categories. In the meantime, researchers should be cautious and, in case of doubt, opt for LMLTA and compare its results to those of LMFA.

By investigating differences in discrete MM changes over time in relation to covariates, LMLTA is a valuable step towards validly studying psychological dynamics. Additionally, as briefly described in the introduction (Section 5.1), the results of LMLTA may also help researchers decide on subsequent analyses. When invariance is clearly untenable, further evaluating dynamics with an approach that builds upon the invariance framework is simply not appropriate. However, observations for which invariance holds can be used to study dynamics in latent processes with standard analyses (e.g., growth models, Muthén, 2002, or dynamic structural equation modeling, Asparouhov et al., 2017), without results being influenced by differences in the underlying MMs. Moreover, if partial invariance holds across states, one may also continue with latent process analyses either by removing items with non-invariant parameters or by allowing for state- (or subject- and time-point-) specific parameters. Finally, we would like to highlight that there is no gold standard yet in how to analyze intensive longitudinal data

and the latent variable framework that LMLTA is based on is only one possibility. There are various other reasonable frameworks for analyzing the data (e.g., network psychometrics; Epskamp, 2020; Marsman et al., 2018) and decisions about the data analysis can considerably impact, for example, clinical recommendations (Bastiaansen, Kunkels, et al., 2020). Therefore, in the future, it would be desirable to compare perspectives about psychological phenomena from various modeling approaches.

6

Tutorial for the R Package *lmfa*

This chapter is submitted to Behavior Research Methods as Vogelsmeier, L. V. D. E., Vermunt, J. K., & De Roover, K. *How to explore within-person and between-person measurement model differences in intensive longitudinal data with the R package lmfa.*

Abstract

Intensive longitudinal data (ILD) have become popular for studying within-person dynamics in psychological constructs (or between-person differences therein). Prior to investigating what the dynamics look like, it is important to examine whether the measurement model (MM) is the same across subjects and time and, thus, whether the measured constructs have the same meaning. If the MM differs (e.g., because of changes in item interpretation or response styles), observations cannot be validly compared. Exploring differences in the MM for ILD can be done with latent Markov factor analysis (LMFA), which classifies observations based on the underlying MM (for many subjects and time-points simultaneously) and thus shows which observations are comparable. However, the complexity of the method or the fact that no open-source software for LMFA existed until now may have hindered researchers from applying the method in practice. In this chapter, we introduce the new user-friendly software package *lmfa*, which allows researchers to perform the analysis in the freely available software R. We provide a step-by-step tutorial for the *lmfa* package so that researchers can easily investigate MM differences in their own ILD.

6.1 Introduction

In recent years, researchers have shown an increased interest in intensive longitudinal data (ILD) for studying dynamics of psychological latent constructs (or "factors"), such as depression or affective well-being, for many subjects over a longer period of time. The ILD are commonly obtained by means of experience sampling methodology (ESM; Scollon et al., 2003) where many subjects repeatedly complete small questionnaires—containing items intended to measures the latent factors—at random (or event-based) time-points, several times a day for several days or weeks via a smartphone app. State-of-the-art analyses to model dynamics in psychological factors for many subjects over time range from basic random effect models (for studying individual differences in the dynamics or average levels of the measured factors; Hamaker, Ceulemans, Grasman, & Tuerlinckx, 2015; Myin-Germeys et al., 2018), over multilevel autoregressive models (for studying individual differences in lagged relationships between measured factors; e.g., Bringmann et al., 2013), to dynamic structural equation modeling (DSEM) that, in addition to the methods just mentioned, also allow for the estimation of more complex models (e.g., models containing multiple outcome variables; McNeish & Hamaker, 2020).

While the technology to gather ILD and approaches to analyze dynamics in the measured constructs are readily available, an important point of concern of many researchers before they start their analyses is whether the latent factors actually have the same meaning across subjects and time-points and, thus, whether observations are comparable. For this, the measurement model (MM) needs to be invariant across observations, that is, measurement invariance (MI) must hold. The MM indicates which factors are measured by which indicators and, for continuous item responses, is traditionally obtained with factor analysis (FA; Lawley & Maxwell, 1962). In the resulting MM (or "FA model"), factor loadings indicate the extent to which items measure the factors and item intercepts indicate the expected item scores when scores on the factors are equal to zero. If the loadings, the intercepts, or the number of factors differ within or across subjects, MI is violated and factors cannot be meaningfully compared (Adolf et al., 2014). However, invariance within and between subjects is easily violated because of differences and changes in response styles (Moors, 2003; Paulhus, 1991) or item interpretations (Oort et al., 2005). Thus, the MM may be different across subjects but also change within a person over time.

To clarify possible non-invariance of MMs, consider the following example. A research team conducts an ESM study to investigate between-subject differences regarding dynamics in affective well-being of employees in the travel sector during the Covid-19 pandemic. On the one hand, the underlying MM may differ across employees because people generally differ in their ability to label emotions in a granular way

(Barrett, Gross, Christensen, & Benvenuto, 2001; Erbas, Kalokerinos, Kuppens, van Halem, & Ceulemans, 2020; Kashdan, Barrett, & McKnight, 2015). The "high differentiators" differentiate more between specific emotions, such as feeling content or happy, than the "low differentiators", who focus more on the valence of a feeling and, thus, whether an emotion is positive or negative (Barrett, 1998; Erbas, Ceulemans, Koval, & Kuppens, 2015). A result could be that several factors underlie the responses of the high differentiators (say, four factors pertaining to high- and low-arousal positive and negative affect) while only one factor underlies the responses of the low differentiators (say, a bipolar "valence of affect" factor). On the other hand, the MM may change within some employees over time. For instance, employees who are normally high differentiators may also switch to a MM with a single "valence of affect" factor when being exposed to a stressful situation (e.g., learning about government restrictions extending temporary part-time work) because stress triggers a valence focus (Erbas et al., 2018). Because the low differentiators respond according to a single valence of affect factor, regardless of experienced stress, the same MM would be underlying their responses during the entire participation.

Undetected measurement non-invariance is a threat to valid inferences from ILD analyses. Therefore, detecting non-invariance is crucial. Until recently, researchers could only test whether the MM is invariant across (groups of) subjects and/or time-points (e.g., by means of traditional MI tests that are, for example, available in the R package *lavaan*; Rosseel, 2012). However, if the results indicate that invariance is untenable, researcher cannot automatically identify for which subjects or time-points the MMs differ and what the different MMs look like without conducting pairwise comparisons of subject- or time-point specific MM parameters. This quickly becomes unfeasible for ILD that usually contain many observations from many subjects. Furthermore, it is only possible to investigate non-invariance across subjects (assuming invariance over time) or to investigate invariance over time (assuming invariance across subjects) and not to investigate both at the same time. These problems were solved by Vogelsmeier, Vermunt, van Roekel, et al. (2019), who developed latent Markov factor analysis (LMFA), which allows researchers to conveniently explore all kinds of MM differences, both across subjects and time. LMFA is a mixture modeling approach that combines a latent Markov model (LMM; Bartolucci et al., 2014; Collins & Lanza, 2010) with mixture FA (McLachlan & Peel, 2000; McNicholas, 2016): First, the LMM clusters observations according to their underlying MM into dynamic latent states. Note that the latent states are equivalent to latent classes in a latent class analysis or mixture model but are called states in a LMM because subjects can transition between latent classes over time. Second, for each state, FA reveals what the underlying MM looks like. Summarized, LMFA classifies observations into different states that pertain to different MMs and invariance holds for observations in the same state but is violated for observations in different states. Researchers can then

decide how to continue with their data analysis (e.g., retaining observations from one state or removing non-invariant items; see Section 6.4 for a more elaborate discussion on this). Researchers can also learn from subjects' transitions between MMs by including time-varying or time-constant covariates as predictors of the state memberships (e.g., "stress" could be included when analyzing the changes in the MM in our employee example).

Although more and more researchers are eager to explore MI in their ILD (Horstmann & Ziegler, 2020), many researchers are still unfamiliar with LMFA, how it is applied, or how results must be interpreted. Furthermore, until now, LMFA was only available in the commercial software Latent GOLD (Vermunt & Magidson, 2016) and, thus, not all researchers had access to the novel method. The latter has now changed with the release of the package *lmfa* (Vogelsmeier & De Roover, 2021) that allows researchers to perform all necessary steps in the open-source software R (R Core Team, 2020). The aim of this chapter is to provide a tutorial for the *lmfa* package that guides users through the different steps of performing the analysis and interpreting the results with the goal to increase researchers' confidence and ease in using LMFA. This can also indirectly improve research of applied or substantive researchers. The tutorial is targeted at an audience that has a basic understanding of R but not necessarily of the LMFA model.

The remainder of this chapter is organized as follows: First, in Section 6.2, we describe the data structure, introduce an example dataset, recap the LMFA method, and describe how it is estimated in *lmfa*. Then, in Section 6.3, we guide the reader through the different analysis steps by means of annotated R code. Next, in Section 6.4, we describe how to proceed with ILD analyses based on the results of LMFA and, finally, in Section 6.5, conclude with a discussion about current limitations and possible future extensions of *lmfa*.

6.2 Method

6.2.1 Data Structure

We assume typical ILD and, thus, repeated measures data containing several continuous and normally distributed[55] variables that were assessed for multiple subjects at multiple measurement occasions and which measure one or more latent factors. In addition to

[55] Note that one may also investigate ordinal data as long as the data has more than a few response categories (say at least five; Dolan, 1994) and the distribution across the categories is approximately normal. However, the influence on the parameter estimation has not yet been investigated for LMFA in particular and results should be therefore interpreted with caution. Highly skewed responses with only a few response categories may lead to convergence problems and local optima (for a description of local optima, see Section 6.3.2.3). For such data, the LMFA extension to categorical data should be used, that is, latent Markov latent trait analysis (Vogelsmeier, Vermunt, Keijsers, & De Roover, 2021).

these variables, the data may contain time-constant or time-varying explanatory variables (i.e., covariates), which may be dichotomous, ordinal, or continuous and which may explain transitions between the underlying MMs. For the mathematical notation of the data structure, see Appendix N.1. To clarify the data structure, consider the following example dataset that will be used throughout this tutorial.

6.2.2 Constructed Data Example

The data is a synthetic dataset that was inspired by a real ESM dataset, which was used in Vogelsmeier, Vermunt, van Roekel, et al. (2019) to illustrate how to explore MM changes by means of a LMFA without covariates. Every evening for about three months, multiple subjects (suffering from anhedonia, one of the core symptoms of depression; Van Roekel et al., 2017) reported their affect and the unpleasantness of the most unpleasant event they experienced since the previous measurement occasion (in the following just "negative event"). Affect was measured by means of ten positive affect (PA) and eight negative affect (NA) items (see LMFA output box 1) and a single item was used to assess the negative event. All items were assessed on a Visual Analogue Scale ranging from 0 = "Not at all" to 100 = "Very much". Moreover, after the first month, subjects were randomly assigned to receive an intervention to reduce anhedonia or not.[56] The results of LMFA indicated that most subjects transitioned between three MMs that differed with regard to the number and nature of the factors. Descriptive statistics showed that there was a relation between the states and the two covariates "had an intervention" and "negative event".

For the tutorial in this chapter, we created a dataset with MMs similar to the ones found in the real data application (yet, somewhat adjusted and simplified) and with the two time-varying covariates "had an intervention" (coded as 1 = "yes" and 0 = "no") and "negative event" having an effect on the transitions between the states. The dataset contains data for 100 subjects with a mean of 47.76 observations and an SD of 6.56, resulting in a total number of observation equal to 4776. The intervals between measurement occasions differ within and across subjects with an average length of 1.22 days and an SD of 1.02. The negative event scores differ within and across subjects with a mean of 49.65 and an SD of 15.11. Of all subjects, 50 receive no intervention and 50 receive one intervention after approximately 1/3 of their total participation duration. The dataset will be used throughout the article to ease understanding of the method and the steps of the tutorial.

[56] The intervention was either a personalized lifestyle advice or the advice in combination with a skydive. For simplicity in this tutorial, we do not distinguish between different types of interventions.

6.2.3 Latent Markov Factor Analysis (LMFA)

LMFA consists of two building blocks. The first one pertains to the state-specific MMs and, thus, a FA model for each state that indicates which constructs are measured by which items. Note that the state-specific FA assumes continuous item responses (or responses that can be treated as such, say, with five or more categories; Dolan, Oort, Stoel, & Wicherts, 2009). It is possible to use a latent variable model for items with categorical item responses (Vogelsmeier et al., 2021) but this option is not included in the *lmfa* package. The second building block is the LMM that models the transitions between MMs over time (Bartolucci, Farcomeni, et al., 2015; Zucchini et al., 2016). Note that there are two types of LMMs. First, the discrete-time- (DT-)LMM (Bartolucci, Farcomeni, et al., 2015; Zucchini et al., 2016) assumes the intervals between time-points to be equal across subjects and time. In contrast, the continuous-time- (CT-)LMM (Böckenholt, 2005; Jackson & Sharples, 2002) accommodates unequally-spaced observations, which is usually more realistic in ILD (e.g., due to random beeps or skipped measurement occasions). However, the CT-LMM also works for equal intervals. In fact, estimating a CT-LMM with equal intervals is similar to estimating a DT-LMM but the parameter interpretation differs (Vermunt & Magidson, 2016), which will be clarified in Section 6.2.3.2. The *lmfa* package uses CT-LMM because it is more generally applicable.

LMFA can be estimated with a full information maximum likelihood (FIML) estimation (Vogelsmeier, Vermunt, van Roekel, et al., 2019) or with a three-step (3S) estimation (Vogelsmeier et al., 2020). The latter breaks down the estimation of LMFA into three steps, which makes the analysis faster and more convenient for the investigation of covariate effects. Therefore, *lmfa* uses the 3S estimation. For a detailed discussion about the estimation procedures, we refer to Vogelsmeier et al. (2020).

In the next section, we explain and illustrate the method using LMFA results for our example data. Note that the true number of states and factors (i.e., like in the data generating model) was applied. Of course, the number of states and factors and the relevant covariates are not known in empirical practice. The required model selection and covariate selection procedures are explained in Section 6.3.2. In the following, we first explain the state-specific MMs (i.e., the FA models; Section 6.2.3.1). Then, we describe the (covariate-specific) transition model (i.e., the CT-LMM; Section 6.2.3.2). Next, we summarize the finding of applying LMFA to our example data (Section 6.2.3.3). Thereafter, we explain the 3S estimation (Section 6.2.4).

6.2.3.1 *The State-Specific Measurement Models*

In LMFA, the MMs are determined by state-specific FA models, which consist of three types of parameters. Depending on which parameters differ across states, different levels

of MI are violated. In this section, we first explain the parameters of the FA model, the different levels of (non-)invariance, and the exploratory FA that is used to obtain the parameter estimates. Thereafter, we demonstrate potential non-invariance of the FA models across states with our example data.

6.2.3.1.1 *State-Specific FA Models and (Non-)Invariance Across States*

The first type of parameters are the factor loadings, which determine the item-factor relations and, hence, the degree to which an item measures a factor or, stated differently, to what extent an item is predicted by the underlying factor. Thus, items with stronger loadings are better measures of a factor than items with lower loadings. Second, item intercepts are the expected scores for an item when the factor scores are equal to zero. Third, the items' unique variances indicate the variance of an item that is unique to the item and, hence, that is not explained by the factors (for the mathematical notation and the technical details, see Appendix N.2.2). The three types of parameters can take on different values across states and inform us about violations of four different levels of MI (Meredith, 1993). These levels are configural invariance (invariance of the number of factors and the pattern of non-zero loadings), weak invariance (invariance of the non-zero loadings), strong invariance (invariance of the intercepts), and strict invariance (invariance of the unique variances). It is important to note that strict invariance is assumed to hold within each state, since the states capture differences in loadings, intercepts, and unique variances.

For obtaining the state-specific MMs, LMFA uses exploratory FA (EFA) and not confirmatory FA (CFA). CFA is too restrictive because it imposes a priori assumptions about the presence or absence of item-factor relations by setting certain loadings equal to zero. Thus, CFA cannot detect MM differences pertaining to the configural model, such as the number and nature of the underlying factors in our previous employee example. In contrast, EFA detects all types of loading differences, including configural non-invariances like differences in cross-loadings. Note, however, that the EFA model is not identified without setting constraints. Firstly, one needs to set the scale of the factors. To this end, *lmfa* sets the factor (co-)variances equal to an identity matrix (with dimensions equal to the state-specific number of factors), which means that factors are initially uncorrelated. This initial solution is usually not well interpretable because many items may have high loadings on more than one factor (i.e., there is no "simple structure"; Thurstone, 1947). In order to achieve a more interpretable solution, *lmfa* applies a rotation of the factors for each state. An oblique rotation (i.e., one that allows factors to be correlated) results in the best simple structure and is usually more valid for psychological constructs (Clarkson & Jennrich, 1988; De Roover & Vermunt, 2019; Kiers, 1997). Finally, the factor means are set equal to zero per state. This implies that the state-specific intercepts are in fact state-specific item means.

6.2.3.1.2 *State-Specific FA Models for the Example Data*

To illustrate possible measurement non-invariance across states, consider the state-specific MMs resulting from LMFA with three states and three, two, and three factors ("[3 2 3]") of the synthetic dataset described in Section 6.2.2. The results are displayed in LMFA output box 1.[57]

LMFA output box 1

```
## Estimation converged after 168.67 seconds and 40 iterations.
##
## LL = -353166.81
##
## Number of states: 3
##
## Number of factors: [3 2 3]
##
## ------------------------------------------------------------
##
## Obliquely rotated standardized loadings:
##
##               S1F1  S1F2  S1F3     S2F1  S2F2     S3F1  S3F2  S3F3
## Interested    0.66  0.04  0.00     0.68  0.01     0.57 -0.01  0.02
## Joyful        0.60  0.02  0.02     0.65 -0.01     0.88  0.01  0.06
## Determined    0.37  0.03 -0.55     0.61  0.00     0.84  0.02 -0.01
## Calm          0.37 -0.58 -0.01     0.59  0.00     0.18 -0.15  0.82
## Lively        0.63  0.03  0.03     0.65  0.00     0.88 -0.01  0.01
## Enthusiastic  0.65 -0.01  0.02     0.64  0.00     0.89  0.02  0.00
## Relaxed       0.64  0.02  0.00     0.64  0.01     0.16 -0.14  0.85
## Cheerful      0.63  0.07  0.01     0.63 -0.01     0.91  0.01  0.02
## Content       0.61  0.00  0.03     0.67  0.02     0.93  0.02  0.01
## Energetic     0.64 -0.01  0.00     0.63 -0.01     0.90  0.05 -0.01
## Upset         0.09  0.62 -0.01     0.00  0.53     0.03  0.83 -0.03
## Gloomy       -0.24  0.39  0.44    -0.01  0.53     0.02  0.82 -0.01
## Sluggish      0.07 -0.01  0.73    -0.01  0.50    -0.29  0.34  0.77
## Anxious       0.09  0.70 -0.02     0.00  0.52     0.05  0.79 -0.01
## Bored         0.07 -0.01  0.74    -0.01  0.52     0.04  0.47 -0.04
## Irritated     0.06  0.51 -0.05     0.01  0.58     0.04  0.85 -0.02
## Nervous       0.08  0.73 -0.04     0.00  0.51     0.03  0.74  0.01
## Listless      0.06 -0.05  0.73     0.01  0.54     0.02  0.46 -0.03
##
##
## ------------------------------------------------------------
##
## Factor correlations after oblique rotation:
##
## S1
##       F1    F2    F3
## F1  1.00 -0.08 -0.24
## F2 -0.08  1.00  0.07
## F3 -0.24  0.07  1.00
##
## S2
```

[57] Note that, for clarity, loadings with absolute values larger than .3 are printed in boldface in this tutorial. However, the normal output in R does not include any boldface loadings.

```
##       F1    F2
## F1  1.00 -0.37
## F2 -0.37  1.00
##
## S3
##       F1    F2   F3
## F1  1.00 -0.04 0.23
## F2 -0.04  1.00 0.02
## F3  0.23  0.02 1.00
##
##
## ------------------------------------------------------------
##
## Intercepts:
##
##                 S1    S2    S3
## Interested   49.24 61.46 51.98
## Joyful       48.92 61.12 49.95
## Determined   46.60 61.20 50.35
## Calm         46.25 61.14 54.76
## Lively       49.29 60.85 50.57
## Enthusiastic 48.99 61.16 50.24
## Relaxed      49.00 61.12 54.90
## Cheerful     49.03 61.02 50.42
## Content      49.39 60.84 49.98
## Energetic    49.35 60.90 50.41
## Upset        44.12 26.54 36.42
## Gloomy       45.88 27.09 35.93
## Sluggish     44.95 26.54 33.26
## Anxious      45.81 26.48 35.83
## Bored        44.98 26.75 29.94
## Irritated    43.48 26.69 35.66
## Nervous      46.39 26.50 35.94
## Listless     45.35 26.84 29.67
##
## ------------------------------------------------------------
##
## Unique variances:
##
##                   S1    S2     S3
## Interested   273.26 53.37  96.43
## Joyful       273.82 48.67  92.81
## Determined   261.88 49.93  92.70
## Calm         265.99 51.75  99.17
## Lively       286.09 48.20 104.04
## Enthusiastic 257.06 50.09 107.07
## Relaxed      270.54 49.55  99.75
## Cheerful     284.69 50.47  83.05
## Content      271.52 41.15  92.38
## Energetic    271.12 53.24  95.55
## Upset        278.71 46.13  92.89
## Gloomy       256.03 46.46  73.85
## Sluggish     245.57 51.70  82.24
## Anxious      276.61 45.65  87.14
## Bored        253.52 47.69 103.11
## Irritated    267.30 44.56  84.99
## Nervous      261.57 49.30  86.21
## Listless     269.07 47.29  92.10
```

Intercepts. We first look at the intercepts because, for this data example, it allows us to give labels to the states that can be used throughout the interpretation of the other parameters. Specifically, we see that, in all three states, the positive emotions (i.e., "interested", "joyful", "determined", etc.) are larger than the negative emotions (i.e., "upset", "gloomy", "sluggish", etc.). However, the states differ regarding their overall positive and negative emotions. More specifically, the positive emotions are lowest in state 1 and the negative emotions are highest in state 1, followed by state 3 and then state 2. Therefore, in the following, we label the first state the "displeasure" state, the second one the "pleasure" state, and the third one the "neutral" state.

Loadings. Next, we inspect the loadings. Note that the default output displays standardized[58] obliquely rotated factor loadings.[59] The reason is that unstandardized values can be difficult to interpret as they often exceed an absolute value of 1 (especially when a large rating scale is used like in our example dataset; Section 6.2.2) and, hence, rules of thumb to evaluate which items have strong loadings on a factor cannot be applied. In contrast, for standardized loadings, rules of thumb are available (e.g., loadings with an absolute value larger than or equal to 0.3 can be seen as considerable, which is also the threshold used in our example).

Looking at the loadings, we see that, in all states, the first factors correspond to a positive affect (PA) factor containing loadings of most or all positive emotion items. However, in the displeasure state (i.e., state 1), the loadings of the items "determined" and "calm" are somewhat lower and, in the neutral state (i.e., state 3), the loadings "calm" and "relaxed" are even lower than the chosen threshold. Furthermore, the second factors are, broadly speaking, negative emotion factors but with even more prominent differences across states (especially between the displeasure and the pleasure state; i.e., state 1 and state 2). While the pleasure state has a clear negative affect (NA) factor with high loadings of all negative emotions, the displeasure state has a bipolar "distress" factor with loadings of the high arousal negative emotions and a reversed loading of the item "calm". The second factor in the neutral state (i.e., state 3) lies in between the factors of the pleasure and displeasure states in that it has considerable loadings of all the items but relatively low loadings of the low arousal emotions "sluggish", "bored", and "listless". The most striking difference is that the displeasure state (i.e., state 1) contains a third bipolar "drive" factor, whereas the neutral state (i.e., state 3) contains a third "serenity" factor. More specifically, the drive factor (or rather lack-of-drive factor) has high loadings of the low arousal negative emotions "gloomy", "sluggish", "bored" and "listless" and a

[58] More specifically, they are standardized by means of the standard deviations of the item scores across all states.
[59] If desired, however, the user can also request unstandardized and unrotated loadings.

reversed loading of the item "determined". The serenity factor has high loadings of the low arousal emotions "calm", "relaxed", and "sluggish". In conclusion, when subjects are in the displeasure or neutral state, they have a more differentiated representation of their emotions than when they are in the pleasure state. The drive factor in the displeasure state is especially interesting because it is in line with research showing that drive differs from general PA when persons are anhedonic (Berridge et al., 2009; Treadway & Zald, 2011). Moreover, it is noticeable that the loadings in the neutral state (i.e., state 3) are generally higher than the loadings in the other states. Thus, considering the overall item variances, the item-factor relations are strongest in the neutral state. Note that overall larger loadings may also be a result of larger factor variances as a consequence of constraining the factor variances to 1 (Section 6.2.3.1).

In addition to the loadings, it is interesting to also inspect the factor correlations that result from the oblique rotations—which are not part of the MM. First, in the displeasure state (i.e., state 1), we see a small negative correlation between PA and the lack-of-drive factor (i.e., factors 1 and 3). In the neutral state (i.e., state 3), we see a small positive correlation between PA and the serenity factor (i.e., factors 1 and 3). In the pleasure state (i.e., state 2), PA and NA (i.e., factors 1 and 2) are moderately negatively correlated. All other correlations are close to zero, indicating that the other factors are rather independent of each other.

Unique variances. Finally, looking at the unique variances, we see that they are largest in the displeasure state (i.e., state 1), followed by the neutral and the pleasure state (i.e., state 3 and 2). The large emotion-specific variability in the displeasure state is in line with findings that depression and emotional complexity are related (Grühn et al., 2013).

6.2.3.2 ***The Transition Model***

After examining the MMs, the next step is to investigate what the transitions between the MMs look like by means of the CT-LMM. As previously stated, the CT-LMM is a latent class model that allows subjects to transition between latent states over time. Specifically, we will inspect the probability to start in a state at the first time-point (i.e., "initial state probabilities") and the probabilities of transitioning to other states from one time-point to the next (i.e., "transition probabilities"). In this section, we first explain the initial state parameters and show what they look like for our example data. Then we describe the transition parameters and show the corresponding results for our example.

6.2.3.2.1 *Initial State Parameters*

The initial state probabilities pertain to the probability to start in a certain state at the first time-point. The probabilities sum to one and are stored in a vector with the number

of elements being equal to the number of states. For example, the vector $\boldsymbol{\pi} = (.42 \quad .34 \quad .24)$ shows that the probability of starting in state 1 is equal to .42, the probability of starting in state 2 is .34, and the probability of starting in state 3 is .24. In *lmfa*, logit models are used to model the initial state probabilities (similar as in logistic regression; Agresti, 1990). The inherent logit values (or "log-odds") indicate the relative chance to start in a state compared to a reference state (in *lmfa*, this is state 1). Note that a separate logit model is required for all states but the reference state. These logit values do not have to be interpreted because the initial state probabilities can be calculated from these logit models (see Appendix N.2.1).[60]

Finally, the initial state parameters may be related to covariates, which could be, for instance, scores on a baseline questionnaire (e.g., a depression score or a score for the general ability to differentiate between emotions). Note, however, that including covariates on the initial state parameters only makes sense if the dataset contains data of more than a few subjects.[61] Otherwise, there is not enough information to investigate the covariate effects. The covariates are related to the initial state parameters by means of regression and they affect the logits and not the probabilities directly (for details, see Appendix N.2.1). However, in order to see what effect the covariates have on the initial state probabilities, one can convert logits into probabilities for different covariate values and compare them. For example, for a categorical covariate with two categories, one could compare the initial state probabilities for both categories. For continuous covariates, one could compare the initial state probabilities corresponding to the sample mean plus the standard deviation of the covariate to the probabilities corresponding to the sample mean minus the standard deviation (or compare probabilities for different quantiles of the covariate) while setting other covariates equal to their (sample) means.

6.2.3.2.2 *Initial State Parameters for the Example Data*

To illustrate the interpretation, consider the results of a LMFA without covariates on the initial state parameters. The results are presented in LMFA output box 2.

LMFA output box 2

```
## Model estimation:
##
## Estimation converged after 1159.95 seconds.
##
## LL = -3712.4847
##
## ------------------------------------------------------------
##
```

[60] Note that *lmfa* users do not have to calculate any probabilities themselves as the package provides them.
[61] For an estimate of the required sample size, users may consult guidelines for multinomial logistic regression (e.g., de Jong et al., 2019).

```
## Wald tests:
##
##                   Wald df p-value
## intervention  213.3821  6       0
## negativeEvent  55.7629  6       0
##
## ----------------------------------------------------------------
##
## Parameter estimates:
##
##                                coef   s.e. z-value p-value
## initial state parameters 2 -0.1864 0.2408 -0.7741  0.4389
## initial state parameters 3 -0.5479 0.2949 -1.8577  0.0632
## transition parameters 1|2  -1.1725 0.3207 -3.6555  0.0003
## transition parameters 1|3  -1.5951 0.4235 -3.7661  0.0002
## transition parameters 2|1  -1.6011 0.4260 -3.7588  0.0002
## transition parameters 2|3  -0.4188 0.4183 -1.0013  0.3167
## transition parameters 3|1  -0.5761 0.2975 -1.9368  0.0528
## transition parameters 3|2  -1.3186 0.7413 -1.7788  0.0753
## intervention 1|2            0.6000 0.1786  3.3604  0.0008
## intervention 1|3            0.3228 0.2565  1.2582  0.2083
## intervention 2|1           -0.9528 0.2624 -3.6308  0.0003
## intervention 2|3           -0.4081 0.2474 -1.6497  0.0990
## intervention 3|1           -1.0119 0.1809 -5.5946  0.0000
## intervention 3|2            0.4767 0.4931  0.9668  0.3336
## negativeEvent 1|2          -0.0194 0.0057 -3.3903  0.0007
## negativeEvent 1|3          -0.0096 0.0081 -1.1931  0.2328
## negativeEvent 2|1           0.0153 0.0091  1.6812  0.0927
## negativeEvent 2|3          -0.0071 0.0079 -0.9001  0.3681
## negativeEvent 3|1           0.0104 0.0060  1.7364  0.0825
## negativeEvent 3|2          -0.0142 0.0132 -1.0753  0.2823
##
## Note: For the initial state parameters, state 1 is the
## reference category. The transition intensity parameters
## are sorted by rows of the transition matrix and the
## staying rates serve as references.
##
## ----------------------------------------------------------------
##
## Probabilities:
##
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##     S1     S2     S3
## 0.4153 0.3446 0.2401
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 0.4139
## negativeEvent score: 49.6505
##
##        S1     S2     S3
## S1 0.7923 0.1032 0.1046
## S2 0.2542 0.5388 0.2070
## S3 0.3909 0.1073 0.5019
##
```

```
##
## Note: The probabilities are calculated for covariate scores
## equal to the sample means (and a unit time interval). Use
## the function probabilities() to calculate initial state and
## transition probabilities for any covariate score (and
## interval) of interest.
##
## ------------------------------------------------------------
##
## State proportions:
##
##     S1     S2     S3
## 0.5517 0.2567 0.1916
```

For now, we focus only on the parts "`Parameter estimates`" and "`Initial state probabilities`", starting with the former. The "`coef`" and "`s.e.`" columns indicate the point estimates and standard errors, respectively. The "`z-value`" and "`p-value`" columns show the corresponding z-statistic and p-values. Since no covariates were included for the initial state parameters, there are only two "`initial state parameters`". These parameters always correspond to the logit values for covariate scores being equal to zero. In case of covariate effects, they would be shown below the initial state parameters. As previously described, it is more convenient to interpret the corresponding initial state probabilities. More specifically, in order to obtain a good impression of what the probabilities look like for the average person, it makes most sense to inspect the initial state probabilities for covariates being equal to the sample means. These probabilities can be found in the "`Initial state probabilities`" part further below in LMFA output box 2. Of course, if no covariates are defined (as in our model), the probabilities do not depend on the values of a covariate. The probabilities indicate that starting in the displeasure state was most likely, followed by the pleasure state and the neutral state.

6.2.3.2.3 *Transition Parameters*

The transition probabilities are stored in a matrix with dimensions equal to the number of states and the elements within a row of the transition probability matrix sum to one (Bartolucci, Farcomeni, et al., 2015; Zucchini et al., 2016). In order to clarify how to read a transition probability matrix, consider the following matrix:

$$\mathbf{P} = \begin{pmatrix} p_{11} = .66 & p_{12} = .18 & p_{13} = .16 \\ p_{21} = .20 & p_{22} = .49 & p_{23} = .31 \\ p_{31} = .32 & p_{32} = .17 & p_{33} = .51 \end{pmatrix}. \tag{6.1}$$

The rows indicate the state memberships at the previous time-point and the columns indicate the state memberships at the current time-point. This implies that the values on the diagonal specify the probabilities to stay in a state and the off-diagonal elements refer to the probabilities to transition to another state. For example, the first row of the matrix

shows that the probability to stay in state 1 is equal to .66 and the probabilities to transition from state 1 to state 2 and from state 1 to state 3 are equal to .18 and .16, respectively.

As described before, the transition probabilities depend on the interval between two consecutive measurement occasions. The larger the interval, the larger the probabilities to transition to another state. To accommodate the interval length, LMFA (using CT-LMM) does not estimate the transition probabilities directly. Instead, transition intensities (or "rates"; i.e., transition probabilities per very small time unit)[62] are estimated and the transition probabilities are computed based on the transition intensities and the intervals (Böckenholt, 2005; Jackson & Sharples, 2002).[63] The transition intensities are also captured in a matrix with dimensions equal to the number of states. However, intensities are only estimated for the transitions away from the origin state and, hence, for the off-diagonal entries. The diagonal entries are equal to the sum of the off-diagonal transition intensities, which implies that rows sum to zero (Cox & Miller, 1965). For example, consider the matrix that corresponds to the transition probabilities in Equation (6.1):

$$\mathbf{Q} = \begin{pmatrix} -q_{12} - q_{13} = -.51 & q_{12} = .31 & q_{13} = .20 \\ q_{21} = .20 & -q_{21} - q_{23} = -.86 & q_{23} = .66 \\ q_{31} = .56 & q_{32} = .28 & -q_{31} - q_{32} = -.84 \end{pmatrix}. \quad (6.2)$$

The rate to transition from state 1 to state 2 is $q_{12} = .31$ and the rate to transition from state 1 to state 3 is $q_{13} = .20$. Larger rates are related to larger transition probabilities away from a state.

The transition intensities are modeled by means of a log-linear model such that the parameters are not intensities but log intensities (thus, the parameterization differs from the logit parameterization of the initial state parameters). For example, the estimates for the log intensities corresponding to intensities for the first row in Equation (6.2) are $\log(q_{12} = .31) = -1.17$ and $\log(q_{13} = .20) = -1.60$. Intensities can be obtained from the log intensities by exponentiation (e.g., $e^{-1.60} = .20$).

Finally, like the initial state parameters, the transition parameters may be related to covariates, which may be either time-constant, such as scores from baseline

[62] For readers familiar with survival models, note that the intensities are actually equivalent to hazard rates (Cox & Miller, 1965; Kalbfleisch & Lawless, 1985; Kleinbaum & Klein, 2012).
[63] More specifically, the probabilities are equal to the matrix exponential of the product of the intensities and the interval (see Appendix N.2.1).

questionnaires, or time-varying[64], such as the negative event scores and the intervention that some subjects receive during their participation in our example data. The covariates are related to the transition parameters by means of regression (as is the case for the initial state parameters; for details, see Appendix N.2.1). Because the parameters of the transition model are log intensities, the regression effects have to be exponentiated to obtain the effects of the covariates on the transition intensities. However, it is more convenient to interpret the covariate effects on the transition probabilities. To this end, one can convert the intensities into probabilities for a certain interval length and different covariate values and compare them (as for the initial state probabilities).

6.2.3.2.4 *Transition Parameters for the Example Data*

To illustrate the interpretation of the transition parameters, consider again the LMFA results for our data, which are presented in LMFA output box 2. Note that we included the two covariates "had an intervention" and "negative event" on the transition parameters. This time, we focus on the parts "`Parameter estimates`" and "`Transition probabilities`". The "`transition parameters`" in the "`Parameter estimates`" part correspond to the log intensities for covariate scores being equal to zero. However, for better interpretability, we inspect the corresponding transition probabilities for a unit time interval for the average person in the sample and, thus, for covariate scores being equal to their sample means. These probabilities are displayed in the "`Transition probabilities`" part in LMFA output box 2. We can see that the sample mean for "had an intervention" is equal to .41 and the sample mean for "negative event" is equal to 49.65. The probabilities indicate that the probabilities to transition to another state are generally lower than to stay in a state, especially when staying in the displeasure state (i.e., state 1). The transition probabilities from the displeasure state (i.e., state 1) to the pleasure and the neutral state (i.e., state 2 and 3) are approximately equal. The transition probability from the pleasure state (i.e., state 2) to the neutral state (i.e., state 3) is smaller than from the pleasure to the displeasure state (i.e., state 1). Finally, the transition probabilities from the neutral state (i.e., state 3) to the displeasure state (i.e., state 1) are larger than to the pleasure state (i.e., state 2).

Next, we investigate the covariate effects on the transition probability matrix (for a unit interval). To this end, we make two comparisons. First, we keep the "negative

[64] Note that, for time-varying covariates, the score at time-point t is used to predict the transition probabilities from time-point $t-1$ to time-point t. This makes most sense for LMFA because MMs are typically triggered by momentary circumstances (e.g., social interactions). Even when assuming temporal precedence (e.g., the effect of perceiving a negative event prior to time-point t), questionnaires are usually designed in a way that they ask subjects about such covariates at the current time-point (e.g., "Please rate the unpleasantness of the most unpleasant event you have experienced since the previous measurement occasion").

event" score equal to the sample mean and compare the probabilities for both categories of "had an intervention". The probabilities pertaining to the "no intervention observations" are shown in LMFA output box 3.

LMFA output box 3

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 0
## negativeEvent score: 49.65
##
##      S1   S2   S3
## S1 0.84 0.07 0.08
## S2 0.37 0.44 0.19
## S3 0.54 0.08 0.39
```

The probabilities for the "intervention observations" are displayed in LMFA output box 4.

LMFA output box 4

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 1
## negativeEvent score: 49.65
##
##      S1   S2   S3
## S1 0.71 0.16 0.13
## S2 0.14 0.66 0.20
## S3 0.23 0.15 0.61
```

Comparing the transition probabilities, we see that having had an intervention is related to relatively smaller probabilities to transitioning to and staying in the displeasure state (i.e., state 1).

Second, we compare the transition probabilities for the sample mean of "negative event" minus the standard deviation (i.e., 34.54) to the transition probabilities for the sample mean of the covariate plus the standard deviation (i.e., 64.76), thereby keeping

the "had an intervention" score equal to the sample mean. The probabilities for a "negative event" score of 34.54 are displayed in LMFA output box 5.

LMFA output box 5

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 0.41
## negativeEvent score: 34.54
##
##      S1   S2   S3
## S1 0.74 0.14 0.13
## S2 0.21 0.55 0.24
## S3 0.33 0.13 0.53
```

The probabilities for a "negative event" score of 64.76 are shown in LMFA output box 6.

LMFA output box 6

```
## 1. Initial state probabilities:
##
## (no covariates defined)
##
##   S1   S2   S3
## 0.42 0.34 0.24
##
## 2. Transition probabilities:
##
## interval length: 1
## intervention score: 0.41
## negativeEvent score: 64.76
##
##      S1   S2   S3
## S1 0.84 0.08 0.09
## S2 0.31 0.51 0.18
## S3 0.45 0.08 0.46
```

Comparing the probabilities, we see that higher scores on "negative event" are related to larger probabilities of transitioning to and staying in the displeasure state (i.e., state 1).

6.2.3.3 *Summary of the LMFA Findings for our Example Data*

Summarized, based on the finding of LMFA, we conclude the following. First, the number and nature of the factors changed, which implies that configural invariance is violated for our example data. Second, subjects transitioned rather frequently between the states. However, transitioning to and staying in the displeasure state (i.e., state 1) was most likely, especially when experiencing negative events. However, the probabilities for

transitioning to and staying in the neutral or pleasure state (i.e., state 3 and state 2) increased after receiving an intervention.

6.2.4 Estimation

In *lmfa,* the maximum likelihood (ML) parameter estimates are obtained by means of the 3S estimation (Vogelsmeier et al., 2020), which builds on Vermunt's (2010) ML method and its extension for DT-LMM by Di Mari et al. (2016). The 3S estimation separates the estimation of the state-specific MMs and the CT-LMM as follows:

1. The state-specific MMs are estimated while disregarding the transitions between the latent states at consecutive measurement occasions and the covariate effects on these transitions.
2. Each observation is assigned to the state with the highest state-membership probability, that is, "modal state assignment" is applied.[65] Furthermore, the inherent classification uncertainty is calculated. Note that there is always uncertainty unless all observations are assigned to a state with a probability of 1.
3. The MMs (i.e., the factor parameters) are kept fixed and the state assignments from step 2 are used as single indicators for the estimation of the CT-LMM (with covariates), while correcting for step 2's assignment uncertainty. This correction is necessary to prevent underestimation of the relations between the states (i.e., the transition probabilities) and the covariate effects. Also note that the final state assignments will slightly differ from the step 2 state assignments (for details, see Appendix N.4 and N.5). Usually, the assignments improve because the step 3 estimation benefits from additional information from the transition model (with covariates) to classify the observations (Vogelsmeier et al., 2020).

For technical details about the steps, their likelihood functions, and the algorithms to maximize them, see Appendix N.3–N.5.

6.3 How to Conduct LMFA with the *lmfa* Package

In the following, we guide the readers through the different steps of conducting LMFA in the package *lmfa*. These steps are based on the three estimation steps described in Section 6.2.4: step 1 is investigating the MMs, step 2 is obtaining the state assignments and classification errors, and step 3 is investigating the transition model. Note that we introduce an additional step 0, which pertains to checking the data requirements prior to

[65] Note that it is theoretically also possible to use a "proportional assignment", which assigns the state memberships according to the posterior state-membership probabilities. However, the proportional assignment is unfeasible for data that contain a large number of measurement occasions for many subjects (Di Mari et al., 2016) and, thus, for ILD.

performing LMFA. Furthermore, as mentioned in Section 6.2.3, the best model complexity in terms of the number of states and factors is not known in advance and has to be evaluated in step 1. Additionally, depending on the subsequent analyses to investigate dynamics in psychological construct, researchers require factor scores corresponding to the state-specific MMs. Therefore, step 1 is divided into selecting the number of states and factors (step 1a), interpreting the MMs (step 1b), and attaching factor scores to the dataset (step 1c). Moreover, one has to decide which covariates should be included in the final transition model. Therefore, step 3 is divided into selecting covariates (step 3a) and interpreting the transition model and updating the final state assignments (step 3b). Figure 6.1 provides a summary of the steps with references to the required *lmfa* functions.[66] In the following, we describe the steps and functions by means of our example data introduced in Section 6.2.2. To follow the steps of this tutorial, the *lmfa* package and the example data have to be loaded into R. Before using the package for the first time, it has to be installed once, which can be done using the following command:

```
install.packages("devtools")
library("devtools")
install_github("LeonieVm/lmfa@0.1.0")
library("lmfa")
```

Note that the package *devtools* is required to install packages from the GitHub repository. The dataset used in this tutorial can be loaded into the R environment with the command:

```
data("ESM")
```


[66] Note that the package consists of only six functions in total (next to the general `summary()` and `plot()` functions). An overview of these functions is provided in Appendix M.

Step 0: Checking data requirements

- use a long data format for the data (e.g., called "`ESM`")
- remove empty records (i.e., skipped measurement occasions) from the data
- remove (or impute) records with partly missing observations

Step 1: Investigating measurement models

Step 1a: Selecting the number of states and factors

- run `step1()` with the modelselection option and store the results (e.g., into the object "`modelselection`")
- inspect the models using `summary(modelselection)` and rerun non-converged models using `step1()` without the model selection option
- choose best model(s) based on the BIC and CHull using `plot(modelselection)` and `chull_lmfa(modelselection)`
- store the best model(s) (e.g., into the object "`measurementmodel`")

Step 1b: Interpreting the measurement models

- use `summary(measurementmodel)` to obtain the measurement model parameters from the chosen model(s)
- investigate state-specific (obliquely rotated and standardized) loadings
- investigate state-specific intercepts
- investigate state-specific unique variances

Step 1c: Attach factor scores to the dataset

- store the outcome of `factorscores_lmfa(ESM, measurementmodel)` into a new dataset

Step 2: Obtaining state assignments & classification errors

- run `step2()` to calculate classification errors, posterior state-membership probabilities, and modal state assignments and store the results (e.g., into the object "`classification`")
- obtain the results with `summary(classification)`
- if desired, attach posterior state-membership probabilities and the modal state assignments to the dataset by storing `classification$data` into a new dataset

Step 3: Investigating transition model

Step 3a: Selecting the covariates for the transition model

- run `step3()` using the posterior state-membership probabilities of the `classification` object from `step2()` and all covariates of interest and store the results (e.g., into the object "`transitionmodel`")
- inspect Wald test results with `summary(transitionmodel)` and look at the p-values to decide which covariates should be included in the final transition model

Step 3b: Interpreting the transition model & updating state assignments

- use `summary(transitionmodel)` to obtain the transition model parameters and probabilities for covariates being equal to their sample means
- use `probabilities(transitionmodel)` to obtain initial state and transition probabilities for any covariate value (and interval length) of interest
- if desired, attach posterior state-membership probabilities and the modal state assignments to the dataset by storing `transitionmodel$data` into a new dataset

Figure 6.1. Summary of the three steps to conduct latent Markov factor analysis with *lmfa*.

6.3.1 Step 0: Checking Data Requirements

The first step is to check the data requirements with regard to the format (Section 6.3.1.1) and missing values (Section 6.3.1.2).

6.3.1.1 *Data Format*

In line with the assumed data structure, the data has to be in long format, that is, with rows equal to the number of total observations. Furthermore, next to the columns with the indicators of the latent factors (in our case, `"Interested"`, `"Joyful"`, `"Determined"`, `...`) and possibly covariates (in our case, `"intervention"`, `"negativeEvent"`), the data must contain a column with the subject identification numbers (in our case, `"id"`). Moreover, if observations should not be treated as equidistant, a column must be specified with the time intervals between two consecutive observations (in our case, `"deltaT"`). Regarding the latter, a proper unit should be used. For instance, if there is approximately only one observation per day, the unit "days" is appropriate (e.g., with an interval of 1.42 days representing an interval of one day and 10 hours). If there are several observations per day, say nine, "hours" is an appropriate unit. With "minutes" or "seconds" as unit, the intervals for these examples would take large values that likely lead to numerical problems when estimating the model.[67] Furthermore, measurement occasions within subjects must be ordered by time (i.e., intervals must not be negative). Additionally, for obtaining valid results, intervals for consecutive observations within a subject must not be equal to zero. Zero and negative intervals may occur from technical errors during data collection and should be removed (otherwise, an error message is displayed).

6.3.1.2 *Missing Data*

The data should only include the records for the measurement occasions at which the subjects completed the questionnaires because the CT-LMM automatically accounts for differences in the intervals, including skipped measurement occasions. Note that, depending on the data collection software, it may happen that a subject started a questionnaire but did not finish it such that some indicators or covariates contain missing values. These cases must be imputed (e.g., by means of the *mice* package in R; van Buuren & Groothuis-Oudshoorn, 2011) or removed before running *lmfa* (otherwise, an error message is displayed).

[67] The reason is that the value of the transition intensities (for covariate scores being equal to zero) are directly related to the size of the unit (note that the model fit is not influenced by the unit, however). More specifically, the larger the size of the unit, the smaller the intensities and the more likely numerical problems occur.

6.3.2 Step 1a: Selecting the Number of States and Factors

When estimating a LMFA model, the number of underlying states and factors per state has to be specified. However, in an exploratory approach like LMFA, the best model complexity is not known in advance and has to be determined by estimating a number of plausible models and comparing their results in terms of fit and parsimony. To this end, one can use criteria that balance the loglikelihood and number of parameters, such as the Bayesian information criterion (BIC; Schwarz, 1978) and the convex hull (CHull; Ceulemans & Kiers, 2006) method (Bulteel et al., 2013; Vogelsmeier, Vermunt, van Roekel, et al., 2019). In the following, we first describe the two criteria (Section 6.3.2.1). Then, we explain how to decide what range of states and factors to include in the model selection procedure (Section 6.3.2.2). Thereafter, we describe how to increase the chance of finding the "global" maximum and how to assess convergence of the estimation procedure (Section 6.3.2.3). Finally, we show how to perform the model selection with *lmfa* (Section 6.3.2.4).

6.3.2.1 *BIC and CHull*

First, the BIC considers model fit and model complexity of the model by penalizing models with more parameters (see Appendix N.6.2). Second, the CHull is a generalized scree test (Bulteel et al., 2013; Ceulemans & Kiers, 2006) that automatically identifies models at the higher boundary of the "convex hull" (or CHull) in a "loglikelihood vs. number of parameters" plot (Cattell, 1966) and that chooses the best model by finding the point (or "elbow") in this scree plot (or CHull plot) at which improvement in fit levels off when adding additional parameters to the model. Detecting this elbow is done by means of comparing "scree ratios" (see Appendix N.6.3) for all models on the upper boundary and the model with the largest ratio is chosen. In this way, the CHull also balances complexity and parsimony.

Both the BIC and the CHull offer valuable information about which model should be selected. However, for many real datasets, it is possible that the BIC keeps increasing when adding additional states and/or factors to the model (Bauer, 2007; McNeish & Harring, 2017). Then, investigating the relative improvement in the loglikelihood value by means of the CHull is especially important. Additionally, the CHull does not make distributional assumptions and may therefore perform better for many empirical datasets. The CHull method, however, has two drawbacks that should be accounted for. First, the least and the most complex models at the higher boundary of the CHull cannot be chosen because no scree ratios can be computed (see Bulteel et al., 2013). Therefore, it is always advisable to also inspect the CHull plot visually (e.g., the most complex model might still fit considerably better than the preceding model on the hull). The *lmfa* package will remind the user of this by displaying a note. Second, for some cases, it is possible that

the scree ratio is artificially inflated, even though the more complex model does not add much in terms of the fit. Specifically, when adding additional parameters hardly increases the fit anymore, both the numerator and denominator of the scree test ratio (Appendix N.6.3) approach zero, which results in a very large scree test ratio although the hull is pretty much a straight and horizontal line at that point (for a detailed explanation, see Wilderjans, Ceulemans, & Meers, 2013). The *lmfa* package displays a note if there are signs of artificial inflation. When the note is displayed, the user should inspect the CHull plot visually and also consider the next best model(s). Finally, it is best practice to look at the results of competing models and take the interpretability into account.

6.3.2.2 ***Range of States and Factors***

For the model selection, one must decide on the range of states and factors to be considered. Regarding the former, one may start with a few states (say, 1–3). If models with three states barely improve model fit (i.e., according to the BIC and CHull) or if the estimation of three states already causes estimation problems, there is no point in adding more states. Otherwise, one may increase the number of states. Moreover, the maximum number of states is restricted by the number of observations (i.e., one should have at least 1000 observations for each state; Vogelsmeier, Vermunt, van Roekel, et al., 2019). For instance, for our example dataset (with 4776 observations), we should not include more than four states. In order to decide on the number of factors, one should think about theoretically plausible factor structures and consider that each factor should ideally be measured by at least three items. Otherwise, the factors may not be well measured or "determined", which may cause convergence problems, Heywood cases (Van Driel, 1978) or less reliable parameter estimates. For example, if the data consist of six indicators, of which three are intended to measure PA and three are intended to measure NA, no more than two factors should be included. Additionally, similarly to the number of states, one should begin with a small number of factors and examine the increase in fit and convergence problems for the most complex factor structure.

6.3.2.3 ***Increasing the Chance to Find the Global Maximum and Assessing Convergence***

For estimating the state-specific FA models, the algorithm searches for the maximum of the loglikelihood function (Appendix N.3), that is, the solution with the largest loglikelihood value. However, it is possible that the solution is not a "global" maximum but a "local" one. To clarify this, consider the loglikelihood function as a landscape with multiple hills. Each hill has its own local maximum (i.e., the top) but only one hill is the highest and thus has the global maximum. To start searching for a global or local maximum, the algorithm requires initial parameter values. Different start values may lead to finding different (local) maxima (comparable to searching for the highest hill

starting from different locations in the landscape). Therefore, it is essential for the algorithm to use multiple start sets with different initial values and, in the end, provide the solution with the best loglikelihood value (Appendix N.3.5). Users should choose at least 25 start sets but the larger the number of start sets, the more likely it is to obtain the solution pertaining to the global maximum.

Moreover, it is possible that the model estimation does not converge at all. This means that the algorithm did not find a (local or global) maximum in a pre-specified number of maximum iterations (Appendix N.3). Especially for more complex models, it is possible that the algorithm requires more iterations to achieve convergence. However, it could also be a sign that the model is not suited for the data at hand (e.g., too many factors). The user may decide to re-estimate corresponding models once, allowing for more iterations, before continuing with the model selection procedure. The *lmfa* package displays this advice as a reminder.

6.3.2.4 *Model Selection with lmfa*

In order to select the "best" model among the models with different numbers of states and factors, we have to use *lmfa*'s `step1()` function. In the following, we compare models with one to four states and two to three factors per state (i.e., 14 models in total). Note that the order of factors does not matter because they result in the same fit and estimates. For instance, model [3 3 2] is the same as model [3 2 3] and model [2 3 3]. In the *lmfa* package, the permutation of the states of the estimated models is always determined based on the size of the states, starting with the largest. The function can be used as follows (because the estimations start from random state-membership assignments (see Appendix N.3.5), we set a seed for reproducibility):

```
set.seed(1000)
modelselection <- step1(data = ESM,
                        indicators = c(
                          "Interested",
                          "Joyful",
                          "Determined",
                          "Calm",
                          "Lively",
                          "Enthusiastic",
                          "Relaxed",
                          "Cheerful",
                          "Content",
                          "Energetic",
                          "Upset",
                          "Gloomy",
                          "Sluggish",
                          "Anxious",
                          "Bored",
                          "Irritated",
```

```
                          "Nervous",
                          "Listless"),
                        modelselection = TRUE,
                        n_state_range = 1:4,
                        n_fact_range = 2:3,
                        n_starts = 25,
                        max_iterations = 1000)
```

There are five mandatory arguments that we have to specify. First, we have to provide the data via the `data` argument (in our case, `ESM`). Second, via the `indicators` argument, we specify the variable names of the indicators in the same order as they appear in the data. These are `c("Interested", "Joyful", "Determined",...)`. Third, we indicate that we want to perform model selection via the argument `modelselection` (i.e., `modelselection = TRUE`). Fourth and fifth, we determine the range of states and factors that should be included in the model selection with `n_state_range = 1:4` and `n_fact_range = 2:3`. Additionally, we could change the default values for the number of start sets and the number of maximum iterations after which the estimation terminates regardless of whether convergence has been reached or not, but we simply use the default values `n_starts = 25` and `max_iterations = 1000`.[68]

When the estimation is terminated, we obtain the model-selection results as follows:

```
summary(modelselection)
```

Note that the model selection for our example data took about 3 hours. In order to follow the next tutorial steps in R, readers can simply load the model selection object with the command: `data("modelselection")`. The output is displayed in LMFA output box 7.

LMFA output box 7

```
##                LL      BIC convergence n_par
## [323]  -353166.8 708485.3           1   254
## [333]  -353149.0 708602.3           1   272
## [3322] -353071.7 708913.6           1   327
## [3233] -353065.6 709053.8           1   345
## [3333] -353016.6 709108.4           0   363
## [3222] -353316.0 709249.7           1   309
## [322]  -353855.3 709709.8           1   236
## [33]   -354421.0 710375.3           1   181
## [2222] -353962.6 710390.5           1   291
## [32]   -355010.3 711401.4           1   163
## [222]  -354986.3 711819.4           1   218
## [22]   -356377.4 713983.1           1   145
## [3]    -361759.6 724281.6           1    90
```

[68] Note that the function contains additional arguments related to the estimation procedure for which default values are provided. These values may be changed by the user if desired. For an explanation of the additional arguments, see Appendix M.1 and the function documentation, which can be called with `?step1`.

```
## [2]    -363744.0 728098.0           1    72
##
## Note: When re-estimating models that did not converge, the
##    number of maximum iterations should be increased.
```

The first column (i.e., "LL") pertains to the loglikelihood value. The second column (i.e., "BIC") shows the value of the BIC. The third column (i.e., "convergence") indicates whether the model estimation converged (with 1 = "convergence" and 0 = "non-convergence"). The fourth column (i.e., "n_par") shows the total number of parameters.[69] The models are ordered by the value of the BIC, starting with the lowest value and thus the model with the best fit according to this criterion.

Before continuing with the model selection, we check if models have to be re-estimated due to non-convergence. Indeed, the estimation of model [3 3 3 3] did not converge. For estimating single models, we use the `step1()` function but without model selection (i.e., with `modelselection = FALSE`). The code to estimate model [3 3 3 3] is:

```
set.seed(1000)
model3333 <- step1(data = ESM,
                   indicators = c(
                     "Interested",
                     "Joyful",
                     "Determined",
                     "Calm",
                     "Lively",
                     "Enthusiastic",
                     "Relaxed",
                     "Cheerful",
                     "Content",
                     "Energetic",
                     "Upset",
                     "Gloomy",
                     "Sluggish",
                     "Anxious",
                     "Bored",
                     "Irritated",
                     "Nervous",
                     "Listless"),
                   modelselection = FALSE,
                   n_state = 4,
                   n_fact = c(3,3,3,3),
                   n_starts = 25,
                   max_iterations = 2000)
```

[69] Note that the number of parameters are equal to the sum of the state-specific intercepts, unique variances, and loadings and the state proportions minus 1 (minus 1 because one state is treated as a reference state; see Appendix N.6.1).

When `modelselection = FALSE`, it is mandatory to provide a single number of states via the argument `n_state` (i.e., `n_state = 4`) and a vector with state-specific numbers of factors via the argument `n_fact` (i.e., `n_fact = c(3,3,3,3)`). As previously described, when re-estimating models that did initially not converge, it is wise to increase the number of maximum iterations. Therefore, we set `max_iterations = 2000`. In order to replace the old by the new models, the following command can be used:

```
modelselection$`[3333]` <- model3333
```

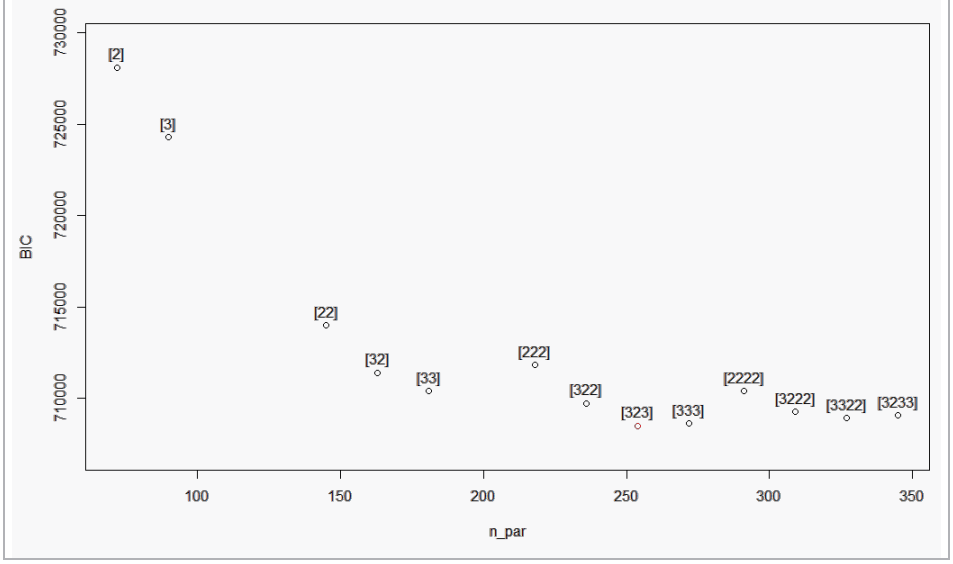
However, the model did not converge (it might simply not be suitable for the data) and, therefore, we continue with the original model selection object.

From the summary in LMFA output box 7, we can already see that the best model according to the BIC is model [3 2 3] and, thus, the true model. For an easier inspection of the results, we also plot the BIC of the converged models against the number of free parameters. The plot is readily available by running the following command:

```
plot(modelselection)
```

The output is shown in LMFA output box 8.

LMFA output box 8



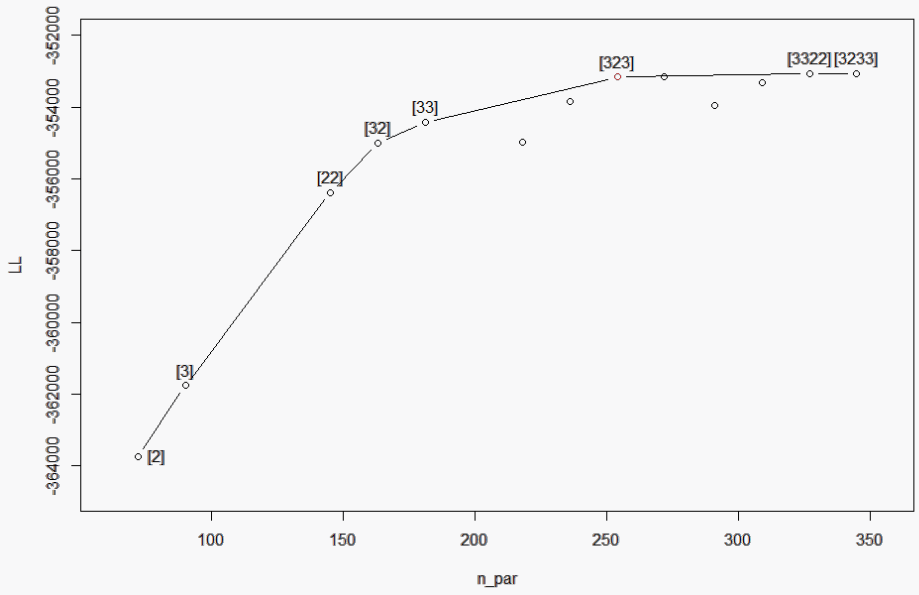
The model corresponding to the lowest BIC value is indicated by a red dot. Note that, for our example, the BIC does not keep increasing for more complex models. Therefore, we would consider it relatively save to choose the model with the lowest BIC value. However,

in order to support our choice, we also investigate the results of the CHull method for the converged models, which can be obtained with the `chull_lmfa()` function as follows:

```
chull_lmfa(x = modelselection)
```

We only have to specify argument `x`, which pertains to the model-selection object (in our case, `modelselection`).[70] The output is shown in LMFA output box 9.

LMFA output box 9




```
## Models on the upper boundary of the CHull:
##
##        n_par        LL        st
## [2]       72 -363744.0        NA
## [3]       90 -361759.6  1.126582
## [22]     145 -356377.4  1.288490
## [32]     163 -355010.3  2.319779
## [33]     181 -354421.0  1.905623
## [323]    254 -353166.8 13.191076
## [3322]   327 -353071.7  3.812062
## [3233]   345 -353065.6        NA
##
## Selected model(s):
##
##       n_par        LL
## [323]   254 -353166.8
##
## Note 1: The least and most complex models cannot be selected.
## Therefore, it is advisable to also visually inspect the CHull plot.
```

[70] For details about the function, see Appendix M.4 or call the documentation file with `?chull_lmfa`.

```
##
## Note 2: The st value(s) of the best model(s) might be artificially
## inflated. Therefore, it is advisable to also consider the next best model(s).
```

The output consists of three parts, the CHull plot, the summary of the models on the upper boundary of the CHull (including their scree-test values "st"), and the selected model(s). We can see that the model [3 2 3] was selected. However, we got the note that the scree value might be artificially inflated. Therefore, in practice, one should also consider the results of the next best model(s). Because we know the true model, we continue with model [3 2 3].[71] In order to inspect the model, we have to extract it from the model-selection object `modelselection` and store it as follows:

```
measurementmodel323 <- modelselection$`[323]`
```

The parameters can be displayed with the command:[72, 73]

```
summary(measurementmodel323)
```

6.3.3 Step 1b: Interpreting the Measurement Models

Once the best model in terms of the number of states and factors is selected, we can interpret the state-specific MMs. The output (obtained with `summary(measurementmodel323)`) was already displayed in LMFA output box 1 and the results were already interpreted in Section 6.2.3.1. To briefly summarize them, we found three states (a displeasure, a neutral, and a pleasure state) that all contained a positive affect and a negative affect (or distress) factor but the displeasure state was additionally characterized by a drive factor and the neutral state by a serenity factor.

6.3.4 Step 1c: Attach Factor Scores to the Dataset

Before proceeding with step 2, we can attach state-specific factor scores to our dataset for each observation in the dataset.[74] The factor scores are estimates of the latent

[71] The second-best model according to the scree value is model [3 3 2 2]. However, visual inspection of the CHull shows that adding a fourth state does not considerably improve the fit. The next best model is model [3 2]. If we would inspect model [3 2], we would see that the smallest state (i.e., the neutral state) would be divided into the pleasure and the displeasure state.

[72] Note that, in the `summary()` function, the user can specify an additional argument to change the number of decimals to which the parameters should be rounded. The default for the summary of the MM parameters is `rounding = 2`.

[73] Note that the model object (i.e., `measurementmodel323`) contains additional information that is not directly relevant for the interpretation but that may be interesting for some users (e.g., unrotated and unstandardized loadings). For an overview of all available output, see Appendix M.1.

[74] In *lmfa*, the factor scores are calculated by means of the regression method (Thomson, 1934; Thurstone, 1935), which is one of the most commonly used approaches. For the exact computation, see Appendix N.6.8. Note that the calculation of factor score estimates is generally considered controversial because different methods can result in (very) different scores (which is referred to as the problem of factor score

constructs and can be used for subsequent analyses to investigate dynamics in psychological constructs (for suggestions on how to proceed in the presence of non-invariance, see Section 6.4). A copy of the dataset with the factor scores attached can be obtained with:[75]

```
ESM_fs <- factorscores_lmfa(data = ESM, model = measurementmodel323)
```

In this function, two arguments are required. First, via the argument `data`, we have to provide the data that was used for the `step1()` estimation (in our case, `ESM`). Second, via the argument `model`, we have to specify the `step1()` object with the state-specific MMs (in our case, `measurementmodel323`). In the resulting dataset (i.e., `ESM_fs`), the columns are called "`S1F1`", "`S1F2`", etc., where the "`S`" refers to the state and "`F`" to the factor.

6.3.5 Step 2: Obtaining state assignments & classification errors

The next step is to obtain information about the classification and the (modal) state assignments. In this section, we first describe how to obtain the results for our example data with *lmfa* and then, based on the output, we explain the different classification statistics.

In order to obtain the classification information, we use the `step2()` function as follows:

```
classification <- step2(data = ESM_fs, model = measurementmodel323)
```

The function contains two arguments that we have to specify. First, we have to provide the data that was used for the `step1()` estimation via the argument `data`. It is most convenient to use the version including the factor scores estimates (in our case, `ESM_fs`) because we will add additional columns later on and this allows us to obtain a complete dataset for further analyses. Second, we need to specify the `step1()` object with the state-

indeterminacy; for discussions on this and possible solutions to account for biases in subsequent analyses, see, e.g., Devlieger et al., 2016; Green, 1976; Grice, 2001).

[75] Note that, in the `factorscores_lmfa()` function, the user can specify two additional arguments. The first one indicates whether the factor score estimates should be obtained for the obliquely rotated loadings. The default is `oblique = TRUE`. Otherwise, the factor score estimates corresponding to the unrotated factor loadings are obtained. The second argument pertains to the number of decimals to which the factor score estimates should be rounded. The default is `rounding = 4`. For details about the function, see Appendix M.5 or call the documentation file with `?factorscores_lmfa`.

specific MMs via the argument `model` (in our case, `measurementmodel323`).[76] The following code prints the results:[77]

```
summary(classification)
```

The output is shown in LMFA output box 10.

LMFA output box 10

```
## R2_entropy: 0.86
##
## Total classification error: 0.05
##
## Classification errors:
##
##         S1      S2     S3
## S1 2568.11    4.12  59.50
## S2    2.71 1146.53  44.37
## S3   50.19   89.35 811.13
##
## Classification-error probabilities:
##
##      S1   S2   S3
## S1 0.98 0.00 0.02
## S2 0.00 0.96 0.04
## S3 0.05 0.09 0.85
##
## State proportions:
##
##   S1   S2   S3
## 0.55 0.26 0.19
```

First, the R-square measure $R^2_{entropy}$ (called "`R2_entropy`" in the output) indicates how well the states are separated (and thus how much the MMs differ) with values ranging from 0 (bad separation) to 1 (good separation; Lukočienė et al., 2010). Note that a larger state separation implies less classification error. It is important to inspect the $R^2_{entropy}$ value because a bad state separation (with $R^2_{entropy} < .5$) can lead to an incorrect classification-error correction [78] and, in turn, to an underestimation of transition probabilities and the covariate effects (Vermunt, 2010). When observing a bad state separation, which is rather unlikely in practice, it is advised to use the FIML estimation, which is currently only available in Latent GOLD. The $R^2_{entropy}$value for our example data

[76] For an additional explanation of the function arguments, see Appendix M.2 or call the documentation file with `?step2`.

[77] Again, the user can adjust the number of decimals to which the parameters should be rounded. The default for the summary of the classification outputs is `rounding = 2`.

[78] The reason is that the $R^2_{entropy}$ tends to be overestimated for bad state separations (e.g., if we find a value of .4, the real state separation is probably even lower). In turn, the classification errors are underestimated, leading to an incorrect correction in the final step of the analysis (Vermunt, 2010).

indicates that the states are well separated, which explains the small total classification error (called "`Total classification error`" in the output).

Second, information about the classification errors can be obtained from the classification error matrix (called "`Classification errors`" in the output), which cross-classifies the modal state assignments by the "true" state assignments and which is used to correct for the error in step 3 of the analysis (for details, see Appendix N.4). Higher values on the diagonal and lower values on the off-diagonal are indicative of less classification error. For an easier interpretation, the counts can be translated into proportions (called "`Classification-error probabilities`" in the output). Inspecting the classification error matrices, we see that the classification error is lowest in the displeasure state (i.e., state 1), followed by the pleasure state (i.e., state 2) and the neutral state (i.e., state 3). Thus, the classification into the neutral state was accompanied by the most uncertainty, which is not surprising, given that the neutral state lies somewhat in between the displeasure and pleasure state.

Third, the state proportions (also called like that in the output; i.e., "`State proportions`") pertain to the overall state sizes. Looking at the state proportions for our data, we see that the displeasure state is the largest, followed by the pleasure and the neutral state.

Finally, the state assignments are not displayed in the output because R cannot display all assignments simultaneously. However, we can simply obtain a copy of our dataset with additional columns corresponding to the state assignments with the following command:

```
ESM_fs_cl <- classification$data
```

Specifically, the columns with the posterior state probabilities (called "`State1`", "`State2`" etc. in the dataset) indicate the probabilities for an observation to belong to a certain state and, thus, that the state-specific MM underlies the responses for this observation. As explained in Section 6.2.4, the modal state assignments (called "`Modal`" in the dataset) correspond to the state with the largest probability and, hence, to the most likely state membership.

6.3.6 Step 3a: Selecting the Covariates for the Transition Model

When the state-specific MMs are obtained and the observations are assigned to the states, we can continue with investigating the transitions between the states and what may cause them by means of estimating a LMFA with covariates on the initial state and/or transition parameters. In order to test if a covariate is significantly related to the transition model parameters (and thus whether it should be included in the model) can

be evaluated with Wald tests (Agresti, 1990). In the following, we first explain the covariate selection with the Wald tests (Section 6.3.6.1) and then show how to perform the covariate selection for our example data with *lmfa* (Section 6.3.6.2).

6.3.6.1 *Covariate Selection Procedure Using Wald Tests*

Every covariate in the model is accompanied by separate covariate effects on the initial state or transition parameters (e.g., the covariate "had an intervention" has six effects, one on each of the transition parameters). The Wald tests in *lmfa* are omnibus tests that show whether including a covariate is significant overall (i.e., across the initial and transition parameters). Thus, for every covariate, there is one Wald-test statistic. In order to select which covariates to include, one can start with a LMFA with all covariate candidates in it. Then, the least significant covariate is removed and the model gets re-estimated. This "backward selection" continues until only significant covariates are left (say, according to an alpha level of .05). When only significant covariates are left in the model, one can continue with the interpretation of covariate effects on the transition probabilities (as we did for our example in Section 6.2.3.2). Note that, instead of using such a data-driven approach, a more theory-driven approach is also possible (e.g., investigating the significance and effects of a set of covariates that was previously found to be significantly related to the transition model parameters).

6.3.6.2 *Covariate Selection with lmfa*

In the following, we estimate a transition model with covariate effects of "had an intervention" and "negative event" on the transition parameters. In order to estimate the transition model, we use the `step3()` function as follows (because the estimation starts from random values for the transition parameters (see Appendix N.5.4), we set a seed for reproducibility):

```
set.seed(1000)

transitionmodel <- step3(data = ESM_fs,
                         identifier = "id",
                         n_state = 3,
                         postprobs =
                           classification$classification_posteriors[,-1],
                         timeintervals  = "deltaT",
                         initialCovariates = NULL,
                         transitionCovariates =
                           c("intervention", "negativeEvent"),
                         n_starts = 25,
                         max_iterations = 1000)
```

There are four mandatory arguments that we have to specify. First, we provide the data via the `data` argument. We use the dataset `ESM_fs` and, thus, the data including the factor

scores but without the state assignments from step 2 because they are updated in step 3. Second, we specify the name of the column with the subject identification numbers via the argument `identifier` (in our case, `"id"`). Third, we define the number of states with `n_state = 3`. Fourth, we specify the posterior state probabilities by means of the argument `postprobs`. The probabilities can be extracted from the `step2()` classification output with the command `classification$classification_posteriors[,-1]`, where `[,-1]` indicates that we leave out the column with the modal state assignments.

The following three arguments are not required but have to be specified if the model should account for differences in intervals and if covariate effects on the transition model parameters should be included. Both applies to our example. Thus, first, via the argument `timeintervals`, we provide the function with the name of the column in the dataset that contains the time intervals. In our case, this is `"deltaT"` (if no such column name is provided, observations would be assumed to be equidistant). Next, via the arguments `transitionCovariates` and `initialCovariates`, we can provide (a vector of) column names that contain the covariate scores (the default for both arguments is `NULL`, i.e., no covariates are used). Thus, for our analysis, we provide the vector `c("intervention", "negativeEvent")` as input for `transitionCovariates`.

Finally, similarly to the `step1()` function, the users may decide to change the default values pertaining to the number of start sets[79] via the argument `n_starts` and the number of maximum iterations via the argument `max_iterations`[80]. However, for our analysis, we simply use the default values, that is, `n_starts = 25` and `max_iterations = 1000`.[81] After termination of the estimation, the results are obtained as follows:[82]

```
summary(transitionmodel)
```

The estimation for our example data took about 20 minutes. Again, readers who want to follow the rest of the tutorial can also load the results with the command: `data("transitionmodel")`. The results are shown in the "`Wald tests`" part in LMFA output box 2. For each covariate, we get a significance test with the corresponding Wald-test statistic (i.e., "`Wald`"), the degrees of freedom (i.e., "`df`"), and the p-value (i.e., "`p-`

[79] Note that the results of the `step3()` function are sensitive to start values for the transition intensities (see Appendix N.5.4). Therefore, one should use at least 25 start sets.
[80] If the maximum number of iterations is reached without convergence, *lmfa* displays a note with the advice to repeat the estimation with an increased number of `max_iterations`.
[81] Note that the user may change the defaults of additional arguments pertaining to the estimation procedure. For an explanation of these arguments, see Appendix M.3 and the function documentation that can be called with `?step3`.
[82] Again, the user can change the number of decimals to which the parameters should be rounded. The default for the summary of the transition model is `rounding = 4`. Thus, by default, there are two more decimals than for the other representations. This is because some parameters can get very small such that differences would vanish too easily when using less decimal points.

value"). We see that both covariates have significant effects on the transition parameters. Thus, we keep both covariates in the model.

6.3.7 Step 3 b: Interpreting the Transition Model & Updating the State Assignments

After selecting the covariates for the transition model, we can interpret the effects on the probabilities, investigate changes in the final state proportions, and obtain the final state assignments. In the following, we show how to obtain the relevant output in *lmfa* and how to interpret the results for our example data. First, as previously shown, the regression parameters and the probabilities for covariates being equal to their sample means (and a unit interval) can be obtained with `summary(transitionmodel)`. In order to obtain the initial state and transition probabilities for any covariate score and interval of interest, we can use the `probabilities()` function. For example, in order to obtain the probabilities for a "had an intervention" score equal to zero, a "negative event" score equal to the sample mean of 49.65, and a unit interval, we use the following command:

```
probabilities(model = transitionmodel,
              deltaT = 1,
              initialCovariateScores = NULL,
              transitionCovariateScores = c(0, 49.65))
```

Only the first argument is mandatory, that is, we have to provide the output of the `step3()` function (in our case, `transitionmodel`) via the `model` argument. By default, the function prints the probabilities for a unit interval and covariate scores equal to the sample means (i.e., `deltaT = 1`, `initialCovariateScores = NULL` and `transitionCovariateScores = NULL`). In order to print the probabilities for specific covariate scores, we have to provide a vector with these scores in the same order as we included the covariates in the estimation of the transition model with the `step3()` function. In our case, we include `transitionCovariateScores = c(0,49.65)`.[83, 84] The results for the transition model with covariates were interpreted in Section 6.2.3.2. To briefly summarize them, first, most subjects started in the displeasure state. Second, the probabilities to stay in a state were generally higher than to transition to another state (especially for subjects in the displeasure state). Third, the intervention was related to transitions to a more positive state and experiencing negative events was related to transitioning to the more negative state.

[83] An additional explanation of the arguments can be found in Appendix M.6 and in the function documentation, which can be called with `?probabilities`.
[84] The user can determine an additional argument in the `probabilities()` function to change the number of decimals to which the parameters should be rounded. The default is `rounding = 2`.

Next, the final state proportions are shown under "`State proportions`" in the same output as the transition model parameters (see LMFA output box 2). Comparing the results to the state proportions in LMFA output box 10 (and thus to the state proportions resulting from the modal state assignment in step 2), we see that there was no change. This is not surprising considering the small classification errors from step 2.

Third, we can again obtain a copy of our dataset with the posterior state-membership probabilities and the modal state assignment attached. To this end, we use the following command:

```
ESM_fs_cl <- transitionmodel$data
```

These are the state assignments that should be considered for subsequent data analyses because, as described in Section 6.2.4, they may be more accurate than the step 2 assignments. Therefore, we simply overwrite the previous dataset with the step 2 assignments `ESM_fs_cl`.

6.4 Proceeding Based on the Results of LMFA

Once the MM differences and possible explanations are known, the question is of course: Based on the LMFA results, how to proceed with (originally planned) analyses to investigate the dynamics in psychological constructs? The answer to this question largely depends on the findings. It is important to note that a comparison of the state-specific MMs may indicate violations of different levels of invariance and that the required level of invariance depends on the type of comparisons one wants to make. When comparing the state-specific loadings, one may find that the MMs differ a lot across states—specifically, in the number and/or nature (the zero-loading pattern) of measured constructs—which indicates a violation of configural invariance. It may also be that the pattern of (near-)zero loadings appears to be equal across states, but that the non-zero loadings differ in size. This suggests that configural invariance holds but weak invariance fails. When configural or weak non-invariance is indicated, continuing with analyses that assume invariance is not possible for the full dataset because factor scores are not validly comparable. Differences in the means of the constructs or relations between constructs could be due to underlying differences in the MMs. Finding such differences in MMs is interesting in its own right (e.g., the additional drive factor for anhedonic subjects in our data example), however. In any case, it is possible to proceed with factor scores from one

specific state (e.g., the largest state or the state that best corresponds to an a priori assumed MM) and, thus, with observations for which strict invariance holds.[85]

If weak invariance holds across the states—that is, if the (near-)zero and non-zero loadings are highly similar across states—users may examine whether covariances (e.g., regression coefficients or autocorrelations) between latent constructs (e.g., positive affect and negative affect) differ across subjects and/or change across time, because factor covariances are not affected by intercept differences (Oberski, 2017; Steenkamp & Baumgartner, 1998). However, examining whether mean construct scores differ across subjects and time-points calls for strong invariance to avoid mixing up differences in latent means and intercept differences (Meredith & Teresi, 2006). This implies that strict invariance is not necessary for meaningfully comparing latent covariances or means (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). Thus, finding states that differ in the unique variances only does not preclude latent variable comparisons. Note that it is best to allow for non-invariances of intercepts or unique variances (as indicated by LMFA) as much as possible in your follow-up analysis, ideally by including states.[86] Otherwise, the latent means and/or covariances may be estimated incorrectly, especially in the case of large non-invariances (Chen, 2008; Guenole & Brown, 2014). Alternatively, one could perform one analysis per state (using the state-specific factor scores) and weight the observations according to the posterior state-membership probabilities such that observations with larger probabilities receive more weight than observations with lower probabilities. Another option could be to conduct a weighted multilevel analysis, in which the states would be considered as observed groups. Furthermore, if "partial" metric or strong invariance holds (i.e., if only a few loadings or intercepts differ; Byrne et al., 1989), one may exclude non-invariant items or, again, capture the differences by letting parameters differ across states in subsequent analyses or dealing with it by conducting separate analyses with weighted data. Moreover, in order to avoid non-invariance in future studies, one could consider removing the problematic items from the questionnaire or rephrase them. To conclude, LMFA can be viewed as a primary analysis

[85] When proceeding with factor scores from one state, it is important to investigate the certainty of the final state assignments. For instance, the largest posterior state probabilities for some observations might be rather low (say, below .6 or .7), indicating that it is less clear which of the state-specific MMs best fits the scores of the observation. Therefore, it is advisable to remove these observations before conducting further analyses.

[86] Note that this is possible with more advanced analyses like dynamic latent class analysis (DLCA; Asparouhov, Hamaker, & Muthén, 2016). Moreover, in some analyses, it is at least possible to allow for MM differences across subjects like in dynamic structural equation modeling (DSEM; McNeish & Hamaker, 2020; McNeish, Mackinnon, Marsch, & Poldrack, 2021).

step that indicates which observations are actually comparable, what the MMs look like, and that, in turn, facilitates decisions about how to further analyze the ILD.

6.5 Discussion

When studying dynamics in psychological constructs in intensive longitudinal data (ILD), it is crucial to investigate whether the measurement models (MMs) underlying the responses are invariant across subjects and time, which is easily violated due to between-person differences and/or situation-specific changes in item interpretation and response styles. Undetected measurement non-invariance poses a threat to valid inference from state-of-the-art ILD analyses. In this tutorial, we showed how to explore which MMs underlie the data, what transitions between these MMs look like, and how to investigate if covariates are related to such transitions by means of latent Markov factor analysis (LMFA). Most importantly, the method indicates which observations are actually comparable by classifying them into the same MM-state, which helps to safeguard valid inferences. Moreover, researchers gain substantive insights into the dynamics of the underlying MM in their ILD.

The package *lmfa* was implemented in the open source software R to provide researchers with a freely available software option to perform LMFA. Even though this is a huge advantage, it is important to stress that some features are currently not (yet) available. In the following, we will elaborate on the current limitations of the package and ideas for future extensions.[87] Firstly, the state-specific MMs in step 1 of the analysis are currently obtained by means of exploratory factor analysis (EFA). As previously explained, EFA is less restrictive than confirmatory factor analysis (CFA), which implies that it allows to detect all types of non-invariance in the loadings. However, for some datasets, it is certainly interesting to use a CFA model—thus, with fixed patterns of zero factor loadings. For example, researchers may want to rely on LMFA results or other results from previous research showing that the configural factor structure is rather stable across subjects and time. Therefore, in the future, *lmfa* will include an option to perform CFA within the states.

Secondly, the factor analysis models in step 1 assume continuous item responses. If items are measured with only a few categories or if the item responses are heavily skewed, state-specific "latent trait" (or "item response theory") models should be employed in step 1 of the analysis to adequately deal with categorical data, as is done in the extension called latent Markov latent trait analysis (LMLTA; Vogelsmeier et al., 2021). Performing LMLTA is currently only possible in Latent GOLD, but advanced R users could

[87] Note that the commercial software Latent GOLD offers most of the features discussed below.

theoretically specify their own state-specific models for step 1 (for instance, by using mixture models for categorical data from other packages) and use the posterior state-membership probabilities as input[88] for the step 3 analysis with *lmfa*. However, suitable packages are, to the best of our knowledge, currently not available in R. If a package would become available, possibilities to include it in the *lmfa* package (in order to perform LMLTA) will be examined.

Thirdly, prior to performing step 1 in *lmfa*, users have to remove or impute records that contain missing values on some of the indicators, that is, for measurement occasions that were not completely skipped (note that completely omitted measurement occasions are dealt with by the continuous-time approach). Generally, technological advances in many experience sampling methodology apps prevent subjects from submitting incomplete responses. However, sometimes researchers rather have incomplete data than to lose the measurement occasion entirely. Furthermore, missing data may be a result of the increasingly employed "planned missing-data designs", in which researchers deliberately assess only selected items at each measurement occasions while omitting others in order to reduce the burden on the subjects, which, in turn, tends to increase the quality of the responses (Silvia, Kwapil, Walsh, & Myin-Germeys, 2014; van Roekel et al., 2019). In the future, *lmfa* will be extended to be applicable for ILD collected with such innovative missing-data designs and missing data in general.

Fourthly, the inclusion of covariates in step 3 of the analysis helps researchers understand why certain subjects transition between MMs over time, but some researchers might be more (or also) interested in individual transition patterns, especially in case of only a few subjects. Estimating subject-specific transition parameters is currently not possible in *lmfa*. However, one can estimate one transition model per subject. More specifically, step 1 and 2 (i.e., estimating the MMs and obtaining the state assignments and classification errors) would still be conducted for all subjects[89] but step 3 would be performed for each subject individually. Additionally, instead of inspecting subject-specific transition parameters, it might be interesting to investigate whether unobserved subgroups of subjects have similar transition patterns, especially in case of many subjects. Theoretically, it is possible to cluster subjects based on their transition behavior by adding a latent grouping variable to the LMFA in step 3 (e.g., see Crayen et al., 2017; Vogelsmeier et al., 2021). This is not possible with *lmfa* but advanced R users may consider using the *depmix* package in step 3 of LMFA by passing the modal state

[88] As described in Section 6.3.6.2, this would be done by providing a data frame with posterior state-membership probabilities per state via the argument `postprobs`.
[89] One may also perform the entire LMFA for data of a single subject if the number of observations is large enough. For guidelines about the required number of observations, see Vogelsmeier, Vermunt, van Roekel, et al. (2019).

assignments and classification error probabilities as fixed parameters to the `depmix()` function. This package allows for a latent grouping variable in the transition model but uses a discrete-time latent Markov model and, hence, does not account for differences in intervals, which may impair the estimation of the transition model parameters when intervals are unequal (for details about the syntax and about how to fix parameters, see the package documentation; Visser, 2007).

Lastly, *lmfa* users currently have to draw conclusions about (non-)invariance by visually comparing the state-specific MMs. If the number and nature of the factors appear to be the same across states, determining which parameters differ substantially becomes a daunting task, especially when comparing parameters for models with many states and factors. Furthermore, one will always find small parameter differences across states due to sampling fluctuations and error fitting. Deciding which differences are practically or statistically significant is not a trivial problem. On top of that, the states also capture differences in the factor variances (in the loadings) and factor means (in the intercepts) due to the model identification constraints (see Section 6.2.3.1). To obtain loadings that are optimally comparable across states and to enable hypothesis testing for these loadings (using Wald tests), multigroup factor rotation (MGFR; De Roover & Vermunt, 2019) should be applied. MGFR solves rotational freedom by rotating the loadings towards a simple structure within the states and towards agreement across states while unraveling differences in the loadings from differences in the factor variances. MGFR is currently only available in Latent GOLD but possibilities to include it in *lmfa* will be investigated in the future. Similarly, a solution for optimally comparing intercepts (with hypothesis tests) could be to employ multiple group factor alignment (MGFA; Asparouhov & Muthén, 2014b), in which the factors are rescaled and shifted (or "aligned") with respect to their means, thereby disentangling differences in the intercepts from differences in the factor means. However, currently, MGFA is only applicable to CFA models without cross-loadings. If an MGFA extension for EFA would become available, possibilities to include the method in *lmfa* will be examined. Until MGFR and MGFA are implemented, users can inspect whether there appears to be a difference in the scaling of all loadings of a factor and/or a "shift" in all intercepts of items that correspond to the same factor (as indicated by loadings that are not close to zero). If separate loadings or intercepts differ across states, it is unlikely that these differences are caused by differences in the underlying factor variances or factor means, respectively.

7

Epilogue

In this dissertation, we presented latent Markov factor analysis (LMFA) as a powerful method for tracing (non-)invariance of the measurement model (MM) in (intensive) longitudinal data both across subjects and time. In Chapter 2, we introduced the basic version of LMFA for continuous item responses and equal intervals between measurement occasions. Then, in Chapter 3, we extended LMFA to validly accommodate unequally-spaced observations, which are more common in intensive longitudinal data (ILD) assessed by means of experience sampling methodology (ESM). In Chapter 4, we showed how to include explanatory variables to better understand (differences in) the dynamics of the MM. In Chapter 5, we introduced latent Markov latent trait analysis (LMLTA), the extension that adequately handles categorical item responses with only a few categories and/or skewed response distributions, which are often observed when employing Likert-type response scales. In Chapter 6, we presented a tutorial for conducting LMFA in the open-source R package *lmfa*.

Throughout the dissertation, we showed why it is important and interesting to trace MM differences across subjects and time by means of several applications. For example, one of the main findings in the application in Chapter 2 was that the responses by anhedonic individuals with an overall negative mood were not only characterized by the originally assumed positive and a negative affect factors but also by an additional "drive" factor. Furthermore, in Chapter 3, we investigated longitudinal data from patients with a major depression disorder and found that the dimensionality of the underlying factors reduced over the course of participation. Moreover, in the application in Chapter 4, we added to the theoretical debate whether positive and negative affect are two separate factors or rather two bipolar ends of a single factor by showing that, for the adolescents in the analyzed ILD, both perspectives occurred at different points in time, for instance, related to the overall mood.[90]

Despite taking important first steps in developing and evaluating LMFA, a few questions remain open, some limitations of LMFA and this dissertation in general have not been critically evaluated yet, and relatedly, not all relevant directions for future extensions of LMFA have been presented. In the following, we first address these points (Sections 7.1 and 7.2) and then conclude with a remark about LMFA as a framework for studying dynamic psychological phenomena (Section 7.3).

7.1 Answers to Open Questions About LMFA

We see two questions that have not been answered sufficiently throughout the dissertation. That is, (1) how to evaluate which level of invariance holds across states

[90] It is important to note that the analyses were entirely exploratory and that the empirical data examples were solely used to demonstrate the method and its extensions. Therefore, one should not draw substantive conclusions without validating the findings, for instance, by means of replication studies.

while disentangling differences in the MM parameters from differences in factor means and variances, and (2) whether response styles can be identified as such.

7.1.1 Evaluating Higher Levels of Invariance

When applying LMFA to ILD at hand, researchers initially obtain one of two possible results. Either, LMFA indicates that the best model complexity pertains to a single state, implying that the highest level of invariance holds across all observations (i.e., the MM is invariant regarding loadings, intercepts, and unique variances). Or, LMFA concludes that at least two states are underlying the data, in which case researchers need to determine which level of invariance is endorsed across states. This is relevant for valid decisions about subsequent analyses (generally, the higher the level of invariance, the more meaningful comparisons can be made; for a detailed discussion, see Chapter 6). The psychometric literature distinguishes between four successive levels of invariance (Meredith, 1993). The first level is configural invariance, which applies when the number of factors and patterns of (close to) zero loadings are the same across states (i.e., whether the same factors are measured across states). The second level is weak invariance, which is attained when the size of the loadings is also identical across states. The third level is strong invariance, which applies when, additionally, the intercepts are the same across states. The fourth level is strict invariance, which holds when the unique variances are also identical across states.

When configural invariance does clearly not hold (like in our applications where different constructs were measured across states), one can stop with further invariance investigations. However, if configural invariance seems to hold, researchers should continue investigating whether further levels of invariance are attainable. Note that it is possible that "full" invariance cannot be endorsed but, if only a few loadings, intercepts, or unique variances differ across states, "partial" weak, strong or strict invariance would still be tenable (Byrne et al., 1989). As described in Chapter 6, partial invariance does not preclude subsequent analyses as long as non-invariance is accounted for.[91]

In order to evaluate which level of invariance holds in the applications of this dissertation, we used visual inspection of the state-specific MMs. Although visual inspection provides relevant first insights into the tenability of different levels of invariance, determining which parameters differ significantly is cumbersome, especially when examining differences in parameters for many states and factors. One will always find some degree of heterogeneity across states due to sample fluctuations and error fitting. Additionally, rotating towards simple structure per state, differences in loadings

[91] Note that there are no clear guideline about the minimum number of loadings, intercepts, or unique variances that should be invariant (as will be discussed in Section 7.2.1, this should be investigated for state-of-the-art ILD analyses). For now, one may follow the general rule of thumb that the majority of the parameters should be invariant (Putnick & Bornstein, 2016).

across states can be overrated (De Roover & Vermunt, 2019). An alternative to evaluate higher levels of invariance has not yet been adequately addressed in this dissertation.

Generally, it is possible to evaluate different levels of invariance by progressing from the baseline model (i.e., the unrestricted exploratory factor analysis (EFA) models per state) towards more restricted models (i.e., partially or fully invariant models) and comparing (changes in) their goodness of fit to evaluate if invariance holds (e.g., Jöreskog, 1971; Lubke & Muthén, 2005; Sörbom, 1974). One could treat the states as observed groups and apply the increasingly more restrictive models to the states with standard multigroup factor analysis (e.g., Dolan et al., 2009; Jöreskog, 1971; Sörbom, 1974). However, this approach is rather naïve because the uncertainty of the state assignments is disregarded. Moreover, restricting parameters in the MMs changes the nature of the states (i.e., the focus is shifted to other differences) such that the posterior state-membership probabilities should be allowed to change as well, which is not possible when treating the states as observed. For both reasons, the invariance results could be invalid, especially in the presence of a generally large classification uncertainty (i.e., when the posterior state-membership probabilities are not (close to) one for all (or most) observations). A more valid alternative is to extend the LMFA framework by means of a confirmatory factor analysis (CFA) variant that allows for equality restrictions across the states. In the following, we describe this extension and its inherent challenges. It is important to note that it would not yet be possible to use this extension in the R package *lmfa* but only in Latent GOLD.[92]

7.1.1.1 *Evaluation Procedure*

For the LMFA extension, we can build on a procedure suggested by Lubke and Muthén (2005) for evaluating different levels on invariance for mixture factor analysis (where a latent class model with non-dynamic classes is used instead of a latent Markov model with dynamic states). Specifically, the authors proposed estimating increasingly more restrictive class- (or, in our case, state-)specific CFA models (corresponding to different levels of (partial) invariance) and to determine the level of invariance based on a goodness-of-fit measure, for instance, the Bayesian information criterion (BIC; Schwarz, 1978).[93] Thus, the idea is very similar to multigroup factor analysis. The LMFA extension can be considered a latent Markov variant of the procedure for mixture factor analysis. Note that completely[94] re-estimating the LMFA model every time the model is modified is necessary to account for changes in the nature of the states, that is, by allowing

[92] In the future, it would be possible to extend the R package *lmfa* by using a mixture model that allows to specify (partially) invariant factor models in the first step, for instance, by building up on the *pgmm* package (McNicholas, ElSherbiny, McDaid, & Murphy, 2019).
[93] Note that, for simplicity, we will only describe the investigation of the tenable level of invariance by means of the BIC. However, it is debatable which criterion should be considered to decide on the tenability of invariance. The decision may be guided by (comparing) several measures (see, e.g., Lubke & Muthén, 2005).
[94] When using the three-step approach (see Chapter 4), only the first step would have to be re-estimated.

observations to switch to another state when adding equality constraints to the state-specific MMs.

7.1.1.1.1 *Evaluating Configural and Weak Invariance*

When initial results of LMFA indicate that configural invariance is not (obviously) violated (i.e., the number of factors and the non-zero loading pattern appear to be the same across states), the next steps would be to investigate if the configural model indeed holds and if the weak invariance model fits better than the configural invariance or non-invariance LMFA model according to the BIC. To this end, the LMFA model needs to be adjusted by specifying a CFA model with an invariant zero-loading structure within the states (for the configural invariance model) and by allowing to restricting all loadings to be the same across states (for the weak invariance model). As was explained in Chapters 2–6, the differences in the initial loadings of the unrestricted EFA models can be due to differences in the factor variances as a consequence of constraining the variances equal to one for identification.[95] However, when estimating the weak invariance model, it is possible to set aside differences in the loadings from the differences in the factor variances that are irrelevant for assessing weak measurement invariance by freeing up factor variances for all but a reference state (e.g., Millsap, 2012).

Once the models are estimated, one can compare the BIC of the unrestricted model to the BIC of the configural and weak invariance models. Ideally, the weak invariance model has the best fit (i.e., the lowest BIC), implying that invariance of the loadings would be endorsed. If the configural invariance model has the best fit, at least configural invariance holds. If neither weak nor configural invariance is attained, one can investigate the sources of non-invariance. In order to efficiently scrutinize configural and weak non-invariances at the same time, multigroup factor rotation (MGFR; De Roover & Vermunt, 2019) can be applied to the baseline LMFA model (i.e., the model without any equality restrictions).[96] MGFR obtains optimally comparable loadings by rotating them towards simple structure within states and towards agreement across states, while also disentangling the differences in factor variances from the differences in factor loadings. This enables hypothesis tests for these loadings by means of Wald tests. Specifically, the

[95] In Chapter 6, we explained that researchers can get a sense of whether differences in factor variances and factor loadings are mixed up by investigating whether separate loadings differ across states or whether the scaling of all loadings of a factor is affected (i.e., whether all loadings are likewise larger or smaller). In the latter case, loading differences are likely due to differences in the factor variances rather than due to differences in the loadings themselves. For example, if all loadings turn out to be $\sqrt{2}$ larger in one of the states, this can (also) be explained by the factor variance being 2 times larger.

[96] Other approaches exist to trace items with non-invariant loadings, for instance, based on modification indices or item deletion strategies (for an overview, see De Roover, Timmerman, De Leersnyder, Mesquita, & Ceulemans, 2014). However, none of these methods is as efficient as the MGFR as they all require estimating multiple models or following several subsequent steps such that selecting an acceptable partial weak invariance model can become a daunting task, especially in the presence of multiple states, factors, and items.

Wald tests indicate which loadings are significantly different across states. These loadings can then be freed in a partial weak invariance model.

7.1.1.1.2 *Evaluating Strong Invariance*

If at least partial weak invariance holds, one can progress towards evaluating strong invariance by estimating a model with all intercepts being equal across states. Note that, when estimating the strong invariance model, any non-invariant loadings have to remain unrestricted, that is, they need to be allowed to vary across states. As explained throughout Chapters 2–6, the factor means are set equal to zero for identifying the unrestricted baseline model (and also the configural and weak invariance models).[97] However, once equality restrictions are imposed on the intercepts (i.e., in the (partial) strong invariance model), one can disentangle between-state differences in factor means from differences in the intercepts by adding factor means to the model for all but a reference state (e.g., Millsap, 2012).

The fit of the strong invariance model can then be compared to the fit of the (partial) weak invariance model. If the BIC improves, strong invariance holds. When strong invariance is violated, one can go ahead and investigate if at least partial strong invariance is tenable. A method to optimally compare intercepts by means of hypothesis tests—in the absence of cross-loadings—could be multiple group factor alignment (MGFA; Asparouhov & Muthén, 2014b), which minimizes the amount of non-invariance across observed groups while accounting for differences in the factor means. Based on significance tests, MGFA indicates which parameters differ across groups.[98] However, MGFA is currently not applicable to unobserved groups (i.e., classes or states) and, thus, not for the type of model that is needed for LMFA.

An alternative approach to assessing which constraints (e.g., on intercepts) should be relaxed for improving the fit of a model is to examine modification indices (Sörbom, 1989), which are also referred to as score tests (Oberski, van Kollenburg, & Vermunt, 2013). More specifically, for the strong invariance model, the indices indicate which intercepts are most likely non-invariant across states and, thus, which ones should be freed in certain states. Based on this information, the model can be progressively adjusted and the fit compared to the (partial) weak invariance model each time, either until partial strong invariance is achieved or until the point that one can no longer theoretically justify

[97] In these models, differences in the factor means manifested themselves as differences in intercepts. We showed that one could find out whether differences in factor means and item intercepts are possibly mixed up by assessing whether separate intercepts differ or whether entire sets of intercepts appear to differ in a similar way. A "shift" in all intercepts of items corresponding to a factor would be indicative of differences in the underlying means rather than the intercepts themselves.

[98] Note that MGFA starts from the baseline model and also identifies non-invariant loadings (while accounting for differences in factor variances). Thus, the method could also be used to scrutinize non-invariant loadings. However, unlike MGFR, MGFA requires the specification of the configural model, which implies that the method cannot handle cross-loadings. Thus, full configural invariance is assumed to hold, which can be too restrictive.

freeing up more parameters. The latter is important because using this approach to adjust the model progressively until it fits the data well (i.e., until the BIC is lower than for the (partial) weak invariance model) is susceptible to capitalization on chance (De Roover, Timmerman, De Leersnyder, et al., 2014; MacCallum, Roznowski, & Necowitz, 1992).

7.1.1.1.3 *Evaluating Strict Invariance*

Finally, if (partial) strong invariance holds, a model with equal unique variances across states can be fit in order to evaluate strict invariance. As before, if the BIC improves, strict invariance holds. If strict invariance is violated, one can detect non-invariant parameters by means of modification indices for the unique variances. Again, the model can be adjusted progressively until the model fits the data well (i.e., until the BIC is better than for the (partial) strong invariance model) or until concluding that strict invariance in untenable because too many unique variances differ.

7.1.1.2 ***Challenges***

The main challenge of the LMFA extension is that the nature of the states changes every time the model is re-estimated. When starting with an LMFA model with state-specific EFA models, where all parameters may differ across states, the states are initially a mix of differences in the MM parameters and in the factor means and variances (i.e., structural model differences).[99] When fixing MM parameters in more restrictive LMFA models, differences in the factor means and variances get increasingly more dominant. In the extreme case, for the full strict invariance model, the states capture only differences in the means and variances. Therefore, when setting equality constraints, the state memberships could change and, in rare cases, even the selected number of states (i.e., the state enumeration)—as was also shown by Lubke and Muthén (2005) for standard mixture factor analysis. For example, for a non-invariant model, three states may fit best but among a partial invariant model, two states may provide the best fit. Therefore, one may wonder if the number of states should be re-evaluated for every step (i.e., for each more restrictive model), before continuing with the invariance investigation. On the one hand, increasing the number of states would rather pick up differences in the structural model, which is undesirable. On the other hand, reducing the number of states would make sense if the number of states was initially overestimated because of error fitting in the EFA-based baseline model. Ideally, one would extract enough states in the first (baseline EFA-based LMFA) model to have enough states to detect all MM differences but also not more than that in order to avoid capturing differences in the structural model. This will not always work equally well. Future research should investigate possible scenarios under different conditions (e.g., with regard to different degrees of partial invariance and differences in the structural model).

[99] Note that this is also the case for mixture factor modeling by Lubke and Muthén (2005).

7.1.2 Unraveling Response Styles

In this dissertation, we discussed two possible reasons for differences in the underlying MM, that is, substantive differences/changes in item interpretations and response styles. However, the investigation of the latter has been given little attention so far and, therefore, requires additional explanation. Response styles refer to participants' systematic styles or tendencies in responding to questionnaire items, regardless of the content of the items (Moors, 2003; Paulhus, 1991). Some of the most commonly studied types are (1) the extreme response style, where respondents tend to choose the extreme response categories (i.e., the higher and lower ends of a scale; Moors, 2003; Morren et al., 2011), (2) the agreeing (or "acquiescence") response style, where respondents tend to agree to items and thus use the higher categories of the scale (Billiet & McClendon, 2000), and (3) the "content non-responsivity response style", which comprises, among others, careless responding, where respondents do not pay sufficient attention to the questionnaire items, and random responding, where participants randomly pick an answer (Meade & Craig, 2012).

Of particular psychometric concern is that response styles can distort the MM in various ways, that is, response styles may alter factor loadings, intercept, and unique variances (Cheung & Rensvold, 2000; Eisele et al., 2020; Liu, Harbaugh, Harring, & Hancock, 2017) and, in turn, lead to invalid findings if not accounted for (e.g., construct scores could be overestimated; Bolt & Johnson, 2009). Thus, detecting respondents' response styles is important for safeguarding valid inferences. If researchers suspect which response style is at play for certain subjects and/or measurement occasions, they can possibly control for it in subsequent analyses, for instance, by including an additional response style factor (in a CFA model) that separates the effect of the response style(s) and the actual "content" factors on the item responses (Billiet & McClendon, 2000; Bolt & Johnson, 2009; Moors, 2003). Alternatively, researchers can remove inappropriate observations from the dataset to avoid confounding of the content factors. Moreover, when researchers also know possible reasons for response styles, they can try to prevent them from occurring in future studies. For example, if researchers observe a high prevalence of careless responding only after the first month of participation, they may conclude that the burden put on the participants is too high and therefore limit the length of future studies to one month.

The main challenge faced by many researchers is how to identify the presence of a particular response style in ILD. Throughout this dissertation, we advocated LMFA as a promising approach to detect MM changes, which may be due to differences and changes in response styles. One can therefore rightfully come to the intriguing question: Can LMFA indicate if there is a response style involved, and if so, which one? The short answer is: It is not that simple. In the following, we elaborate on this answer and show how LMFA in any case gives important clues for the detection of response styles.

LMFA is indeed able to detect differences in the MM, which potentially include differences caused by response styles. Sometimes, it is also possible to identify a response style and to correct for it. Consider, for instance, that content factors are assessed with balanced scales (i.e., with both positive and negatively worded items). When LMFA finds two states that are comparable with regard to the content factors but, for one state, an additional factor has been extracted with considerable loadings of all items (i.e., both positively and negatively worded items), one can suspect that this factor pertains to an agreeing response style (D'Urso, Tijmstra, Vermunt, & De Roover, 2021).

However, in practice, it is often difficult to disentangle whether differences in the MM are due to response styles or differences in item interpretations for various reasons. To name just a few, firstly, some MM differences may be due to differences in response styles and item interpretations. For instance, larger unique variances may be a sign of careless or random responding as these styles induce measurement error (Meade & Craig, 2012). However, larger unique variances may also be a sign of truly larger item-specific variability. Secondly, response styles can affect the MM parameters differently across studies. For instance, if a rather small agreeing response style is present, the response style may not be captured by an additional factor but alter the factor loadings and/or intercepts. Pinpointing the response style is then difficult. Thirdly, even if an additional factor has been extracted as consequence of a response style, the nature of this factor may be obscured by an inappropriate rotation. More specifically, naively rotating towards simple structure (e.g., by means of oblique rotation) may fail to disentangle the content factors from the response style factors, especially in the case of balanced scales (D'Urso et al., 2021; Ferrando & Lorenzo-Seva, 2010; Lorenzo-Seva, 1999; Schmitt & Sass, 2011).

As becomes apparent, it is difficult to determine if a MM pertains to a response style or not. However, LMFA sets a huge first step by making any differences in the MM visible to the researchers and researchers can apply post hoc analyses to gain insight into the nature of MM differences. Specifically, researchers may relate explanatory variables to the state memberships.[100] To this end, they may choose covariates that previously showed to be related to response styles (e.g., response time; Meade & Craig, 2012). Alternatively, one may calculate indices that are indicative for certain response styles (e.g., the frequency of the highest agreeing category; Bolt & Johnson, 2009) and use these as covariates. Or, already when designing the study, researchers could include checks that flag response styles, in particular the careless or random responding (e.g., by specifically asking to give a certain response to a question; Meade & Craig, 2012), and use these checks as covariates (i.e., whether the correct answer was provided or not). For example,

[100] Note that covariant effects on state memberships can be conveniently explored by means of the three-step approach introduced in Chapter 4 of this dissertation.

by including such a check, one could find out whether large unique variances may be related to careless responding or whether something else is going on.

Moreover, when suspecting that an additional factor in one state might be the consequence of a response style (e.g., based on significant covariate effects), researchers can try to separate content factors from the response style factor by means of applying a suitable rotation. One possibility is the target rotation (Browne, 2001), where researchers can specify (part of) the MM in a target loading matrix (i.e., the part corresponding to the postulated structure [101] of the content factors) and leave the additional factor unspecified. Then, by inspecting the rotated solution, one might conclude that the additional factor is an (agreeing) response style factor if all items load similarly on this additional factor, including items that are theoretically unrelated or with a reverse wording (D'Urso et al., 2021).

In some cases, post hoc analyses may provide insufficient information to identify a response style and to control for it. However, identifying response styles is generally a tedious task in ILD and LMFA at least indicates which observations for which subjects are comparable such that researchers can decide to continue with comparable observations from a state in which the factor structure seems reasonable. Summarized, LMFA certainly yields a contribution towards identifying response styles but further work by the researchers is required to determine if and which response style is at play.

7.2 Current Limitations and Future Research

The first important limitation is that the states in LMFA are currently strongly determined by differences in the intercepts. In Section 7.2.1, we describe the problem in more detail and outline ideas to diminish the influence of intercepts in determining the state memberships by means of extending LMFA. The second limitation is that we did not investigate consequences of ignoring non-invariance in ILD and differences in dealing with non-invariance once LMFA detected any violations of invariance. In Section7.2.2, we elaborate on this limitation and discuss ideas to investigate further how to proceed based on the results of LMFA.

7.2.1 Predominant Influence of Intercepts

One of the main findings in the simulation study in Chapter 2 was that differences in intercepts[102] are very influential in recovering the correct state memberships. On the one hand, this can be an advantage in situations in which the intercepts differ among all of the

[101] This could be based on assumptions about the structure but also based on the obtained structure of the content factors in another state (i.e., where only the content factors and not an additional (response style) factor was extracted).
[102] Note that the intercepts are equal to the item means because the factor means are constrained to be zero in each state. Thus, the structural model also influences the intercepts.

underlying states (as was the case in the simulation study) because the states are then easier to trace by LMFA. On the other hand, intercept differences can be a disadvantage in situations in which not all observations differ with regard to the intercepts (i.e., part of the observations differ only with regard to other MM parameters). It could then happen that intercept differences are so influential that only intercept differences are found (i.e., other parameter differences would be missed).[103] If enough states are selected by the model selection procedure, all MM differences should be detected. However, whether enough states would be selected to recover the correct MMs has not been investigated yet.

The problem of influential intercept differences can get increasingly severe when a questionnaire contains a few items describing, for instance, emotions or symptoms, which rarely occur for some subjects or in some situations. For instance, the emotions "feeling hopeless" or "feeling inferior" are rather extreme for subjects that do not suffer from any mental disorders. When assessing these symptoms on a Visual Analog Scale (e.g., ranging from 0 to 100) in subjects with and without mental disorders, one might obtain an unobserved mix of two response distributions, a normal (or slightly skewed) distribution for subjects with a mental disorder and a highly skewed distribution with most scores being exactly equal to zero for subjects without a mental disorder. Thus, LMFA's normality assumption would be violated. The zero scores can then be extremely influential such that concerned observations would be classified into the same state (possibly regardless of other MM differences, depending on how many states are selected). This could lead to a state with zero variances and loadings on the corresponding items. The severity of this problem has yet to be investigated.

In order to better understand the dominance of intercept differences on state (and thus parameter) recovery, it is important to conduct an extensive simulation study. In such a study, one should also investigate whether the problem can be solved by extracting additional states (possibly more than indicated by model selection criteria). However, this would only solve the problem for datasets that contain enough observations to retain the required number of states and not for datasets in which the state enumeration is bounded by a too small number of observations (for guidelines on the required number of observations, see Chapter 2). If only a limited number of states can be extracted, it might be desirable to put more focus on the differences in loadings because some analyses require only invariant loadings and not intercepts (for a discussion on the

[103] To clarify this latter case, consider that three MMs are underlying the data. MM 1 and MM 2 have the same intercepts and MM 3 has different intercepts for a few items. Furthermore, MM 2 and MM 3 have the same loadings, while MM 1 has higher loadings for a few items. Thus, MM 1 and MM 2 differ only in terms of the loadings. In case of dominant intercept differences, MM 1 and MM 2 might be merged into one state.

required level of invariance, see Chapter 6). Therefore, building up on the results of the simulation studies, LMFA should be extended to adequately deal with the undesired dominance of (zero) intercepts. In the following, we discuss two initial ideas to extend LMFA.

First, in order to lower the impact of zero-scores, we could replace the factor analysis model in LMFA by a model that accommodates skewed data. In Chapter 5, we presented latent Markov latent trait analysis (LMLTA) for non-normal categorical data. For continuous data, one could look into possibilities for implementing mixture factor analysis models for zero-censored data (e.g., Wang, Castro, Lachos, & Lin, 2019), which would take into account that item responses in the zero variance state(s) have many more zero values than can be assumed based on a normal distribution (Vermunt & Magidson, 2016). Thanks to the three-step approach (which was explained in Chapter 4), one would only have to adjust the first step of the analysis, in which the state-specific MMs are obtained. For example, if a package to estimate such a model in R would become available, it could be easily implemented in (or manually combined with step 2 and 3 of) the *lmfa* package (presented in Chapter 6).

Second, in order to lower the impact of intercept differences in general, one could move towards the Bayesian framework[104], which is, for example, implemented in MPlus (Muthén & Muthén, 2005). Using this framework, building on dynamic latent class analysis (DLCA; Asparouhov et al., 2016), it would be possible to add random effects to the intercept parameters to capture intercept differences between (not within) subjects to reduce the impact of (between-subject) intercept differences on the state formation and increase the role of other parameter differences. However, this assumes normally distributed intercept differences among subjects and future research is required to evaluate its feasibility and efficiency in lowering the dominance of intercept differences.

7.2.2 Proceeding Based on the Results of LMFA

Throughout this dissertation, we highlighted the need to detect and account for non-invariance in ILD in order to obtain valid results in subsequent analyses. In Chapter 6, we suggested several, increasingly complex ways to proceed based on the results of LMFA in order to deal with non-invariance. More specifically, we suggested that, if configural invariance is severely violated, one may continue with data of a single state (while only considering observations with a high posterior state-membership probability for this state). If (partial) weak or strong invariance holds (depending on the required level of

[104] Note that the Bayesian framework is generally more flexible in specifying complex models and for relaxing distributional assumptions. However, one should keep in mind that the results may strongly depend on the choice of priors. Additionally, by including too many parameters, one may overfit the data (Kelava & Brandt, 2019).

invariance), one may conduct a multilevel analysis on the state-specific factor scores using states as observed units and data that is weighted by the posterior state-membership probabilities. Alternatively, by means of advanced analyses such as dynamic latent class analysis (DLCA; Asparouhov et al., 2016), one could account for non-invariance by letting parameters differ across latent states or at least across subjects (McNeish et al., 2021).

It would have been interesting to go a step further in this dissertation and evaluate the different strategies to account for non-invariance specifically in the most common analysis techniques for ILD. Firstly, it would be relevant to investigate how far researchers can get with partial invariance (i.e., how much non-invariance can be tolerated before the measured construct changes meaningfully). Secondly, it would be important to understand how much non-invariance is negligible (i.e., how much can the results be trusted when ignoring non-invariance). Thirdly, it would be interesting to learn if the most complex way to deal with non-invariance (i.e., to capture non-invariance by latent states while also capturing dynamics in constructs) is really necessary or whether less advanced methods (i.e., the methods that more researchers are familiar with, like regular multilevel analysis) suffice. Fourthly, especially for replicability of the results across different studies, it would be crucial to have clear guidelines on how to deal with non-invariances. Until such guidelines are provided, researchers should at least be transparent about their choices.

7.3 Concluding Remark About LMFA for Studying Dynamic Phenomena

In the previous section, we have explained that LMFA serves as a precursor for deciding about subsequent data analyses. However, the LMFA framework also opens the doors to obtaining interesting insights into dynamics of psychological phenomena. To give a concrete example, there is a lot of potential in monitoring real-time changes in MM dynamics of patients. Mental illnesses, such as depression or bipolar disorder, are complex systems that are not fully understood yet. Often, sudden changes in symptomology are observed but, presumably, a more gradual internal process precedes this change, which could be detected by "early warning signals" (Kunkels et al., 2021; Wichers, Groot, & Psychosystems, 2016). It would also be worthwhile to devote more research to MM changes as an early warning signal. For example, when studying depression, it was repeatedly found that the dimensionality of the constructs varies as a function of overall severity of depression (for an overview, see Fried et al., 2016), which also seemed to be the case in our application in Chapter 3. Building up on the assumption that the MM is indeed related to the depression symptomology, it would be highly relevant to investigate if the MM changes even before the onset of a depressive episode. In this case, LMFA could possibly aid in anticipating such episodes. More specifically,

LMFA could be incorporated into increasingly popular ESM-based apps (e.g., M-Path[105]) that are used to monitor patients in addition to regular therapy sessions (e.g., for better understanding the patients or as a treatment in itself; Bastiaansen, Ornee, Meurs, & Oldehinkel, 2020; Kauer et al., 2012). In turn, practitioners could directly act on information about MM changes, which could improve the well-being of the patient and possibly reduce treatment costs. Summarized, the potential of LMFA for understanding psychological phenomena has not yet been fully explored.

[105] For more information, see https://m-path.io/landing/

Appendix A

The model is estimated by means of the expectation maximization (EM; Dempster et al., 1977) algorithm that uses the so-called *complete-data loglikelihood* ($\log L_c$), that is, assuming the state assignments of all time-points to be known and thus replacing the difficult maximization by a sequence of easier maximization problems. In the *expectation-step (E-step,* see e.g., Bishop, 2006; Dias, Vermunt, & Ramos, 2008), we assume the parameters of interest $\widehat{\boldsymbol{\theta}}$ (i.e., transition probabilities, initial probabilities, and state-specific MMs) to be given (i.e., by a set of initial values or estimates from the previous iteration $\widehat{\boldsymbol{\theta}}^{\mathbf{old}}$, see De Roover et al., 2017; Vermunt & Magidson, 2016) and calculate the posterior probabilities (i.e., conditional on the data) to belong to each of the states and to make transitions between the states, by means of the *forward-backward* algorithm (Baum et al., 1970). The obtained posterior probabilities are used as expected values of the state assignments to obtain the expected log L_c ($E(\log L_c)$). Next, in the *maximization-step (M-step)*, the parameters $\widehat{\boldsymbol{\theta}}$ are updated such that $E(\log L_c)$ is maximized. The *E-step* (Appendix A.1) and *M-step* (Appendix A.2) are iterated until convergence (Appendix A.3).

A.1 E-Step

The log L_c is given by:

$$\begin{aligned}\log L_c &= \log\left(\prod_{i=1}^{I}\prod_{k=1}^{K}\left[\pi_k^{s_{i1k}}\prod_{t=2}^{T}\prod_{l=1}^{K}a_{ilk}^{s_{it-1,l}\,s_{itk}}\prod_{t=1}^{T}p(\mathbf{y}_{it}|s_{itk})^{s_{itk}}\right]\right)\\ &= \sum_{i=1}^{I}\sum_{k=1}^{K}\Bigg[s_{i1k}\log(\pi_k)+\sum_{t=2}^{T}\sum_{l=1}^{K}s_{it-1,l}\,s_{itk}\log(a_{ilk})\\ &\quad -\frac{1}{2}\sum_{t=1}^{T}s_{itk}\big(J\log(2\pi)+\log(|\boldsymbol{\Sigma}_k|)+(\mathbf{y}_{it}-\boldsymbol{\nu}_k)\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_{it}-\boldsymbol{\nu}_k)'\big)\Bigg]\end{aligned} \tag{A.1}$$

Since the state memberships are in fact unknown, for each subject and time-point, the expected probability of being in a certain state $\gamma(s_{itk}) = p(s_{itk}|\,\mathbf{Y}_i)$ and the expected probability of the occurrence of two consecutive states $\varepsilon(s_{it-1,l}\;s_{itk}) = p(s_{it-1,l}\;s_{itk}|\mathbf{Y}_i)$ are inserted, in order to obtain $E(\log L_c)$:

$$E(\log L_c) = \sum_{i=1}^{I}\sum_{k=1}^{K}\Bigg[\gamma(s_{i1k})\log(\pi_k) + \sum_{t=2}^{T}\sum_{l=1}^{K}\varepsilon\big(s_{it-1,l}, s_{itk}\big)\log(a_{ilk}) - \frac{1}{2}\sum_{t=1}^{T}\gamma(s_{itk})(J\log(2\pi) + \log(|\boldsymbol{\Sigma}_k|) + (\boldsymbol{y}_{it} - \boldsymbol{\nu}_k)\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y}_{it} - \boldsymbol{\nu}_k)')\Bigg]. \quad \text{(A.2)}$$

The E-step for a LMC is achieved with the *forward-backward* (or *Baum-Welch*) algorithm (Baum et al., 1970). The algorithm finds the posterior probabilities by making use of the chain rule and the first-order Markov assumption: The joint probability of being in state *k* at time-point *t* and observing the sequence of observations can be expressed as

$$p(s_{itk}, \mathbf{Y}_i) = p(s_{itk}, \mathbf{y}_{i1}, \dots, \mathbf{y}_{it})p(\mathbf{y}_{it+1}, \dots, \mathbf{y}_{iT}|s_{itk}, \mathbf{y}_{i1}, \dots, \mathbf{y}_{it}). \quad \text{(A.3)}$$

The *first-order Markov assumption* implies that we can remove the dependency of the observations at time-point *t* on all previous observations and let them depend only on the state at time-point *t*. Thus, the equation reduces to

$$p(s_{itk}, \mathbf{Y}_i) = p(s_{itk}, \mathbf{y}_{i1}, \dots, \mathbf{y}_{it})p(\mathbf{y}_{it+1}, \dots, \mathbf{y}_{iT}|s_{itk}). \quad \text{(A.4)}$$

The first factor $p(s_{itk}, \mathbf{y}_{i1}, \dots, \mathbf{y}_{it})$ refers to the *forward probabilities* and the second factor $p(\mathbf{y}_{it+1}, \dots, \mathbf{y}_{iT}|s_{itk})$ corresponds to the *backward probabilities.* On the one hand, the forward probabilities, also indicated by $\alpha(s_{itk})$, are the probabilities of observing $\mathbf{y}_{i1:t}$ and to end in state s_{itk} and are calculated by the *forward algorithm*. First, the initial state probabilities are used to calculate the forward probabilities for subject *i* of being in state *k* at time-point 1:

$$\alpha(s_{i1k}) = \pi_k p(\mathbf{y}_{i1}|s_{i1k}). \quad \text{(A.5)}$$

Then, for the consecutive measurement occasions, we weight the forward probabilities $\alpha\big(s_{it-1,l}\big)$ of the previous time-point by the corresponding transition probabilities. Next, we sum over all possible ways (i.e., transitions) to come to a specific s_{itk} from any $s_{it-1,l}$ and multiply the values with the corresponding response probabilities:

$$\alpha(s_{itk}) = p(\mathbf{y}_{it}|s_{itk}) \sum_{l=1}^{K} \alpha(s_{it-1,l}) a_{ilk} \,. \tag{A.6}$$

On the other hand, the backward probabilities, also indicated by $\beta(s_{itk})$, are the probabilities to be in state s_{itk} and to generate the remaining sequence $\mathbf{y}_{it+1:T}$. Instead of starting at time-point 1, the so-called *backward algorithm* starts at time-point T and backtracks to time-point $t + 1$. The probability for the backward algorithm to be in final state s_{iTk} and, thus, to produce no outcome ($\emptyset$) anymore is assumed to be 1:

$$\beta(s_{iTk}) = p(\emptyset|s_{iTk}) = 1. \tag{A.7}$$

Henceforth, the backward probabilities can be determined in a similar way as the forward probabilities. However, as we go backwards here, we now need to consider the response probabilities of all K states at time-point $t + 1$, multiplying them with the backward probabilities $\beta(s_{it+1,l})$ and the respective transition probabilities, prior to summing over K:

$$\beta(s_{itk}) = \sum_{l=1}^{K} \beta(s_{it+1,l}) p(\mathbf{y}_{it+1}|s_{it+1,l})\, a_{ilk}. \tag{A.8}$$

Next, $p(s_{itk}, \mathbf{Y}_i)$ (Equation (A.4)) is calculated by multiplying the forward and the backward probabilities:

$$p(s_{itk}, \mathbf{Y}_i) = \alpha(s_{itk})\, \beta(s_{itk}) \tag{A.9}$$

Subsequently, we can calculate the conditional probability of being in state s_{itk} given the sequence of observations, as we know that:

$$p(s_{itk}, \mathbf{Y}_i) = p(s_{itk}|\, \mathbf{Y}_i)\, p(\mathbf{Y}_i) \tag{A.10}$$

By inserting this into Equation (A.9), we obtain

$$p(s_{itk}|\, \mathbf{Y}_i) p(\mathbf{Y}_i) = \alpha(s_{itk})\, \beta(s_{itk}). \tag{A.11}$$

It follows that the conditional state probability is equal to

$$p(s_{itk}|\,\mathbf{Y}_i) = \frac{\alpha(s_{itk})\,\beta(s_{itk})}{p(\mathbf{Y}_i)} = \gamma(s_{itk}). \tag{A.12}$$

The denominator can be calculated by marginalizing Equation (A.9) to $p(\mathbf{Y}_i)$ by summing over the latent states for an arbitrarily chosen t:

$$p(\mathbf{Y}_i) = \sum_{k=1}^{K} \alpha(s_{itk})\,\beta(s_{itk}), \tag{A.13}$$

which is in its simplest form for $t = T$, where $\beta(s_{iTk}) = 1$. Thus, the equation reduces to

$$p(\mathbf{Y}_i) = \sum_{k=1}^{K} \alpha(s_{iTk}). \tag{A.14}$$

Finally, we can calculate the joint probability of two successive states by applying Bayes' theorem:

$$\begin{aligned} p(s_{it-1,l}, s_{itk}|\mathbf{Y}_i) &= \frac{p\big(\mathbf{Y}_i\big|s_{it-1,l}, s_{itk}\big)\,p(s_{it-1,l}, s_{itk})}{p(\mathbf{Y}_i)} \\ &= \frac{\alpha\big(s_{it-1,l}\big)\,p(\mathbf{y}_{it}|s_{itk})\,a_{ilk}\,\beta(s_{itk})}{\sum_{k=1}^{K}\alpha(s_{iTk})} = \varepsilon\big(s_{it-1,l}, s_{itk}\big) \end{aligned} \tag{A.15}$$

A.2 M-Step

In the M-step, we maximize $E(\log L_c)$ with respect to $\widehat{\boldsymbol{\theta}}$. To maximize, we set the derivatives with respect to the parameters to zero, making use of Lagrange multipliers whenever a constraint, such as $\sum_{k=1}^{K}\pi_k = 1$, needs to be satisfied. The resulting updates are:

$$\pi_k^{new} = \frac{\sum_{i=1}^{I}\gamma(s_{i1k})}{\sum_{k=1}^{K}\sum_{i=1}^{I}\gamma(s_{i1k})} \tag{A.16}$$

$$a_{ilk}^{new} = \frac{\sum_{t=2}^{T} \varepsilon\left(s_{it-1,l}, s_{itk}\right)}{\sum_{k=1}^{K} \sum_{t=2}^{T} \varepsilon\left(s_{it-1,l}, s_{itk}\right)} \tag{A.17}$$

$$\mathbf{v}_k^{new} = \frac{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk}) \boldsymbol{y}_{it}}{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk})} \tag{A.18}$$

The factor models for the state-specific covariance matrices $\boldsymbol{\Sigma}_k^{new} = \boldsymbol{\Lambda}_k^{new} \boldsymbol{\Lambda}_k^{new} + \mathbf{D}_k^{new}$ are estimated by another maximization algorithm within each M-step. Specifically, each observation is weighted by the corresponding $\gamma(s_{itk})$-value, resulting in K weighted datasets $\mathbf{Y}_k$. Fisher scoring (Lee & Jennrich, 1979; Vermunt & Magidson, 2016) is used to perform factor analysis on these weighted data.

A.3 Convergence

The convergence can be evaluated either with respect to the log L or with respect to the parameter estimates. LG applies the latter approach and assesses convergence by computing the following quantity:

$$\delta = \sum_{r=1}^{R} \left| \frac{\hat{\theta}_r^v - \hat{\theta}_r^{v-1}}{\hat{\theta}_r^{v-1}} \right|, \tag{A.19}$$

which is the sum of the absolute values of the relative parameter changes where $r = 1, \ldots, R$ refers to the parameters. In this chapter, the stopping criterion is $\delta < 1 \times 10^{-8}$. The estimation also stops if the change in the log L becomes smaller than 1×10^{-10} prior to reaching the stopping criterion above.

Appendix B

```
//LG5.1//
version = 5.1
infile 'Sim1.csv' quote = single

model
options
   algorithm
      tolerance=1e-008 emtolerance=1e-008 emiterations=3000
      nriterations=0;
   startvalues
      seed=0 sets=25 tolerance=1e-005 iterations=100 PCA;
   bayes
      latent=0;
   montecarlo
      seed=0 replicates=500 tolerance=1e-008;
   quadrature nodes=10;
   missing  includeall;
   //classification added to output
   output
      profile parameters standarderrors estimatedvalues
      classification probmeans iterationdetails
      WriteParameters = 'results_parameters.csv'
      write = 'results.csv'
      writeloadings='results_loadings.txt';
   outfile
      'classification.csv' classification;

variables
  caseid id;
  dependent
   V1 continuous, V2 continuous, V3 continuous, V4 continuous, V5
   continuous, V6 continuous, V7 continuous, V8 continuous, V9
   continuous, V10 continuous, V11 continuous, V12 continuous, V13
   continuous, V14 continuous, V15 continuous, V16 continuous, V17
   continuous, V18 continuous, V19 continuous, V20 continuous;
  latent
   State nominal dynamic coding=first 2,
   F1 continuous dynamic,
   F2 continuous dynamic;
   independent id nominal;
equations
   // factor variances
  (1) F1| State;
  (1) F2| State;
   // Markov model
   State[=0] <- 1 ;
   State <- (~tra) 1 | State[-1] id;
   //dependent variables determined by state specific Factors
   V1-V20 <- 1 | State + F1 | State + F2 | State;
   //Unique variances
   V1-V20 | State;
end model
```

Appendix C

C.1 Problem of the Additional Simulation Study

To test whether lag-one autocorrelations of factors not captured by the within-state EFA analyses harm the performance of LMFA, we manipulated factors (a) to (g) as in the main simulation study , leaving out the levels with an already inferior performance, and added an eighth factor (h) specifying the autocorrelation.

C.2 Design and Procedure

We crossed eight factors in a complete factorial design:[106]

a. number of factors per state F^k at two levels: 2, 4;
b. number of states K at three levels: 2, 4*;
c. between state difference at two levels:
   medium loading difference & low intercept difference,
   medium loading difference & medium intercept difference*;
d. unique variance e : fixed at .2, .4;
e. frequency of transitions between the states at two levels: frequent, infrequent*;
f. number of subjects N at three levels: 2, 10*;
g. number of observations per subject per state T_{ik} at three levels: 50, 100, 200;
h. autocorrelation ϕ at three levels: 0, .3, .7*

The data was generated by means of the orthogonal dynamic factor model which implies that, at time-point *t,* the factors are uncorrelated with one another but a factor's scores at time-point t are dependent on its scores at time-point $t-1$. Specifically, they are auto-correlated by the coefficient ϕ as follows:

$$\begin{aligned} \mathbf{y}_{it} &= \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \, \mathbf{f}_{it} + \mathbf{e}_{it} \\ \mathbf{f}_{it} &= \phi \mathbf{f}_{it-1} + \boldsymbol{\varepsilon}_{it}, \end{aligned} \tag{C.1}$$

where $\boldsymbol{\varepsilon}_{it}$ is a subject-specific $F^k \times 1$ vector of noise at time-point t which is assumed to be multivariate normally distributed with zero mean and the identity matrix as covariance matrix ($\sim MVN(0; I)$). Thus, to generate the subject-specific datasets $\mathbf{Y}_i$, first, the $\mathbf{e}_{it}$ and $\boldsymbol{\varepsilon}_{it}$ vectors were sampled for each observation. Subsequently, we created the autocorrelated factor score vectors $\mathbf{f}_{it}$ by applying a recursive filter (Hamilton, 1994).

[106] The conditions marked by "*" are the ones that now have less levels than in simulation study 1.

This filter sets the first noise element as the first factor score and computes the remaining factor scores as in Equation (C.1). The resulting factor scores were multiplied by $\sqrt{1-\phi^2}$ to retain an expected variance of 1 (De Roover et al., 2014). Finally, the datasets $\mathbf{Y}_i$ were again merged into one data matrix $\mathbf{Y}$. Note that, for the strength of the autocorrelation (h), we used the values suggested by Cabrieto, Tuerlinckx, Kuppens, Grassmann and Ceulemans (2016). To check how the manipulation played out, we calculated the average autocorrelation across the datasets for each of the three conditions: they amounted to .05, .29 and .68.

For each cell of the factorial design, 20 data matrices $\mathbf{Y}$ were generated as described above. In total 2 (number of states) × 2 (number of factors) × 2 (between-state difference) × 2 (transition frequency between states) × 2 (number of subjects) × 3 (number of observations per subject and state) × 2 (unique variance) × 3 (autocorrelation) × 20 (replicates) = 5760 simulated data matrices were generated. As in Simulation Study 1, the data was generated in R and analyzed in LG with the same settings and the correct number of states and factors per state.

C.3 Results

In general, the recovery was largely unaffected by the autocorrelation conditions (h). Specifically, the recovery of the transition matrices and unique variances was not affected – MAD_{trans} = .05 and MAD_{uniq} = .03 for $\phi = 0, \phi = .3, \phi = .7$ – whereas the recovery of the states, loadings and intercepts was only slightly affected – $ARI = .91$ for $\phi = 0$ and $\phi = .3$ and $.90$ for $\phi = .7$; $GOSL = .99$ for $\phi = 0, \phi = .3$ and $.98$ for $\phi = .7$; $MADint = .07$ for $\phi = 0$, $.08$ for $\phi = .3$ and $.11$ for $\phi = .7$. Note that the mild decrease in intercept recovery with an increased autocorrelation is merely a consequence of the higher variance of the estimated intercepts, since they capture part of the autocorrelation and thus vary more around the population values.

Appendix D

In (CT-)LMFA, the log$L$ is complicated by the unknown latent states and therefore requires non-linear optimization algorithms. LG uses the expectation maximization algorithm (EM algorithm; Dempster et al., 1977) that employs the so-called complete-data loglikelihood ($\log L_c$), which means that the latent state assignments of all time-points are assumed to be known. This is convenient because the latent variables and the model parameters can be estimated separately in an iterative manner as follows: In the expectation-step (E-step; Appendix D.1), the parameters of interest, $\widehat{\boldsymbol{\theta}}$, (i.e., the initial state probabilities, the transition intensities, and the state-specific measurement models (MMs)) are assumed to be given. In the first iteration, initial values for the parameters are used and, for every other iteration, the estimates from the previous iteration $\widehat{\boldsymbol{\theta}}^{old}$ are applied. The time-specific univariate posterior probabilities of belonging to the states and the bivariate posteriors for adjacent measurement occasions, conditional on the data, are calculated by means of the forward-backward algorithm (Baum et al., 1970). These posterior probabilities are in turn used as expected values for the state memberships in order to obtain the expected $\log L_c$ ($E(\log L_c)$). Then, in the maximization-step (M-step; Appendix D.2), the parameters $\widehat{\boldsymbol{\theta}}$ get updated so that they maximize $E(\log L_c)$. This procedure is repeated until convergence (Appendix D.3).

As mentioned in Section 3.2.3, the E-step and the M-step (for all parameter updates but the transition intensities) are largely identical with the steps for DT-LMFA. Therefore, in the following, we only briefly summarize these steps. For more details and derivation of the equations, see Vogelsmeier, Vermunt, van Roekel, et al. (2019). However, we describe the M-step to update the transition intensities in more detail (Appendix D.1.3) because this is the part where CT-LMFA differs from DT-LMFA.

D.1. E-Step

The $E(\log L_c)$ is given by

$$E(\log L_c) = \sum_{i=1}^{I}\sum_{k=1}^{K}\gamma(s_{i1k})\log(\pi_k) + \sum_{i=1}^{I}\sum_{t2}^{T}\sum_{l=1}^{K}\sum_{k=1}^{K}\varepsilon\left(s_{it-1,l}, s_{itk}\right)\log\left(e^{q_{itk}\delta_{ti}}\right) - \frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{k=1}^{K}\gamma(s_{itk})[J\log(2\pi) + \log(|\boldsymbol{\Sigma}_k|) + (\mathbf{y}_{it} - \boldsymbol{\nu}_k)\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_{it} - \boldsymbol{\nu}_k)'] \quad \text{(D.1)}$$

Here, δ_{ti} refers to the time interval between time-point t and $t-1$ for subject i. Furthermore, $\gamma(s_{itk})$ are the expected values to belong to each of the states and $\varepsilon\left(s_{it-1,l}, s_{itk}\right)$ are the expected values to make transitions between the states. Both are computed based on the so-called *forward probabilities* $\alpha(s_{itk})$ —which are the

probabilities of observing the observations for time-point 1 to t, $\mathbf{y}_{i1:t}$, and ending in state s_{itk}—and the *backward probabilities* $\beta(s_{itk})$—which are the probabilities to be in state s_{itk} and to generate the remaining observations for time-point $t+1$ to T, $\mathbf{y}_{it+1:T}$. For time-point $t = 1$, the forward probabilities are computed with

$$\alpha(s_{i1k}) = \pi_k p(\mathbf{y}_{i1}|s_{i1k}) \tag{D.2}$$

and for all for all the remaining time-points with

$$\alpha(s_{itk}) = p(\mathbf{y}_{it}|s_{itk}) \sum_{l=1}^{K} \alpha\left(s_{it-1,l}\right) e^{q_{ilk}\delta_{ti}}. \tag{D.3}$$

The backward probabilities for time-point $t = T$ are computed with

$$\beta(s_{iTk}) = p(\emptyset|s_{iTk}) = 1, \tag{D.4}$$

where $\emptyset$ refers to "producing no outcome". For all the remaining time-points the backward probabilities are computed with

$$\beta(s_{itk}) = \sum_{l=1}^{K} \beta\left(s_{it+1,l}\right) p\left(\mathbf{y}_{it+1}\middle|s_{it+1,l}\right) e^{q_{ilk}\delta_{ti}}. \tag{D.5}$$

Finally, the expected univariate values to belong to each of the states are calculated with

$$\gamma(s_{itk}) = p(s_{itk}|\,\mathbf{Y}_i) = \frac{\alpha(s_{itk})\,\beta(s_{itk})}{\sum_{k=1}^{K} \alpha(s_{iTk})} \tag{D.6}$$

and the expected bivariate values to make transitions between the states with

$$\varepsilon\left(s_{it-1,l}, s_{itk}\right) = p(s_{it-1,l}, s_{itk}|\mathbf{Y}_i) = \frac{\alpha\left(s_{it-1,l}\right)\, p(\mathbf{y}_{it}|s_{itk})\, e^{q_{ilk}\delta_{ti}}\, \beta(s_{itk})}{\sum_{k=1}^{K} \alpha(s_{iTk})}. \tag{D.7}$$

Note that, upon convergence (see Appendix D.3), observations are assigned to the state they most likely belong to (i.e., to the state with the largest probability $\gamma(s_{itk})$).

D.2. M-Step

In the M-step, the parameters get updated so that they maximize ($\log L_c$).

D.1.1 Update Initial State Probabilities and Intercepts

The initial state probabilities and state-specific intercepts are updated as follows:

$$\pi_k^{new} = \frac{\sum_{i=1}^{I} \gamma(s_{i1k})}{\sum_{k=1}^{K} \sum_{i=1}^{I} \gamma(s_{i1k})}, \tag{D.8}$$

$$\boldsymbol{\nu}_k^{new} = \frac{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk}) \mathbf{y}_{it}}{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk})}. \tag{D.9}$$

D.1.2 Update State-Specific Covariance Matrices

In order to find the maximum likelihood estimates for updating the state-specific covariance matrices $\sum_k^{new} = \boldsymbol{\Lambda}_k^{new} \boldsymbol{\Lambda}_k^{new\prime} + \mathbf{D}_k^{new}$, the observations are weighted by the corresponding $\gamma(s_{itk})$ -values and these K weighted datasets $\mathbf{Y}_k$ are in turn factor analyzed by means of Fisher scoring (Lee & Jennrich, 1979).

D.1.3 Update Transition Intensities

In order to calculate the updates for the intensities, we also have to a apply a Fisher algorithm (Kalbfleisch & Lawless, 1985). This algorithm consists of two steps. First, the partial derivatives of the transition probability matrix $\mathbf{P}(\delta_{ti})$ have to be computed and second, a scoring procedure is used to find the maximum likelihood estimate of the parameters in the transition intensity matrix $\mathbf{Q}$, subsequently referred to as $\boldsymbol{\theta}^{\mathbf{Q}}$. For the example of $K = 3$ states, the parameters would be $\boldsymbol{\theta}^{\mathbf{Q}} = (q_{12}, q_{13}, q_{21}, q_{23}, q_{31}, q_{32})$. Note that Kalbfleisch and Lawless (1985) suggest to re-parameterize the parameters to $\boldsymbol{\theta}^{\mathbf{Q}} = \big(\log(q_{12}), \log(q_{13}), \log(q_{21}), \log(q_{23}), \log(q_{31}), \log(q_{32})\big)$ in order to prevent restrictions of the parameter space, which is also what Latent GOLD (LG) does. In LG, the partial derivatives of $\mathbf{P}(\delta_{ti})$ with respect to the parameters $\theta_1^{\mathbf{Q}}$ to $\theta_b^{\mathbf{Q}}$ in $\boldsymbol{\theta}^{\mathbf{Q}}$ are calculated by means of the Padé approximation (Moler & Van Loan, 2003). Once the partial derivatives are obtained, we start the scoring procedure to get the maximum likelihood estimate of $\boldsymbol{\theta}^{\mathbf{Q}}$. This implies that we first calculate the $b \times 1$ vector $\mathbf{S}(\boldsymbol{\theta}^{\mathbf{Q}})$ with entries

$$S(\theta_u^{\mathbf{Q}}) = \frac{\partial logL}{\partial \theta_u^{\mathbf{Q}}} = \sum_{i=1}^{I}\sum_{t=2}^{T}\sum_{k,l=1}^{K} \frac{\varepsilon(s_{it-1,l}, s_{itk})}{p_{lk}(\delta_{ti})}\frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^{\mathbf{Q}}}, \tag{D.10}$$

where $u = 1, \dots, b$. Here, $\varepsilon(s_{it-1,l}, s_{itk})$ are the expected bivariate state-membership probabilities obtained from the E-step (Equation (D.7)). Next, we calculate the $b \times b$ matrix $\mathbf{M}(\boldsymbol{\theta}^{\mathbf{Q}})$ with entries

$$M(\theta_u^{\mathbf{Q}}, \theta_v^{\mathbf{Q}}) = \sum_{i=1}^{I}\sum_{t=2}^{T}\sum_{k,l=1}^{K} \frac{\gamma(s_{it-1,k})}{p_{lk}(\delta_{ti})}\frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^{\mathbf{Q}}}\frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_v^{\mathbf{Q}}}, \tag{D.11}$$

where $v = 1, \dots, b$, just as u. Finally, we put all the elements together to compute the update $\boldsymbol{\theta}_{new}^{\mathbf{Q}}$:

$$\boldsymbol{\theta}_{new}^{\mathbf{Q}} = \boldsymbol{\theta}_{old}^{\mathbf{Q}} + \mathbf{M}(\boldsymbol{\theta}_{old}^{\mathbf{Q}})^{-1}\mathbf{S}(\boldsymbol{\theta}_{old}^{\mathbf{Q}}), \tag{D.12}$$

where $\boldsymbol{\theta}_{old}^{\mathbf{Q}}$ is either the initial parameter vector (for the first iteration) or the previous parameter vector (for all other iterations). This procedure is repeated until convergence within one M-step of the EM algorithm, before the EM algorithm moves on to the next E-step. The convergence criteria for the Fisher algorithm within the M-step are based on the loglikelihood and the change in parameter estimates and are the same as the ones for the "outer" total EM algorithm for CT-LMFA, which is explained in Appendix D.3.

D.3. Convergence

Convergence is evaluated with respect to either the loglikelihood or the change in parameter estimates. Primarily, LG evaluates the sum of the absolute values of the relative parameter changes, that is, $\omega = \sum_{r=1}^{R}\left|\frac{\hat{\theta}_r^{new} - \hat{\theta}_r^{old}}{\hat{\theta}_r^{old}}\right|$, with $r = 1, \dots, R$ referring to the parameters. By default, LG stops when $\omega < 1 \times 10^{-8}$. However, if the change in the loglikelihood gets smaller than 1×10^{-10} prior to reaching the stopping criterion for ω, LG stops iterating as well.

D.4. Start Values

In LG, a specific multistart procedure with multiple (e.g., 25, as used in our simulation study) sets of start values is employed, which decreases the probability of finding a local

instead of the global maximum. The start sets generally consist of random start values but, for loadings and residual variances, they are based on principal component analysis (PCA; Jolliffe, 1986) performed on the entire dataset. More specifically, to get K different start sets, randomness is added to the PCA solution per state k. For more details on the entire multistart procedure, see De Roover et al. (2017) and Vermunt and Magidson (2016).

Appendix E

This appendix provides additional information on the convergence problems inherent to the phantom-variable approach of LMFA, which emerged from an additional simulation study that we conducted. In this extra simulation study, we used the same datasets as in discrete-time- (DT-) and continuous-time- (CT-) LMFA but we put the observations on a 1-hour grid and included the phantom variables. Note that, when missing data is part of the data matrix, the response probabilities $p(\mathbf{y}_{it}|\mathbf{s}_t)$ are changed to $p(\mathbf{y}_{it}|\mathbf{s}_t)^{\kappa_{it}}$, where $\kappa_{it} = 1$ if subject i provides information for time-point t and $\kappa_{it} = 0$ otherwise. While for $\kappa_{it} = 1$ nothing changes, for $\kappa_{it} = 0$, $p(\mathbf{y}_{it}|\mathbf{s}_t)^0 = 1$, so that the missing data do not influence the likelihood (Vermunt et al., 2008).

The overall simulation study results were very much comparable to CT-LMFA (which shows that the theoretical approximation works very well in practice) and are therefore not further discussed. However, while almost all analyses converged in DT-LMFA and CT-LMFA, 10.76 % of the replications in the phantom variable approach exhibited estimation problems, especially for the lowest level of the number of measurement occasions per day (i.e., $T_{day} = 3$). Closer investigation of the non-convergence problems revealed that they were caused by reaching the maximum number of EM iterations without convergence (despite the high number of 10,000 iterations). The problem is that fewer measurement occasions per day increase the amount of phantom variables in the dataset, which hampers convergence. Re-estimating the non-converged models with new starting values or increasing the number of iterations may help. However, it should be noted that also the computation time is influenced. To validly compare the computation times, we re-estimated the first replications for all conditions while allowing for up to 50,000 iterations in the phantom-variable approach to obtain the computation times when estimation is not interrupted by too few iterations. With an average of about 10 minutes, estimation in the phantom variable approach—on an i5 processor with 8GB RAM—took about three times longer for $T_{day} = 3$ than for $T_{day} = 6$. Just to give a reference, the conditions with $T_{day} = 3$ took only about 2 minutes in CT-LMFA and 1 minute in DT-LMFA. Although this computation time is perfectly feasible, the phantom-variable approach can become unfeasible for datasets with highly unequal time intervals and very fine grids (such as the application that was described in Section 3.4), which lead to very large numbers of empty rows with missing values only.

Moreover, we also observed that the percentage of local maxima amounted to 7.24 % for datasets analyzed with the phantom-variable approach, which is much higher than for the other two methods. Here, the local maxima especially occurred for the lowest level of the number of measurement occasions per day, $T_{day} = 3$ and hence again, just as it was the case for the convergence problems, the level with the most phantom variables in a

dataset. More random start sets can reduce the probability of retaining local ML solutions (as briefly outlined in Appendix D.4).

Considering all the disadvantages of the phantom variable approach (i.e., cumbersome data-organization procedure, difficult decisions on the length of the time interval, many required iterations and start sets when the number of phantom variables is large, and results that cannot be easily compared across studies), we advise against using the phantom variable approach, which is why we did not consider this approach in our main simulation study.

Appendix F

In the following, we provide the Latent GOLD syntax that we used to analyze our application data, more specifically, the syntax of the chosen model with two states and respectively two and three factors within the states.

```
model
title '17 [3 2]';
options
   algorithm
      tolerance=1e-008 emtolerance=1e-008 emiterations=6000 nriterations=0;
   startvalues
      seed=0 sets=100 tolerance=1e-005 iterations=100 PCA;
   bayes
      latent=1 categorical=1 poisson=1 variances=1;
   montecarlo
      seed=0 replicates=500 tolerance=1e-008;
   quadrature nodes=10;
   missing  includeall;
   output
      profile parameters standarderrors estimatedvalues classification
      probmeans iterationdetails
      WriteParameters = 'results_parameters17.csv'
      write = 'results17.csv'
      writeloadings='results_loadings17.txt';
   outfile
      'classification17.csv' classification;

variables
  caseid short_ID;
  timeinterval deltaT;
  dependent
   V1 continuous,
   V2 continuous,
   V3 continuous,
   V4 continuous,
   V5 continuous,
   V6 continuous,
   V7 continuous,
   V8 continuous,
   V9 continuous,
   V10 continuous,
   V11 continuous,
   V12 continuous,
   V13 continuous,
   V14 continuous,
   V15 continuous,
   V16 continuous,
   V17 continuous,
   V18 continuous,
   V19 continuous,
   V20 continuous;
  latent
   State nominal dynamic coding=first 2,
   F1 continuous dynamic,
   F2 continuous dynamic,
   F3 continuous dynamic;

  independent condition nominal;

equations
  // factor variances
```

```
(1) F1| State;
(1) F2| State;
(1) F3| State;

// Markov model
 State[=0] <- 1 ;
 State <- (~tra) 1 | State[-1] ;

//Dependent variables determined by state specific
V1  <- 1 | State + (a1)F1 | State + (b1)F2 | State + (c1)F3 | State;
V2  <- 1 | State + (a2)F1 | State + (b2)F2 | State + (c2)F3 | State;
V3  <- 1 | State + (a3)F1 | State + (b3)F2 | State + (c3)F3 | State;
V4  <- 1 | State + (a4)F1 | State + (b4)F2 | State + (c4)F3 | State;
V5  <- 1 | State + (a5)F1 | State + (b5)F2 | State + (c5)F3 | State;
V6  <- 1 | State + (a6)F1 | State + (b6)F2 | State + (c6)F3 | State;
V7  <- 1 | State + (a7)F1 | State + (b7)F2 | State + (c7)F3 | State;
V8  <- 1 | State + (a8)F1 | State + (b8)F2 | State + (c8)F3 | State;
V9  <- 1 | State + (a9)F1 | State + (b9)F2 | State + (c9)F3 | State;
V10 <- 1 | State + (a10)F1 | State + (b10)F2 | State + (c10)F3 | State;
V11 <- 1 | State + (a11)F1 | State + (b11)F2 | State + (c11)F3 | State;
V12 <- 1 | State + (a12)F1 | State + (b12)F2 | State + (c12)F3 | State;
V13 <- 1 | State + (a13)F1 | State + (b13)F2 | State + (c13)F3 | State;
V14 <- 1 | State + (a14)F1 | State + (b14)F2 | State + (c14)F3 | State;
V15 <- 1 | State + (a15)F1 | State + (b15)F2 | State + (c15)F3 | State;
V16 <- 1 | State + (a16)F1 | State + (b16)F2 | State + (c16)F3 | State;
V17 <- 1 | State + (a17)F1 | State + (b17)F2 | State + (c17)F3 | State;
V18 <- 1 | State + (a18)F1 | State + (b18)F2 | State + (c18)F3 | State;
V19 <- 1 | State + (a19)F1 | State + (b19)F2 | State + (c19)F3 | State;
V20 <- 1 | State + (a20)F1 | State + (b20)F2 | State + (c20)F3 | State;

//Variances
V1  | State;
V2  | State;
V3  | State;
V4  | State;
V5  | State;
V6  | State;
V7  | State;
V8  | State;
V9  | State;
V10 | State;
V11 | State;
V12 | State;
V13 | State;
V14 | State;
V15 | State;
V16 | State;
V17 | State;
V18 | State;
V19 | State;
V20 | State;

//constraints:
c1[2,] = 0;
c2[2,] = 0;
c3[2,] = 0;
c4[2,] = 0;
c5[2,] = 0;
c6[2,] = 0;
c7[2,] = 0;
c8[2,] = 0;
c9[2,] = 0;
c10[2,] = 0;
c11[2,] = 0;
```

```
  c12[2,] = 0;
  c13[2,] = 0;
  c14[2,] = 0;
  c15[2,] = 0;
  c16[2,] = 0;
  c17[2,] = 0;
  c18[2,] = 0;
  c19[2,] = 0;
  c20[2,] = 0;

end model
```

Appendix G

In the following, we provide some additional information about the treatment and the Becks Depression Inventory (BDI; Beck et al., 1979) used in the presented application (Section 3.4). Regarding the treatment, all participants were randomly assigned to attend up to 20 sessions of either the cognitive behavior therapy (CBT; see Beck et al., 1979; $n = 60$) or the interpersonal psychotherapy (IPT; Klerman, Weissman, Rounsaville, & Chevron, 1984; $n = 62$). Note that there were also patients who were assigned to medication groups but that we focused on the therapy groups only. Furthermore, we did not distinguish between the two types of therapy to simplify the application, with the main purpose to simply demonstrate the use of CT-LMFA. For the requirements to participate, early termination reasons (e.g., dissatisfaction with treatment), and the explanation of the therapies and the procedure, you are referred to Elkin et al. (1989) where this has been extensively described.

With regard to the BDI measures, note that we removed the two items "weight loss" and the dichotomous item whether this was "wanted" from the original measurement because this distinction cannot be made in factor analysis. Since desired weight loss is not part of depression, we deemed it important to remove the item from our analyses.

Appendix H

H.1. Derivation Step 2

We obtain the conditional probabilities, $p(w_{itm} = 1|s_{itk} = 1)$, starting from the joint probability $p(w_{itm} = 1, s_{itk} = 1, \mathbf{y}_{it})$ applying the product rule:

$$p(w_{itm} = 1, s_{itk} = 1, \mathbf{y}_{it}) = p(w_{itm} = 1|s_{itk} = 1, \mathbf{y}_{it}) \times p(\mathbf{y}_{it}|s_{itk} = 1)p(s_{itk} = 1). \tag{H.1}$$

Next, conditioning on $s_{itk} = 1$ yields

$$\begin{aligned} p(w_{itm} = 1, \mathbf{y}_{it}|s_{itk} = 1) &= p(w_{itm} = 1|s_{itk} = 1, \mathbf{y}_{it})p(\mathbf{y}_{it}|s_{itk} = 1) \\ &= p(w_{itm} = 1|\mathbf{y}_{it})p(\mathbf{y}_{it}|s_{itk} = 1) \end{aligned} \tag{H.2}$$

because $w_{itm} = 1$ is conditionally independent of $s_{itk} = 1$ given $\mathbf{y}_{it}$ (see Figure 4.3). Next, we obtain the marginal probability by integrating out $\mathbf{y}_{it}$:

$$p(w_{itm} = 1|s_{itk} = 1) = \int p(w_{itm} = 1|\mathbf{y}_{it})p(\mathbf{y}_{it}|s_{itk} = 1)\, d\boldsymbol{y}_{it}, \tag{H.3}$$

where the second factor on the right-hand side can be rewritten using Bayes' theorem:

$$p(\mathbf{y}_{it}|s_{itk} = 1) = \frac{p(\mathbf{y}_{it})p(s_{itk} = 1|\mathbf{y}_{it})}{p(s_{itk} = 1)}. \tag{H.4}$$

Inserting this in Equation H.3 leads to

$$p(w_{itm} = 1|s_{itk} = 1) = \frac{\int p(w_{itm} = 1|\mathbf{y}_{it})p(\mathbf{y}_{it})p(s_{itk} = 1|\boldsymbol{y}_{it})\, d\mathbf{y}_{it}}{p(s_{itk} = 1)}, \tag{H.5}$$

where $p(s_{itk} = 1)$ is factored out from the integral because it is independent of $\mathbf{y}_{it}$.

H.2. Derivation Step 3

We consider the joint probability and solve for $p(\mathbf{W}_i|\mathbf{Z}_i)$. The joint probability is

$$p(\mathbf{W}_i, \mathbf{Y}_i, \mathbf{S}_i, \mathbf{Z}_i) = p(\mathbf{W}_i|\mathbf{Y}_i, \mathbf{S}_i, \mathbf{Z}_i)p(\mathbf{Y}_i|\mathbf{S}_i, \mathbf{Z}_i)p(\mathbf{S}_i|\mathbf{Z}_i)p(\mathbf{Z}_i). \tag{H.6}$$

Next, we condition on $\mathbf{Z}_i$ and use the assumption that $\mathbf{Z}_i$ and $\mathbf{Y}_i$ are conditionally independent given $\mathbf{S}_i$ and that $\mathbf{W}_i$ is conditionally independent of $\mathbf{S}_i$ and $\mathbf{Z}_i$ given $\mathbf{Y}_i$, which is also depicted in Figure 4.4:

$$\begin{aligned} &p(\mathbf{W}_i, \mathbf{Y}_i, \mathbf{S}_i|\mathbf{Z}_i) = p(\mathbf{W}_i|\mathbf{Y}_i)p(\mathbf{Y}_i|\mathbf{S}_i)p(\mathbf{S}_i|\mathbf{Z}_i) \\ &= p(\mathbf{w}_{i1}|\mathbf{y}_{i1}) \cdots p(\mathbf{w}_{iT}|\mathbf{y}_{iT})p(\mathbf{y}_{i1}|\mathbf{s}_{i1}) \cdots p(\mathbf{y}_{iT}|\mathbf{s}_{iT})p(\mathbf{s}_{i1}, \dots, \mathbf{s}_{iT}|\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}) \\ &= p(\mathbf{y}_{i1}|\mathbf{s}_{i1})p(\mathbf{w}_{i1}|\mathbf{y}_{i1}) \cdots\ p(\mathbf{y}_{iT}|\mathbf{s}_{iT})p(\mathbf{w}_{iT}|\mathbf{y}_{iT}) \times \\ &\overbrace{p(\mathbf{s}_{i1}|\mathbf{z}_{i1})p(\mathbf{s}_{i2}|\mathbf{s}_{i1}, \mathbf{z}_{i2}) \cdots p(\mathbf{s}_{iT}|\mathbf{s}_{iT-1}, \mathbf{z}_{iT})}^{Markov\ Chain\ (MC)}, \end{aligned} \tag{H.7}$$

where we use $p(\mathbf{w}_{it}|\mathbf{y}_{it})$ with the mth element in $\mathbf{w}_{it}$ equal to 1 and all others equal to 0 as a shorthand notation for $p(w_{itm} = 1|\mathbf{y}_{it})$. We then marginalize over $\mathbf{S}_i$ and $\mathbf{Y}_i$:

$$\begin{aligned} &p(\mathbf{W}_i|\mathbf{Z}_i) \\ &= \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} \int p(\mathbf{y}_{i1}|\mathbf{s}_{i1})p(\mathbf{w}_{i1}|\mathbf{y}_{i1})d\boldsymbol{y}_{it} \cdots \int\ p(\mathbf{y}_{iT}|\mathbf{s}_{iT})p(\mathbf{w}_{iT}|\mathbf{y}_{iT})d\boldsymbol{y}_{it} \\ &\times MC \end{aligned} \tag{H.8}$$

We can then rewrite $p(\mathbf{y}_{it}|\mathbf{s}_{it})$ using Bayes' theorem (see Equation (H.4)), insert it in Equation H.8, and make use of H.5, which leads to

$$\begin{aligned} &p(\mathbf{W}_i|\mathbf{Z}_i) \\ &= \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} \overbrace{\frac{\int p(\mathbf{s}_{i1}|\mathbf{y}_{i1})p(\mathbf{y}_{i1})p(\mathbf{w}_{i1}|\mathbf{y}_{i1})d\mathbf{y}_{i1}}{p(\mathbf{s}_{i1})}}^{classification\ error} \dots \overbrace{\frac{\int p(\mathbf{s}_{iT}|\mathbf{y}_{iT})p(\mathbf{y}_{iT})p(\mathbf{w}_{iT}|\mathbf{y}_{iT})d\mathbf{y}_{iT}}{p(\mathbf{s}_{iT})}}^{classification\ error} \times MC \\ &= \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{w}_{i1}|\mathbf{s}_{i1}) \cdots p(\mathbf{w}_{iT}|\mathbf{s}_{iT}) \times MC \end{aligned} \tag{H.9}$$

$$= \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}|\mathbf{z}_{i1}) \prod_{t=2}^{T} p(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it}) \prod_{t=1}^{T} p(\mathbf{w}_{it}|\mathbf{s}_{it}).$$

Appendix I

I.1. Additional Simulation Study: Autocorrelated Factor Scores

I.1.1. Problem

In order to investigate whether ignoring autocorrelated factor scores is more harmful for the performance of 3S-LMFA than it is for FIML-LMFA (Vogelsmeier, Vermunt, van Roekel, et al., 2019), we conducted a simulation study with selected conditions from the main simulation study (a–d) and, furthermore, manipulated the autocorrelation (e). More specifically, we kept the state-separation conditions (a) and (b) as they had considerable effects on the performances in the main simulation study (Section 4.3) but we kept respectively only one factor of the conditions pertaining to the amount of information (c and d) as these conditions had only minor effects on the performances. For the size of the autocorrelation, we used the coefficients suggested by Cabrieto, Tuerlinckx, Kuppens, Grassmann, and Ceulemans (2017), that were also used in the simulation study to investigate the effect of ignoring autocorrelation in FIML-LMFA (Vogelsmeier, Vermunt, van Roekel, et al., 2019).

I.1.2. Design and Procedure

The conditions were the following:

State-separation
- a. between-state loading differences at two levels: medium loading differences, low loading differences;
- b. between-state intercept differences at two levels: no intercept differences, low intercept differences;

Amount of information
- c. fixed number of subjects N: 70*;
- d. fixed number of days D: 7*;
- e. autocorrelation ϕ at three levels: 0, 0.3, 0.7

The conditions marked with "*" are the ones that are the ones that are now fixed to one value from the manipulated conditions in the main simulation study. This design resulted in $2 \times 2 \times 1 \times 1 \times 3 = 12$ conditions. The data generation was the same as in the main simulation study (again with 100 replicates). However, instead of using an orthogonal regular factor model as shown in Equation (4.1), we used an orthogonal dynamic factor model, where the factor scores at time-point t are correlated with the factor scores at $t - 1$ by the coefficient ϕ (e):

$$\begin{aligned} \mathbf{y}_{it} &= \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \mathbf{f}_{it} + \mathbf{e}_{it} \\ \mathbf{f}_{it} &= \phi \mathbf{f}_{it-1} + \boldsymbol{\varepsilon}_{it}, \end{aligned} \tag{I.1}$$

where $\boldsymbol{\varepsilon}_{it} \sim MVN(\mathbf{0}, \mathbf{I})$ is a subject- and time-point specific $F_k \times 1$ noise vector. The correlated factor scores $\mathbf{f}_{it}$ were generated by means of a recursive filter (Hamilton, 1994), that is, the first factor scores are set equal to the noise elements $\boldsymbol{\varepsilon}_{i1}$ and the remaining scores are computed as in Equation (I.1). In order to retain the expected variance of 1, we multiplied the resulting factor scores by $\sqrt{1-\phi^2}$ (De Roover, Timmerman, Van Diest, et al., 2014). Note that we computed the average autocorrelation across all datasets belonging to the same condition to see how the manipulation played out. The autocorrelations were -0.02, 0.26, and 0.64.

I.1.3. Results

Overall, the state recovery and the parameter recovery of the MMs was unaffected for both 3S- and FIML-LMFA (Table I.1). Only the state recovery after step 3 of 3S-LMFA was slightly worse for increasing autocorrelations. Furthermore, with both procedures, the intercept recovery decreased for increasing autocorrelations. This is because the autocorrelations are partly captured by the intercepts and in turn have a higher variation around the population values.

Regarding the SM (Table I.2), the recovery was also largely unaffected for both 3S- and FIML-LMFA. Only for the strongest autocorrelation ($\phi = .7$) in 3S-LMFA, the RMSE was slightly higher and the SEs slightly more underestimated for the initial state intercepts and the bias was slightly higher for the transition intercepts. Thus, the autocorrelation appears to be partially captured by the step-3 latent state transitions. However, the effect of the autocorrelation on the parameter estimation is negligible.

Table I.1. Goodness of recovery for the states, loadings, intercepts and unique variances conditional on the three autocorrelation factors

		Goodness of Recovery											
		States (ARI)			States RI (RI)			Loadings ($GOSL$)		Intercepts (MAD_{int})		Unique Variances (MAD_{unique})	
		Type of LMFA											
Condition	Factors	3S-1	3S-3	FIML	3S-1	3S-3	FIML	3S	FIML	3S	FIML	3S	FIML
	0	.61	.87	.94	.82	.94	.97	1	1	.02	.02	.01	.01
Auto-correlation ϕ	.3	.61	.86	.94	.82	.94	.97	1	1	.03	.03	.01	.01
	.7	.61	.85	.93	.82	.93	.97	1	1	.04	.04	.01	.01

Note. LMFA = latent Markov factor analysis; 3S-1 = three-step step 1; 3S-3 = three-step step 3; FIML = full information maximum likelihood. The perfect loading recoveries result from the loading matrices that are highly similar across the states.

Table I.2. Parameter bias, RMSE, and SE/SD for the four types of parameters averaged across and conditional on the three autocorrelation factors

		Initial state intercept parameters			Transition intercept parameters			Slope parameters covariate 1			Slope parameters covariate 2		
		Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD
Condition	Factor	**FIML-LMFA**											
	Average	.01	.31	1.01	-.02	.18	1	-.01	.18	0.99	0	.17	0.99
	0	.02	.31	1.01	-.02	.18	1.01	0	.18	0.98	0	.17	0.97
Auto-correlation ϕ	.3	.01	.31	1.02	-.02	.18	1.01	0	.18	0.98	0	.17	0.99
	.7	0	.32	1.01	-.02	.19	0.99	-.01	.18	1	0	.17	1
Condition	Factor	**3S-LMFA**											
	Average	.01	.35	0.98	.01	.23	0.98	0	.23	0.95	0	.20	0.96
	0	0	.33	1	0	.23	0.99	-.01	.23	0.94	-.01	.20	0.96
Auto-correlation ϕ	.3	0	.33	1	-.01	.23	0.99	-.01	.23	0.94	0	.21	0.95
	.7	.02	.37	0.94	.04	.23	0.97	0	.22	0.97	0	.20	0.97

Note. LMFA = latent Markov factor analysis; FIML = full information maximum likelihood; 3S = three-step.

I.2. Additional Simulation Study: Varying Covariate Distributions and Effects

I.2.1. Problem

In order to test whether non-uniform covariate distributions and the strength of the covariate effects influence the performance of 3S- and FIML-LMFA differently, we repeated selected conditions from the main simulation study and additionally manipulated the strength of the covariate effects (e) and the distribution of covariates (f). More specifically, we selected the conditions that affected the performances in the main simulation study the most (Section 4.3). This implied that we kept the state-separation conditions (a and b) while selecting only one factor from the conditions pertaining to the amount of information (c and d).

I.2.2. Design and Procedure

The conditions were the following:

State-separation
- a. between-state loading differences at two levels: medium loading differences, low loading differences;
- b. between-state intercept differences at two levels: no intercept differences, low intercept differences;

Amount of information
- c. number of Subjects *N* at four levels: 70*;
- d. number of days *D* at two levels: 7*;

- e. strength of covariate effects τ at three levels: 0.25, 0.5, 1
- f. distributions of covariate scores at three levels: 70/30, 50/50, 30/70

The conditions marked with "*" are the ones that are now fixed to one value from the main simulation study. This design resulted in $2 \times 2 \times 1 \times 1 \times 3 \times 3 = 36$ conditions. We generated the data as in the main simulation study (again with 100 replicates). However, the effects of the time-varying covariate z_{it1} and time-constant covariate z_{i2} as well as their distributions across observations and/or subjects differed depending on factors (e) and (f). First, with regard to the strength of the covariate effects, a higher value for z_{it1} still lowered the probabilities of transitioning to and staying in state 1 and 3 and increased the probabilities of transitioning to and staying in state 2 but with slope parameters being equal to $\gamma_{12,z_{it1}} = \gamma_{32,z_{it1}} = 1$ and $\gamma_{13,z_{it1}} = \gamma_{21,z_{it1}} = \gamma_{23,z_{it1}} = \gamma_{31,z_{it1}} = -\tau$. Furthermore, a higher value for z_{i2} still increased the probability to transition away from the origin state but with slope parameters being equal to $\gamma_{12,z_{it2}} = \gamma_{13,z_{it2}} = \gamma_{21,z_{it2}} = \gamma_{23,z_{it2}} = \gamma_{31,z_{it2}} = \gamma_{32,z_{it2}} = \tau$. The parameter τ was either 0.25, 0.5, or 1 (see factor e).

Next, with regard to the distributions of the covariate scores -0.5 and 0.5, we included conditions with a uniform distribution (i.e., "50/50") and both a "70/30" and "30/70" condition. The time-varying covariate z_{it1} was assigned such that the score changed from -0.5 to 0.5 after 5 of the 7 days in the $70/30$ condition and after 2 days in the $30/70$ condition. To obtain exactly a $50/50$ condition, the scores changed after 3 days for the first half of the subjects and after 4 days for the other half of the subjects. For the time-constant covariate z_{i2}, the scores -0.5 and 0.5 were randomly selected with probabilities being equal to the three distribution levels (i.e., $70/30$, $50/50$, or $30/70$). Note that we included a $70/30$ and $30/70$ condition to prevent a possible confounding of the results: The covariate scores influence the transition probabilities (i.e., the state memberships become more or less stable) and a higher stability of the state membership previously showed a positive influence on the recovery of the states in FIML-LMFA (Vogelsmeier, Vermunt, van Roekel, et al., 2019). For instance, a covariate score of -0.5 on both covariates would lead to a slightly more stable transition probability matrix than a covariate score of 0.5 on both covariates (e.g., with an average of 96% versus 92% probability to stay in a state with $\tau = 1$ and a one-unit interval). Note, however, that the difference is so small that it might not affect the performance.

I.2.3. Results

The results can be found in Table I.3. The state and MM recovery of 3S- and FIML-LMFA were largely unaffected by the strength of the effect and the distribution of the covariates and, therefore, will not be further discussed. With regard to the SM, there was only a very small effect with regard to the RMSE but it was the same for both estimation procedures. First, the RMSE was slightly higher for the strongest covariate effect (i.e., $\tau = 1$). This is likely due to somewhat larger SE values that are inherent to larger logit parameters. Second, the RMSE for the transition intercepts and transition slopes was slightly higher for non-uniform covariate distributions, which is likely caused by the general loss of information when covariate scores are not uniformly distributed.

Table I.3. Parameter bias, RMSE, and SE/SD for the four types of parameters averaged across and conditional on the manipulated factors

		Initial state intercept parameters			Transition intercept parameters			Slope parameters covariate 1			Slope parameters covariate 2		
		Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD	Bias	RMSE	SE/SD
Condition	Factor						**FIML-LMFA**						
	Average	.01	.31	1.02	-.03	.22	0.98	0	.2	0.98	0	.18	0.98
Strength of covariate effects τ	0.25	.01	.30	1.02	-.03	.21	0.98	-.01	.2	0.97	0	.18	0.98
	0.50	0	.30	1.04	-.02	.21	0.98	0	.2	0.99	0	.18	0.99
	1.00	.02	.32	1	-.03	.23	0.99	0	.21	0.99	-.01	.19	0.97
Distributions of covariate scores	30/70	0	.31	1.01	-.03	.24	0.99	-.02	.21	0.98	-.02	.2	0.98
	50/50	.02	.31	1.02	-.02	.19	0.98	0	.18	0.99	0	.17	0.98
	70/30	.02	.30	1.03	-.03	.22	0.98	.01	.21	0.99	0	.18	0.98
	Factor						**3S-LMFA**						
	Average	.01	.34	1	-.02	.28	0.98	0	.26	0.97	0	.22	0.96
Strength of covariate effects τ	0.25	.02	.33	1.01	-.02	.27	0.98	-.01	.25	0.96	0	.22	0.97
	0.50	.01	.33	1	-.01	.27	0.98	0	.25	0.97	0	.22	0.97
	1.00	0	.34	1	-.02	.30	0.99	.01	.28	0.98	-.01	.24	0.94
Distributions of covariate scores	30/70	.01	.34	0.99	-.02	.31	0.99	-.02	.27	0.98	-.02	.25	0.96
	50/50	-.01	.34	1	-.01	.23	0.97	0	.23	0.96	0	.2	0.96
	70/30	.01	.33	1.02	-.02	.28	0.99	.02	.27	0.98	0	.22	0.95

Note. LMFA = latent Markov factor analysis; FIML = full information maximum likelihood; 3S = three-step.

Appendix J

J.1. Step 1 and 2 Syntax of 3S-LMFA

```
//LG5.1//
version = 5.1
infile 'Dataset.csv' quote = single
model
title 'Step 1 and Sep 2 CT-[2 2 2]';
options
   algorithm
      tolerance=1e-008 emtolerance=1e-008 emiterations=5000
      nriterations=0;
   startvalues
      seed=0 sets=25 tolerance=1e-005 iterations=100 PCA;
   bayes
      latent=1 categorical=1 poisson=1 variances=1;
   quadrature nodes=10;
   missing  includeall;

   output
      parameters=effect
      betaopts=wl
      standarderrors
      classification
      profile
      probmeans=posterior
      bivariateresiduals
      estimatedvalues=model
      iterationdetails
      WriteParameters = 'results_parameters1.csv'
      write = 'results1.csv';

   outfile
      'classification1.csv' classification
   keep id deltaT cov1_vary_D cov2_con_N;

variables
  dependent
   V1 continuous,V2 continuous,V3 continuous,V4 continuous,V5
   continuous,V6 continuous,V7 continuous,V8 continuous,V9
   continuous,V10 continuous, V11 continuous, V12 continuous,V13
   continuous, V14 continuous, V15 continuous, V16 continuous, V17
   continuous, V18 continuous, V19 continuous, V20 continuous;
  latent
   State nominal  coding=first 3,
   F1 continuous,
   F2 continuous;

equations
  (1) F1| State;
  (1) F2| State;
  State <- 1 ;
  V1-V20 <- 1 | State + F1 | State + F2 | State;
```

```
  V1   | State;
  V2   | State;
  V3   | State;
  V4   | State;
  V5   | State;
  V6   | State;
  V7   | State;
  V8   | State;
  V9   | State;
  V10  | State;
  V11  | State;
  V12  | State;
  V13  | State;
  V14  | State;
  V15  | State;
  V16  | State;
  V17  | State;
  V18  | State;
  V19  | State;
  V20  | State;
end model
```

J.2. Step 3 Syntax of 3S-LMFA

Note that the step-3 syntax below is only one option to estimate the third step. Instead of calculating the classification error probability matrix manually and inserting it into the syntax ('w = …') to tell LG that the matrix should be used as fixed response probability matrix, it is also possible to use the "step3" option in LG ('step3 ml modal;'). When using this option, LG automatically calculates the classification error probability matrix from the input file (i.e., the step 1 posterior probabilities and the modal state assignments, here 'classification1.csv') and uses it as fixed response probability matrix. However, when using the step3 option, LG does not yet provide the user with the final latent state-assignments. This is because the classification is often not the primary focus of interest in other three-step analyses where researchers rather focus on parameter estimates such as covariate effects. Since classification is certainly of interest in LMFA, we suggest to use the manual syntax version.

```
//LG5.1//
version = 5.1
infile 'classification1.csv' quote = single
model
title 'Step 3-[2 2 2]';
options
   algorithm
      tolerance=1e-008 emtolerance=1e-006 emiterations=5000
      nriterations=500 expm=pade;
   startvalues
      seed=0 sets=25 tolerance=1e-005 iterations=100 PCA;
   bayes
      latent=1 categorical=1 poisson=1 variances=1 ct=1;
   quadrature nodes=10;
   missing includeall;
   output
      parameters=effect
      standarderrors
      classification
      profile
      iterationdetails
      estimatedvalues=model
      WriteParameters = 'results_parameters2.csv'
      write = 'results2.csv';
   outfile
      'classification2.csv' classification;

variables
  caseid id;
  independent cov1_vary_D nominal, cov2_con_N nominal;
   timeinterval deltaT;
  latent
   State3 nominal dynamic coding=first 3;
   dependent State#;
```

```
equations
   State3[=0] <- 1;
   State3 <- (~tra) 1 | State3[-1]
             + (~tra) cov1_vary_D | State3[-1] + (~tra)
             cov2_con_N | State3[-1];
   State# <- (w~wei) State3;
   w = values obtained from first syntax;
end model
```

Appendix K

K.1. Model Selection Procedure with the CHull Method

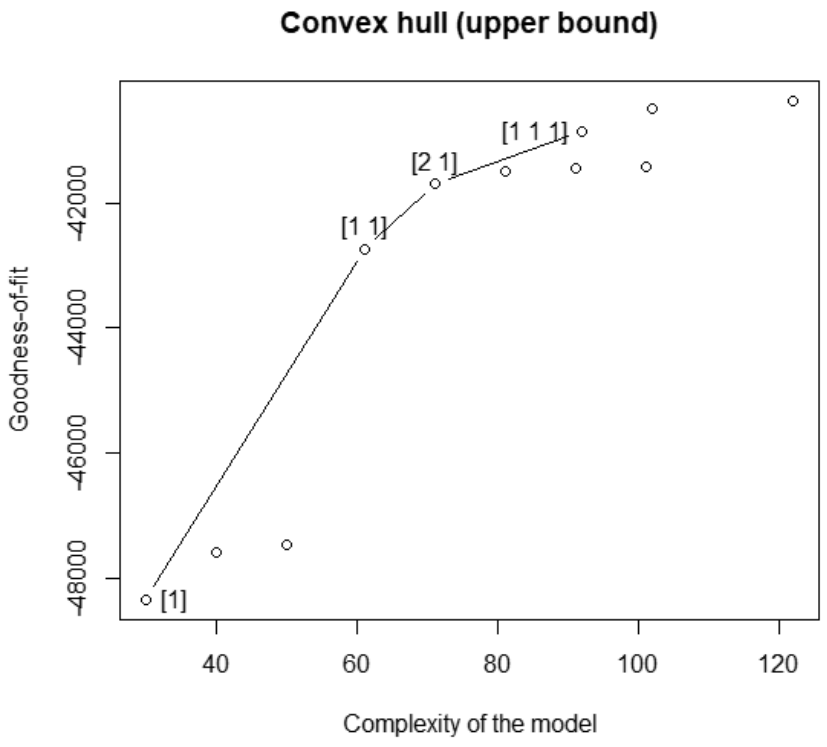
For the model selection, we ran all models five times to see whether the maximum likelihood solutions were indeed global solutions. We considered the solutions to be global when the absolute differences between the loglikelihood values of the 5 solutions was respectively smaller than 0.01. As a result, 11 out of 19 models were passed to the model selection procedure with the CHull method, which was conducted with the R-package "multichull" (note that we also did a sensitivity check by doing the CHull test including possible local optima and the selected model was always the same). The CHull can be considered an automated generalized scree-test (Bulteel et al., 2013; Ceulemans & Kiers, 2006; Ceulemans & Van Mechelen, 2005). The method identifies the models in a "loglikelihood versus number of parameters" plot that are at the higher boundary of the convex hull (Cattell, 1966) and identifies the optimal model by evaluating the elbow in the scree plot (i.e., the point where the improvement in fit with additional parameters levels off). During the CHull procedure, following Wilderjans et al. (2013)'s recommendation, we discarded models for which the fit was almost equal to the fit of a less complex model (i.e., when it fitted less than 1 percent better than the less complex model, which is also the default value in the R-package). The model with 2 states and respectively 2 and 1 factors ("[2 1]") was the best (see output in K.1.1). The second-best model was the model with two states and 1 factor in both states ("[1 1]"). From the grouping of points, corresponding to the different number of states, it can also be seen that the improvement in fit from 1 to 2 states is much larger than the one from 2 to 3 states. Model [2 1] was also better than model [1 1] according to the BIC values (see output K.1.2) and furthermore, better interpretable (model [1 1] was comparable to model [2 1] only that the second factor in the first state was clearly missing as was evident from tremendously high unique variance proportions for the items that had high loadings on the second factor in model [2 1]. Therefore, for this application, we chose model [2 1].

K.1.1. Output CHull

Output from the CHull method performed by the R-package "multichull" shows the models considered, the models on the upper bound of the convex hull, the selected model [2 1], and the CHull-figure plotting the number of free parameters against the loglikelihood value.

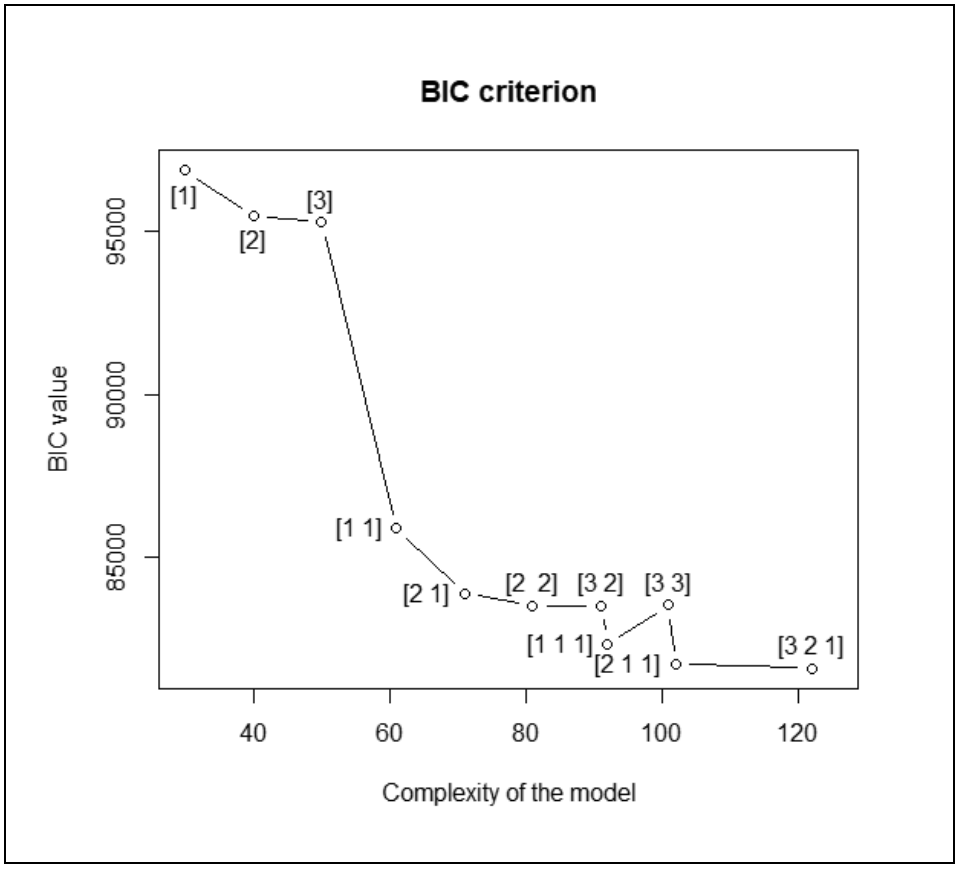
```
SETTINGS BY USER:
Optimalization: upper bound
Required improvement in fit: 1%
Number of considered models: 11

RESULTS:
Number of selected models: 1
```

```
SELECTED MODEL:
       complexity      fit
[2 1]          71 -41695.7

ALL MODELS ON upper BOUND:
         complexity       fit       st
[1]              30 -48338.88       NA
[1 1]            61 -42738.06 1.733280
[2 1]            71 -41695.70 2.572494
[1 1 1]          92 -40844.78       NA


ORIGINAL MODELS
         complexity       fit
[1]              30 -48338.88
[2]              40 -47601.71
[3]              50 -47468.42
[1 1]            61 -42738.06
[2 1]            71 -41695.70
[2  2]           81 -41477.40
[3 2]            91 -41430.23
[1 1 1]          92 -40844.78
[3 3]           101 -41406.83
[2 1 1]         102 -40495.49
[3 2 1]         122 -40367.42
```



K.1.2. Output BIC

This is an extra BIC-figure that, comparable to the CHull-figure, plots the number of free parameters against the BIC value, showing that model [2 1] fits considerably better than model [1 1] not only according to the CHull but also according to the BIC value.



K.2. Differentiation of Emotional Experience

Dutch translation of the Range Subscale form the Range and Differentiation of Emotional Experience Scale (RDEES; Kang & Shaver, 2004); Answer format: 1 = Totally disagree; 2 = disagree; 3 = disagree a little, 4 = neutral; 5 = agree a little; 6 = agree; 7 = Totally agree

"De volgende stellingen hebben betrekking op jezelf en je emoties. Sommige formuleringen lijken op elkaar. Probeer je daar niet aan te storen en probeer elke stelling te beantwoorden."

- Ik ben me bewust van de verschillende nuances en subtiliteiten van een gegeven emotie
- Elke emotie heeft een aparte en unieke betekenis voor mij
- Ik heb de neiging om subtiel onderscheid te maken tussen soortgelijke emoties (bijv. depressief en neerslachtig; geërgerd en geïrriteerd)
- Ik ben mij ervan bewust dat iedere emoties een compleet verschillende betekenis heeft
- Als emoties verschillende kleuren waren, dan zou ik zelfs de kleinste verschillen binnen één soort kleur (emotie) kunnen opmerken
- Ik ben me bewust van de subtiele verschillen tussen de gevoelens die ik ervaar
- Ik ben goed in het onderscheiden van subtiele verschillen in de betekenis van nauw verwante emotie-woorden

K.3. Emotion Clarity Deficit

Dutch translation of the Emotion Clarity Questionnaire (ECQ; Flynn & Rudolph, 2010). Answer format: 1 = Totally disagree; 2 = disagree; 3 = neutral; 4 = agree, 5 = Totally agree

"De volgende stellingen hebben betrekking op hoe jij je gevoelens ervaart. Geef antwoord in hoeverre je het eens bent met elke stelling."

- Ik weet meestal hoe ik mij voel[1]
- Meestal begrijp ik mijn gevoelens[1]
- Ik ben vaak in de war over mijn gevoelens
- Mijn gevoelens zijn meestal logisch voor mij[1]
- Ik vind het vaak lastig om te begrijpen hoe ik mij voel
- Ik ben meestal zeker over hoe ik mij voel[1]
- Ik weet meestal hoe ik mij voel[1]

[1] reversed coded

K.4. Example of Display of Mood Items in the Ethica Data App

Display of the instructions and the first item (Now, I am feeling joyful) in the Ethica Data app (Ethica Data Services Inc, 2018). The participants could scroll through all the Affect items from the Dutch version of the Positive and Negative Affect Schedule for Children (PANAS-C; Ebesutani et al., 2012; Watson et al., 1988) and give their answer by sliding on the Visual Analog Scale (VAS) from 0 (“not at all”) to 100 (“very much”). Initially the slider was set to zero. The participants had to move the slider before they could continue with the next items. Translation of instruction: The following questions are about how you are feeling. The answer scale from “not at all” to “very much”.

Appendix L

L.1. Complication in Estimating LMLTA

The main complication in estimating LMLTA is the lack of a closed form expression for the R_k-dimensional integral in the marginal density in Equation (5.9), $p(\mathbf{y}_{it}|s_{itk}=1)$. This is different in LMFA: As the factors and observations are both normally distributed continuous variables, the marginal density in Equation (5.9) can be written as multivariate normal distribution with means $\boldsymbol{\nu}_k$ and covariance matrices $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k' + \boldsymbol{\Psi}_k$, where $\boldsymbol{\Lambda}_k$ is the state-specific $J \times R_k$ loading matrix and $\boldsymbol{\Psi}_k$ contains the unique variances ψ_{kj} on the diagonal and zeros on the off-diagonal. In LMLTA, LG approximates the integral using Gauss-Hermite quadrature with M quadrature nodes per factor. For instance, with $M = 10$ and $R_k = 2$, there are 10^2 nodes in total. The integration in Equation (5.9) is then substituted by R_k summations (Vermunt & Magidson, 2016):

$$p(\mathbf{y}_{it}|s_{itk}=1) = \sum_{m=1}^{M} \dots \sum_{o=1}^{M} \left(\prod_{j=1}^{J} p\big(y_{ijt} = g|m, \dots, o, s_{itk}=1\big) A_m \cdots A_o \right). \quad \text{(L.1)}$$

Here, $m, o = 1, \dots, M$ indicate the nodes, which are the M roots of the Mth-order Hermite polynomial, and A_m indicates their corresponding weights. The values of the nodes and weights can be found in Abramowitz and Stegun (1970).[107] Note that usually at least 10 nodes per factor are used (Lesaffre & Spiessens, 2001). As the number of nodes and thus the computational effort increases exponentially, specifying models with more than three factors is often unfeasible.

L.2. Summary of the 3S Estimation

In the following, we provide a summary of the 3S estimation of the LMLTA model. In step 1, the state-specific MMs are obtained by estimating a mixture GPCM (e.g., Vermunt & Magidson, 2016). To this end, all repeated observations are treated as independent, which is valid because they are assumed to be conditionally independent given the state memberships at consecutive time-points. Hence, the relations between the states (i.e., the transitions) and the covariate effects that influence the transitions (as well as any latent class variable that clusters subjects by their transition patterns) are ignored in this step. The parameters of interest encompass (1) the state proportions, that is, the proportions

[107] Note that the formula in Equation (L.1) assumes that the factor scores are uncorrelated. When covariances are non-zero, Cholesky decomposition of the covariance matrices is used to orthogonalize the factors and obtained parameters in LG are not covariance matrices but Cholesky decomposed covariance matrices (Vermunt & Magidson, 2016).

of the observations that belong to each state, which is denoted as $p(s_k = 1)$, where s_k now refers to the state memberships across all subjects and time-points, and (2) the state-specific response probabilities $p(\mathbf{y}_{it}|s_{itk} = 1)$. The mixture GPCM is

$$p(\mathbf{y}_{it}) = \sum_{k=1}^{K} p(s_k = 1)p(\mathbf{y}_{it}|s_{itk} = 1) \tag{L.2}$$

with $p(\mathbf{y}_{it}|s_{itk} = 1)$ as in Equation (5.9). The loglikelihood function is

$$\log L_{STEP1} = \sum_{i=1}^{I}\sum_{t=1}^{T} \log p(\mathbf{y}_{it}). \tag{L.3}$$

In order to find the ML estimates for the mixture model, LG combines an expectation maximization algorithm with Newton-Raphson iterations.

Subsequently, in step 2, we consider the posterior state probabilities, $p(s_{itk} = 1|\mathbf{y}_{it}) = p(s_k = 1)p(\mathbf{y}_{it}|s_{itk} = 1)/[\sum_{k'=1}^{K} p(s_{k'} = 1)p(\mathbf{y}_{it}|s_{itk'} = 1)]$, which are the probabilities for every subject and time-point to belong to each of the states. For every observation, we assign a state membership $p(w_{itm} = 1|\mathbf{y}_{it}) = 1$ to the state k with the highest posterior probability (i.e., the most likely state membership), which implies a weight of zero for all other states.[108] The indicators w_{itm} are collected in a new variable $\mathbf{w}_{it} = (w_{it1}, \dots, w_{itK})'$ and, instead of the original observations $\mathbf{y}_{it}$, will be used for the estimation of the CT-LMM in step 3. As the highest posterior state probability is typically not equal to 1 for all observations, there will be classification error, which would lead to underestimation of the relation between the states and the covariates and the states at consecutive time-points if not accounted for in step 3. In order to calculate the errors we need to account for in step 3, we condition the assigned state memberships on the expected true state memberships $p(w_{itm} = 1|s_{itk} = 1)$, for all $k, m = 1, \dots, K$, and collect them in a $K \times K$ "classification error probability matrix". The entries of the matrix are calculated as (for details, see Di Mari et al., 2016; Vogelsmeier et al., 2020):

$$p(w_{itm} = 1|s_{itk} = 1) = \frac{\frac{1}{I \times T}\sum_{i=1}^{I}\sum_{t=1}^{T} p(w_{itm} = 1|\mathbf{y}_{it})p(s_{itk} = 1|\mathbf{y}_{it})}{p(s_k = 1)}. \tag{L.4}$$

[108] Note that this so-called "modal" assignment is the only feasible assignment procedure for a LMM with many subjects and time-points (Di Mari et al., 2016)

Note that the diagonal elements (i.e., where $k = m$), correspond to the correctly classified observations and the off-diagonal elements to the classification errors.

Finally, in the third step, we estimate the (mixture) CT-LMM based on the state memberships that were determined in the previous step and correct for the inherent classification error. As was shown by Di Mari et al. (2016) and Vogelsmeier et al. (2020), this is done by treating the state assignments $\mathbf{w}_{it}$ as error-containing observed indicators of the error-free latent states $\mathbf{s}_{it}$ that are inferred through ML estimation and used to determine the parameters of the CT-LMM. To this end, the following loglikelihood with the classification-error probabilities $p(\mathbf{w}_{it}|\mathbf{s}_{it})$ as fixed response probabilities is maximized (Vogelsmeier et al., 2020):

$$\log L_{STEP3} = \sum_{i=1}^{I} \log \left(\sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}|\mathbf{z}_{i1}) \prod_{t=2}^{T} p_{\delta_{ti}}(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it}) \prod_{t=1}^{T} p(\mathbf{w}_{it}|\mathbf{s}_{it}) \right). \quad \text{(L.5)}$$

Note that the loglikelihood of a mixture CT-LMM as used in our application (Section 5.3), where both the initial state and the transition probabilities may depend on a time-constant or time-varying latent class variable, has a slightly different form (e.g., Vermunt et al., 2008). In the simpler case of time-constant latent classes one gets:

$$\log L_{STEP3,mixture} = \sum_{i=1}^{I} \log \left(\sum_{\mathbf{c}_i} \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{c}_i)\, p(\mathbf{s}_{i1}|\mathbf{z}_{i1}, \mathbf{c}_i) \prod_{t=2}^{T} p_{\delta_{ti}}(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it}, \mathbf{c}_i) \prod_{t=1}^{T} p(\mathbf{w}_{it}|\mathbf{s}_{it}) \right), \quad \text{(L.6)}$$

where $\mathbf{c}_i = (c_{i1}, \ldots, c_{iV})'$ denotes the class memberships and $p(\mathbf{c}_i)$ the latent class or "mixture" proportions. LG obtains the ML parameter estimates by means of a combination of the forward-backward algorithm and the Newton-Raphson algorithm. For details on the mixture with a time-constant latent class variable, see Vermunt et al. (2008). The generalization to a mixture with a time-varying latent variable is straightforward and can be found in Crayen and colleagues (2017).

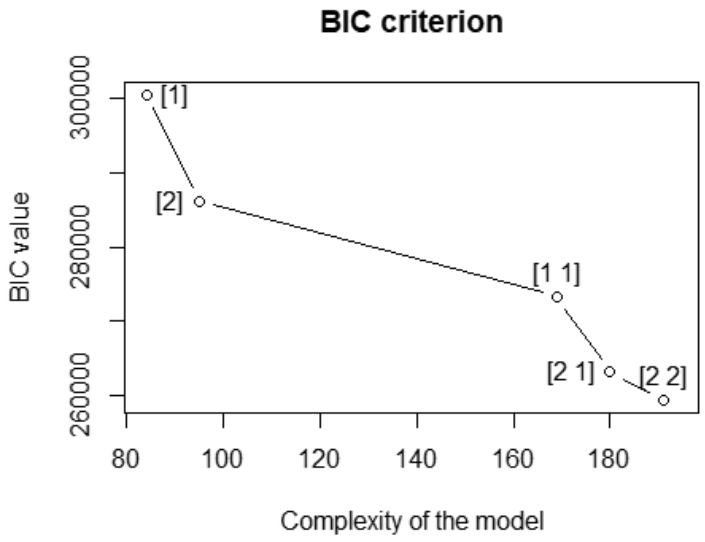
L.3. Model Selection Procedure Step 1

In the following, we provide detailed information about the model selection procedure in step 1 of the 3S approach to estimate the LMLTA model. In order to see if the ML solutions of the nine estimated models were indeed global solutions, we estimated all models five times. The ML solutions were considered global solutions (at least, as far as we know) when the absolute difference between the solutions was smaller than 0.01. This was the case for the five one- and two-state models, but not for the three-state models. First, we compared the BIC values of the stable models. As can be seen from the BIC output below (L.3.1), the two-state model with two factors per state was the best (i.e., the model "[2 2]"), because it had the lowest BIC value.

Second, with the R-package "multichull", we conducted the CHull model selection procedure, which can be considered an automated scree test that identifies which models in a "loglikelihood versus number of parameters" figure are at the higher boundary of the convex hull (Cattell, 1966) and points out where the improvement in fit levels off when adding additional parameters (Bulteel et al., 2013; Ceulemans & Kiers, 2006; Ceulemans & Van Mechelen, 2005). Note that we also included the best ML solutions of the three-state models in the CHull procedure because the method entails that the most complex and most simple model cannot be chosen and the most complex model of the stable models would have been the best fitting model according to the BIC (i.e., the model [2 2]). However, sensitivity checks using all five local optima solutions revealed that the CHull would always come to the same conclusion. As can be seen from the CHull output below (L.3.2), the two best models were the one-state model with two factors (i.e., model [2] with a "scree test value" $st = 4.52$), and the two-state model with two factors in each state (i.e., model [2 2] with $st = 3.17$). Looking at the grouping of points that correspond to the different number of states in the convex hull figure below, it can be seen that the improvement in fit is largest from one to two states, but that the improvement from two to three states is still substantial. For the application, we chose the two-state model, [2 2], because it was among the best two models according to the CHull, better than the one-state model [2] according to the BIC, parameters differed considerably across the states (as is illustrated in Section 5.3), and finally, because it was well interpretable.

L.3.1. Output BIC

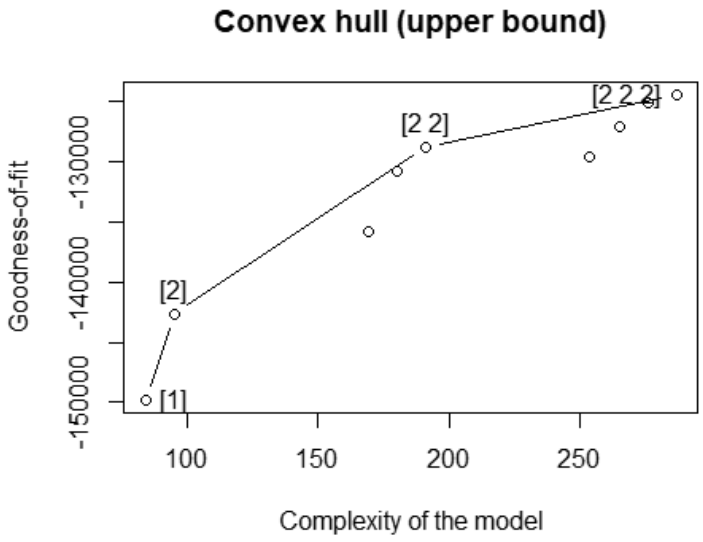
```
      number of parameters BIC value
[1]                     84  300418.1
[2]                     95  286193.3
[1 1]                  169  273333.9
[2 1]                  180  263239.3
[2 2]                  191  259446.7
```



L.3.2. Output CHull

```
SETTINGS BY USER:
Optimalization: upper bound
Required improvement in fit: 1%
Number of considered models: 9

RESULTS:
Number of selected models: 1

SELECTED MODEL:
    complexity       fit
[2]         95 -142641.7

ALL MODELS ON upper BOUND:
        complexity       fit       st
[1]             84 -149806.8       NA
[2]             95 -142641.7 4.520469
[2 2]          191 -128808.7 3.174246
[2 2 2]        287 -124450.8       NA
```

```
ORIGINAL MODELS
         complexity       fit
[1]              84 -149806.8
[2]              95 -142641.7
[1 1]           169 -135857.7
[2 1]           180 -130757.7
[2 2]           191 -128808.7
[1 1 1]         254 -129577.7
[2 1 1]         265 -127080.1
[2 2 1]         276 -125147.9
[2 2 2]         287 -124450.8
```



L.4. Additional Tables for the Application

In the following, we provide two additional tables related to the application (Section 5.3). First, we investigated whether the state-specific MM patterns (i.e., that state 1 consists of the two factors PA and NA and that state 2 consists of the two factors HA-PA and PA/NA) only emerged as a consequence of constraining specific item loadings (i.e., setting the loadings of "happy" and "unhappy" equal to zero). To this end, we re-estimated the model [2 2] with factor covariance matrices $\mathbf{\Phi}_k$ set to identity matrices (i.e., uncorrelated factors that have variances of 1) and with the loadings of the item "irritated" set to equal to 0 for respectively the first factor. The resulting loadings are shown in the first table below, Table L.1. It can be seen that the first state again consists of the two factors PA and NA and the second state again consists of the two factors HA-PA and PA/NA. The only difference compared to the initial solution given in Section 5.3 (Table 5.1) is that, in state

2, the items “happy” and “excited” now have cross-loadings (i.e., they have large loadings on both factors) and not only on the first factor. Cross-loadings can be a result from constraining truly correlated factors to be uncorrelated. We conclude that the state-specific MM patterns emerge regardless of the chosen identification constrains.

Second, Table L.2, shows the differences in the $G-1$ category intercept parameters for the 12 items across the two states. As can be seen, all intercepts differ significantly across the two states. In Section 5.3, we investigate the between-state differences in the mean item scores (Table 5.1; an explanation of how to calculate the item means is provided in the same section). These differences directly follow from the intercept differences but are easier to interpret.

Table L.1. Differences in item means and factor loadings across the two states with the covariance matrices being constrained to identity matrices

	State 1 loadings λ_{jr1}		State 2 loadings λ_{jr2}		Between-state loading difference statistics						Item means	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$	$r = 1$			$r = 2$			State 1	State 2
Item j	PA	NA	HA-PA	PA/NA	Wald	df	p-value	Wald	df	p-value		
relaxed	**2.88**	-1.45	0.60	**-1.67**	20.01	1	< 0.01	0.61	1	0.43	5.72	6.89
content	**4.47**	-2.47	1.02	**-2.57**	15.13	1	< 0.01	0.04	1	0.84	5.76	6.92
confident	**2.29**	-1.22	0.59	**-1.29**	6.74	1	< 0.01	0.03	1	0.86	5.66	6.85
happy	**3.67**	-2.06	1.78	**-2.80**	19.53	1	< 0.01	6.08	1	0.01	5.62	6.81
energetic	**1.70**	-0.93	**1.94**	-1.74	0.27	1	0.60	5.05	1	0.02	5.21	6.41
excited	**2.28**	-1.28	2.29	**-2.34**	0.00	1	1.00	14.08	1	< 0.01	5.27	6.60
sad	-0.18	**3.31**	-0.34	**1.42**	2.56	1	0.11	13.88	1	< 0.01	1.09	1.03
unhappy	-0.34	**4.20**	-0.49	**1.78**	1.19	1	0.28	21.39	1	< 0.01	1.06	1.02
disappointed	-0.09	**4.26**	-0.31	**1.72**	3.13	1	0.08	22.16	1	< 0.01	1.07	1.04
angry	0.11	**3.81**	-0.25	**1.83**	17.64	1	< 0.01	25.09	1	< 0.01	1.04	1.02
nervous	-0.17	**1.80**	-0.09	**0.76**	1.12	1	0.29	6.38	1	0.01	1.24	1.09
irritated	<u>0.00</u>	**1.96**	<u>0.00</u>	**0.86**	/	/	/	8.54	1	< 0.01	1.24	1.16

Note. PA = Positive Affect; NA = Negative Affect; HA = High Arousal; LA = Low Arousal; j refers to items, and r to factors. For identification purposes, we set the underlined loadings of the item "irritated" on the first factors ($r = 1$) equal to 0. For each item and state, the loading with the largest absolute value is printed in boldface. Note that in state 2 the three items "happy", "energetic", and "excited" have high loadings on both factors.

Table L.2. Differences in the $G-1$ category intercept parameters ν_{jgk} for the 12 items across the two states

	State 1 intercepts ν_{jg1} for $g=(1,\ldots,6)$						State 2 intercepts ν_{jg2} for $g=(1,\ldots,6)$						Between-state difference statistics		
Item j	1	2	3	4	5	6	1	2	3	4	5	6	Wald	df	p-value
relaxed	-14.38	-6.97	-1.21	4.07	6.99	7.97	-8.24	-5.94	-2.43	1.23	3.08	5.01	284.20	6	< 0.01
content	-21.85	-10.89	-1.59	6.36	10.87	12.08	-13.16	-8.83	-3.41	1.71	5.04	8.06	224.20	6	< 0.01
confident	-9.63	-4.83	-0.68	2.76	4.81	5.71	-6.21	-5.09	-2.15	0.77	2.50	4.01	199.00	6	< 0.01
happy	-21.30	-10.01	-0.71	6.53	10.64	11.21	-10.8	-6.98	-2.83	1.78	4.51	6.27	295.50	6	< 0.01
energetic	-9.03	-3.54	0.35	3.76	5.05	4.70	-8.32	-5.48	-1.58	2.25	3.76	4.16	230.40	6	< 0.01
excited	-13.02	-5.02	0.39	5.03	6.95	6.52	-9.62	-6.45	-2.32	2.27	4.28	5.33	227.20	6	< 0.01
sad	17.79	15.5	10.34	4.34	-4.04	-12.87	8.27	4.63	2.51	0.56	-2.35	-5.58	131.70	6	< 0.01
unhappy	23.40	20.61	13.56	5.06	-5.51	-17.20	9.90	6.06	3.32	0.69	-2.91	-7.00	344.10	6	< 0.01
disappointed	22.34	19.77	13.13	4.89	-5.97	-18.77	9.24	5.72	3.25	0.54	-2.75	-6.42	61.47	6	< 0.01
angry	21.17	17.93	11.30	3.86	-6.04	-16.70	9.43	5.57	3.01	0.59	-2.62	-6.52	79.08	6	< 0.01
nervous	8.78	7.45	4.84	1.61	-2.46	-7.10	5.21	2.12	1.06	0.30	-1.32	-2.95	53.37	6	< 0.01
irritated	9.61	8.32	5.56	2.18	-2.79	-8.23	4.62	1.97	0.94	0.31	-1.30	-2.80	34.28	6	< 0.01

Note. j refers to items; g refers to the item categories; G refers to the number of item categories, which is 7 in the Grumpy or Depressed study.

L.5. Syntax for Running the Models

In the following, we show the Latent GOLD (LG) syntax files to obtain the application results. Note that only one syntax is required for step 1 and 2 and one separate syntax is required for step 3.

L.5.1. Step 1 and 2 Syntax

In this syntax, the regular ESM dataset is used as input. Note that all variables in the dataset that are not necessary in step 1 and 2 but that are necessary in step 3 (i.e., all the covariates, time intervals, and subject IDs) have to be listed under "keep". The variables are then added to the "classificationS1.csv" output file containing the posterior state probabilities. This is important because the file serves as only input for step 3 of the analysis as described next.

```
options
   algorithm tolerance=1e-008 emtolerance=0.01 emiterations=250
     nriterations=50;
   startvalues seed=0 sets=100 tolerance=1e-005 iterations=100;
   bayes latent=1 categorical=1;
   quadrature nodes=10;
   missing includeall;
output
   parameters=effect
   standarderrors
   profile
   estimatedvalues=model
   iterationdetails;

outfile
   'classificationS1.csv' classification
   keep ID deltaT depression family_c classmates_c friends_c
NEWWAVE;

variables
  psuid ID ;
  dependent
     PA_LA1, PA_LA2, PA_LA3, PA_HA1, PA_HA2, PA_HA3, NA_LA1,
     NA_LA2, NA_LA3, NA_HA1, NA_HA2, NA_HA3;
  latent
      State nominal coding=first 2,
      F1 continuous,
      F2 continuous;

equations
      (c1)F1| State;
      (c2)F2| State;
      (c3)F1 <-> F2 | State;
```

```
      State <- 1 ;
      PA_LA1<- 1 | State + (a1)F1  | State + (b1)F2   | State;
      PA_LA2<- 1 | State + (a2)F1  | State + (b2)F2   | State;
      PA_LA3<- 1 | State + (a3)F1  | State + (b3)F2   | State;
      PA_HA1<- 1 | State + (a4)F1  | State + (b4)F2   | State;
      PA_HA2<- 1 | State + (a5)F1  | State + (b5)F2   | State;
      PA_HA3<- 1 | State + (a6)F1  | State + (b6)F2   | State;
      NA_LA1<- 1 | State + (a7)F1  | State + (b7)F2   | State;
      NA_LA2<- 1 | State + (a8)F1  | State + (b8)F2   | State;
      NA_LA3<- 1 | State + (a9)F1  | State + (b9)F2   | State;
      NA_HA1<- 1 | State + (a10)F1 | State + (b10)F2  | State;
      NA_HA2<- 1 | State + (a11)F1 | State + (b11)F2  | State;
      NA_HA3<- 1 | State + (a12)F1 | State + (b12)F2  | State;

  //Constraints (on "I feel happy/unhappy")
      a4[1,]=1;
      a8[1,]=0;
      b4[1,]=0;
      b8[1,]=1;
      a4[2,]=1;
      a8[2,]=0;
      b4[2,]=0;
      b8[2,]=1;
```

L.5.2. Step 3 Syntax Full Model

In this syntax, the "classificationS1.csv" file is used as input. When using LG's "step3" option, the software automatically calculates the classification error probability matrix from the posterior state probabilities. In order to specify the columns of the classificationS1.csv file in which LG can find the posterior probabilities, the user has to provide the column names as "posterior = (State.1 State.2)". Note that the column names depend on the name that was used to define the state variable in the step 1 and 2 syntax. Also note that we used a CT-LMM for the latent states and a DT-LMM for the latent classes. While subjects were allowed to transition between the states with every new observation, subjects were allowed to transition between the classes only at the beginning of every new wave. Therefore, we added a variable ("NEWWAVE") that indicated whether a record concerned a new wave (NEWWAVE = 1) or whether a record was another observation from the same wave (NEWWAVE = 0). By means of constraints on the logits, all transition probabilities for NEWWAVE = 0 were set to zero. Moreover, 19 subjects skipped wave 2. If this was ignored, LG would assume that all intervals between the waves were the same (i.e., approximately 3 months) although there were 19 longer intervals (i.e., approximately 6 months), which could lead to inaccurate parameter estimates. To solve this problem, 19 empty records (i.e., with missing values on all variables but the ID and the NEWWAVE variable) were added to the "classificationS1.csv" file. By choosing to including all records with missing observation ("missing includeall"), LG accounts for the

fact that that the second wave has been skipped and corrects for this when estimating the transition probabilities. Finally, note that the final latent state assignments may differ from the initial state assignments (i.e., the single indicators) when the classification error is rather large. In order to see the final state assignments, the user has to add the command “noignoreclassification” to the “step3” option.

```
options
  algorithm tolerance=1e-008 emtolerance=0.01 emiterations=250
    nriterations=50 expm=pade;
  startvalues seed=0 sets=10 tolerance=1e-005 iterations=100;
  bayes latent=1 categorical=1 ct=1;
  missing includeall;
  step3 ml modal noignoreclassification;

output
  parameters=effect
  standarderrors
  profile
  iterationdetails
  estimatedvalues=model
  classification;

variables
  caseid ID;
  independent family_c nominal coding=first, classmates_c nominal
    coding=first, friends_c nominal coding=first, NEWWAVE nominal
    coding=first, depression2 nominal coding=first;
  timeinterval deltaT;

  latent State nominal dynamic posterior=(State.1 State.2)
    coding=first, Class nominal dynamic dt 3 coding=first;

equations
   Class[=0] <- 1 + depression;
   Class <- (b~tra) 1 | Class[-1] NEWWAVE
                      + (~tra) depression| Class[-1];

   State[=0] <- 1;
   State <- (~tra) 1 | State[-1]
                      + (~tra) Class | State[-1]
                      + (~tra) family_c  | State[-1] Class
                      + (~tra) classmates_c   | State[-1] Class
                      + (~tra) friends_c | State[-1] Class;

   b[1] = -100;
   b[2] = -100;
   b[3] = -100;
```

L.5.3. Step 3 Syntax Reduced Model

This syntax is the same as the step 3 syntax for the full model but without the covariate effects of depression on the initial class and class transition probabilities and without the effect of being with friends on the state-transitions conditional on the class. Instead, the unconditional effect of being with friends on the state-transitions was added. Below, we only stated the changed equations.

```
equations
   Class[=0] <- 1;
   Class <- (b~tra) 1 | Class[-1] NEWWAVE;

   State[=0] <- 1;
   State <- (~tra) 1 | State[-1]
                     + (~tra) Class | State[-1]
                     + (~tra) family_c  | State[-1] Class
                     + (~tra) classmates_c   | State[-1] Class
                     + (~tra) friends_c | State[-1];

   b[1] = -100;
   b[2] = -100;
   b[3] = -100;
```

L.6. Model Selection Procedure Step 3

In the following, we provide information about the model selection procedure that was used to determine the number of latent classes in step 3 of the estimation (note that a description of the CHull procedure is provided in Appendix L.3). In contrast to the model selection in step 1 (Appendix L.3), the models in step 3 were estimated only once because local maxima are very unlikely when the MMs are fixed. First, we estimated the full model ("F") as specified in Equation (5.11) with 1–3 classes. Investigating the models with 2 and 3 classes, we saw that depression did neither predict the initial state probabilities nor the transition probabilities for the classes. Furthermore, the effect of being with friends on the transition intensities for the states appeared to be significant but did not significantly differ across classes. Since it was already apparent form the BIC that the full 3-class model fitted better than the 1- and 2-class models, we also examined a 4-class model in order not to overlook a relevant class. However, the full model with 4 classes did not converge and was therefore not considered in the model selection procedure.

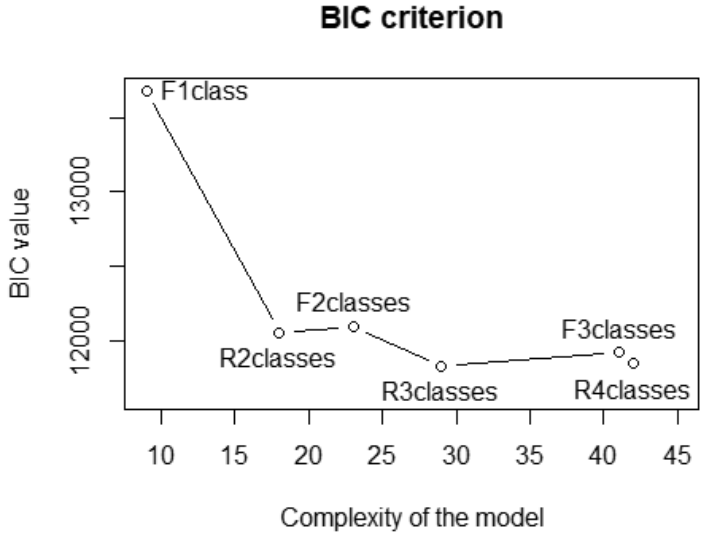
Subsequently, we re-estimated the models with multiple classes (including the 4-class model), leaving out the effects that were non-significant in the full models (i.e., the effect of depression on the initial class and transition probabilities between classes and

the effect of being with friends depending on the class), while including the unconditional effect of being with friends (i.e., not conditional on the class). All reduced ("R") models converged. As can be seen from the BIC and CHull outputs below, the reduced model with 3 classes (i.e., the model "R3classes") had the best fit according to the BIC, as it has the lowest BIC value, and was under the best three models according to the CHull when considering all converged full and reduced models.

Although we chose the reduced model with three classes, we also investigated the reduced model with two classes as the improvement in fit when adding a third class was rather small (as can be seen from the BIC and CHull plots in L.6.1 and L.6.2). Similar to the three-class solution, the two-class solution had one stable class. In the other class, adolescents had a high probability to transition between to states with a slightly higher probability to move to and stay in state 2. Therefore, if we would have considered the 2-class solution, we would have missed the third class, in which adolescents frequently transition between the states but are more likely to transition to and stay in state 1.

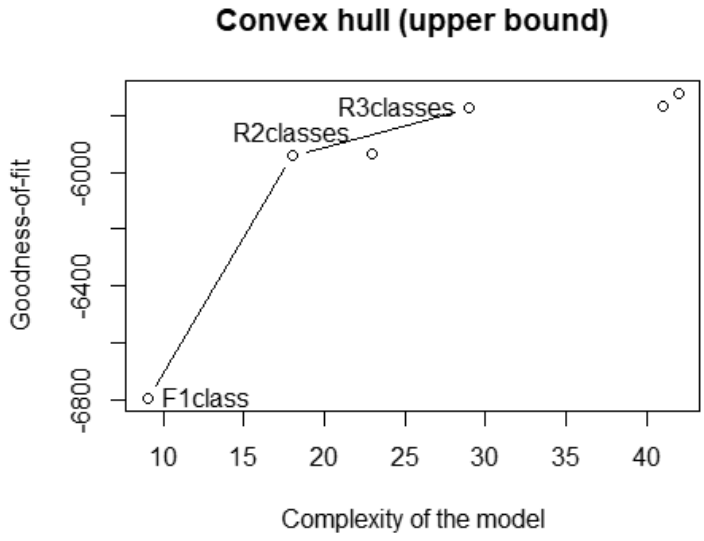
L.6.1. Output BIC

```
           number of parameters BIC value
F1class                       9  13677.23
R2classes                    18  12056.45
F2classes                    23  12097.09
R3classes                    29  11828.40
F3classes                    41  11923.62
R4classes                    42  11847.99
```



L.6.2. Output CHull

```
SETTINGS BY USER:
Optimalization: upper bound
Required improvement in fit: 1%
Number of considered models: 6

RESULTS:
Number of selected models: 1

SELECTED MODEL:
           complexity       fit
R2classes          18 -5942.029

ALL MODELS ON upper BOUND:
           complexity       fit       st
F1class             9 -6795.519       NA
R2classes          18 -5942.029 6.257659
R3classes          29 -5775.329       NA


ORIGINAL MODELS
           complexity       fit
F1class             9 -6795.519
R2classes          18 -5942.029
F2classes          23 -5938.405
R3classes          29 -5775.329
F3classes          41 -5765.476
R4classes          42 -5722.873
```



L.7. R Code to Calculate Transition Probabilities

In the following, we show how to calculate transition probabilities between the states for a given class membership and covariate value and for any interval of interest. More specifically, we calculate the transition probability matrix for being with family in class 2 and a median interval length (i.e., 2.25 hours). As explained in Section 5.2.2, the log intensities can be calculated as $\log q_{lk} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk}\mathbf{z}_{it}$ and the transition probabilities $\mathbf{P}_{2.25}$ are the matrix exponential of $\mathbf{Q} \times 2.25$.

```
library(expm)

# fill the estimates from table 2 into equation (12) from the article.
# note that the 0's and 1's are the values on the dummy variables.
# for example, -0.63 * 0 implies that we calculate the
# transition intensity when not being with friends.

q12 <- -0.55 +      # transition intercept state
        0.00 * 1 +  # effect of class for class 2
       -7.21 * 0 +  # effect of class for class 3
       -0.48 * 0 +  # effect of family x class 1
       -0.10 * 1 +  # effect of family x class 2
       -1.11 * 0 +  # effect of family x class 3
       -2.62 * 0 +  # effect of classmates x class 1
       -0.75 * 0 +  # effect of classmates x class 2
       -2.70 * 0 +  # effect of classmates x class 3
       -0.63 * 0    # effect of friends

q21 <- -0.08 +      # transition intercept state
       -1.71 * 1 +  # effect of class for class 2
       -8.74 * 0 +  # effect of class for class 3
       -0.63 * 0 +  # effect of family x class 1
       -1.12 * 1 +  # effect of family x class 2
       -2.27 * 0 +  # effect of family x class 3
       -1.30 * 0 +  # effect of classmates x class 1
        0.51 * 0 +  # effect of classmates x class 2
       -0.96 * 0 +  # effect of classmates x class 3
       -0.39 * 0    # effect of friends

# put log intensities in a matrix:
LogIntensities <- matrix(c(0, q12,
                           q21, 0),
                         nrow = 2, ncol = 2, byrow = TRUE)

# exponentiate to obtain intensities:
Intensities <- exp(LogIntensities)

# at this point, we already have the intensities of the
# off-diagonal elements. the diagonal elements are equal to
# the negative row sums of the off-diagonal elements:
for (i in 1:ncol(Intensities)){
  Intensities[i, i] <- -(sum(Intensities[i, -i]))}
```

```r
# get the probabilities for the median interval d = 2.25
# (note that we can specify any interval of interest here):
d <- 2.25
TransitionProbabilities <- expm(Intensities * d)

# transition probabilities when being with family (in class 2).
# as can be seen, there is a high probability to transition to
# state 2 (66%) and to stay in state 2 (93%):
round(TransitionProbabilities, digits = 2)
```

```
##      [,1] [,2]
## [1,] 0.34 0.66
## [2,] 0.07 0.93
```

Appendix M

In the following, we summarize the arguments and the output for each function. Note that additional documentation files are available for all functions. These can be called by typing a questionnaire mark followed by the function name (e.g., `?step1`).

M.1. `step1()` function

Arguments

`data`	The dataset with the indicators.
`indicators`	The variable names of the indicators.
`n_state`	The number of states that should be estimated when `modelselection = FALSE`
`n_fact`	The number of factors that should be estimated when `modelselection = FALSE`
`modelselection`	The indication whether model selection should be performed or not. The default is `FALSE`.
`n_state_range`	The range of states that should be estimated when `modelselection = TRUE`.
`n_fact_range`	The range of factors that should be estimated when `modelselection = TRUE`.
`n_starts`	The number of start sets. Multiple start sets are required in order to increase the chances of finding the global maximum (for details, see Appendix N.3.5). The default is 25.
`n_initial_ite`	The number of initial iterations, that is, the number of iterations that is performed for each start set (for an explanation, see Appendix N.3.5). The default is 15.
`n_m_step`	The number of maximization-step iterations inside the implemented expectation maximization algorithm (for details, see Appendix N.3). The default is 10.
`em_tolerance`	The estimation convergence criterion (for details, see Appendix N.3.4). The default is 1e-8.
`m_step_tolerance`	The criterion for stopping the maximization-step iterations. The default is 1e-3. Thus, the maximization-step iterations stop when either `m_step_tolerance` or `n_m_step` has been reached.
`max_iterations`	The maximum number of iterations after which the estimation terminates regardless of whether convergence has been reached or not. The default is 1000 iterations.
`n_mclust`	The number of mclust start sets (for details, see Appendix N.3.5). The default is 5.

Output

n_it	The number of iterations.
seconds	The time in seconds that was required to reach convergence.
convergence	Indicates whether the model estimation converged prior to reaching the maximum number of iterations. A convergence of 1 indicates that the model converged.
LL	The value of the loglikelihood.
BIC	The value of the BIC.
intercepts	The state-specific intercepts.
loadings_stand_obli	The state-specific standardized obliquely rotated loadings. If the number of factors is equal to one, the loadings are equal to the ones in loadings_stand_list.
unique_variances	The state-specific unique variances.
state_proportions	The state proportions.
n_obs	The total number of observations across all subjects and time-points.
n_par	The total number of free parameters (for details, see Appendix N.6.1).
explained_variance	The amount of explained variance weighted by the state sizes (for details, see Appendix N.6.4).
n_state	The number of states.
n_fact	The state-specific number of factors.
intercepts_list	List of state-specific intercepts.
loadings_list	List of state-specific loadings.
loadings_stand_list	List of state-specific standardized loadings.
loadings_obli_list	List of state-specific obliquely rotated loadings. If the number of factors is equal to one, the loadings are equal to the ones in loadings_list.
loadings_stand_obli_list	List of state-specific standardized obliquely rotated loadings. If the number of factors is equal to one, the loadings are equal to the ones in loadings_stand_list.
unique_variances_list	List of state-specific unique variances.
factor_correlations_stand_obli_list	List of state-specific factor correlations resulting from the rotations of the standardized loadings.
factor_correlations_obli_list	List of state-specific factor correlations resulting from the rotations of the loadings.
activated_contraints	The number of activated constraints (for details, see Appendix N.3.4).
classification_posteriors	The posterior state-membership probabilities and the modal state assignments.
classification_errors	The classification errors when using the modal state assignment (for details, see Appendix N.4).

classification_errors_prob	The classification-error probabilities when using the modal state assignment (for details, see Appendix N.4).
R2_entropy	The entropy-based R-squared measure (for details, see Appendix N.6.6)
warning_loadings	A message indicating whether convergence for rotating the loadings was reached or not.
warning_loadings_stand	A message indicating whether convergence for rotating the standardized loadings was reached or not.
raw_data	The data corresponding to the indicator items that were used in the analysis.

M.2. step2() function

Arguments

data	The dataset used in step1().
model	The model estimated with step1().

Output

classification_posteriors	The posterior state-membership probabilities and the modal state assignments.
classification_errors	The classification errors when using the modal state assignment (for details, see Appendix N.4).
classification_errors_prob	The classification-error probabilities when using the modal state assignment (for details, see Appendix N.4).
R2_entropy	The entropy-based R-squared measure (for details, see Appendix N.6.6).
totoal_classification_error	The total classification error.
state_proportions	The state proportions.
data	The data with the posterior state-membership probabilities and the modal state assignments attached.

M.3. step3() function

Arguments

data	The dataset (including the covariate values).
timeintervals	The name of the column containing the intervals between measurement occasions. The default is NULL, which means that the measurement occasions are assumed to be equidistant.
identifier	The name of the column containing the subject identifiers.
n_state	The number of states that was used for the estimation with step1().
postprobs	The posterior state-membership probabilities of step2().

`transitionCovariates`	The covariate(s) for the transition intensities. The default is `NULL`, which means that no covariate effects are estimated.
`initialCovariates`	The covariate(s) for the initial state probabilities. The default is `NULL`, which means that no covariate effects are estimated.
`n_starts`	The number of random start sets (for details, see Appendix N.5.4). The default is 25.
`n_initial_ite`	The number of initial iterations that should be performed for each start set. The default is 10.
`method`	The estimation method. The default is `"BFGS"`, which is usually faster and more stable when including covariates. The alternative is `"CG"`.
`max_iterations`	The maximum number of iterations after which the estimation stops regardless of whether convergence has been reached or not. The default is 1000.
`tolerance`	The convergence tolerance (for details, see Appendix N.5.3). The default is 1e-10. When the message occurs that the model has likely not converged because the Hessian is not positive definite, it is advisable to set the argument to a lower value and repeat the analysis (e.g., 1e-16; Jackson, 2011).
`scaling`	A scaling parameter for the loglikelihood that can prevent numerical problems from occurring, which is internally passed to the optimization function `optim()`. An appropriate scale value is close to -2 times the loglikelihood, but the loglikelihood is of course unknown prior to estimating the model. Therefore, by default, *lmfa* uses an approximation, which is based on the loglikelihood value of a CT-LMM without transitions. Next to this default (i.e., `scaling = "proxi"`), it is also possible to specify own scale values.

Output

`seconds`	The time in seconds that was required to reach convergence.
`convergence`	Indicates whether the model estimation converged prior to reaching the maximum number of iterations. A convergence of 1 indicates that the model converged. Note that it is not possible to obtain the number of iterations because this information is not returned by the `optim()` function.
`LL`	The value of the loglikelihood.
`WaldTests`	The Wald test output.
`estimates`	The parameter estimates of the transition model.
`classification_posteriors`	The posterior state-membership probabilities and the modal state assignments.
`state_proportions`	The state proportions.
`n_initialCovariates`	The number of covariates for the initial state probabilities.
`n_transitionCovariates`	The number of covariates for the transition intensities.
`n_initial_covariates`	The number of covariates specified for the initial state parameters.
`transition_covariate_means`	The number of covariates specified for the transition parameters.
`n_state`	The number of states.

`data`	The data with the posterior state-membership probabilities and the modal state assignments attached.

M.4. `chull_lmfa()` function

Arguments

`x`	The model-selection output of the function `step1()`.

Output

Prints the models on the upper boundary of the CHull, the corresponding scree-test values, and the selected model(s).

M.5. `factorscores_lmfa()` function

Arguments

`data`	The dataset used in `step1()`.
`model`	The model estimated with `step1()`.
`oblique`	The indication whether the factor scores should be obtained for the obliquely rotated loadings or unrotated loadings. The default is `TRUE`, indicating that the obliquely rotated loadings are considered.
`rounding`	The number of decimals to which the results should be rounded. The default is 4.

Output

Attached the state-specific factor scores to the dataset.

M.6. `probabilities()` function

Arguments

`model`	The transition-model output of the function `step3()`.
`deltaT`	The interval for which the transition probabilities should be calculated.
`initialCovariateScores`	The covariate scores for which the probabilities should be calculated. The default is `NULL`, which implies that any scores are set equal to the sample means.
`transitionCovariateScores`	The covariate scores for which the probabilities should be calculated. The default is `NULL`, which implies that any scores are set equal to the sample means.
`rounding`	The number of decimals to which the results should be rounded. The default is 2.

Output

Prints the initial state and transition probabilities for specified covariate values (and intervals).

Appendix N

In this appendix, we provide all technical information about latent Markov factor analysis (LMFA) and the three-step (3S) estimation with the R package *lmfa*. In the following, we first introduce relevant data notation (N.1). Then, we explain LMFA (N.2). Thereafter, we describe the three estimation steps (as presented in Section 6.2.4) and how they are implemented in *lmfa* (N.3–N.5). Finally, we provide equations for relevant statistics, such as the Bayesian information criterion (BIC; N.6).

N.1. Data Notation

The observations are denoted by y_{ijt} with $i = 1, \dots, I$ referring to the subjects, with $j = 1, \dots, J$ indicating the variables, and with $t = 1, \dots, T$,[109] referring to the measurement occasions. The observation are collected in the $J \times 1$ vectors $\mathbf{y}_{it} = \left(y_{i1t}, y_{i2t}, \dots, y_{iJt}\right)'$ that are themselves stored in the $T \times J$ data matrices $\mathbf{Y}_i = (\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, \dots, \mathbf{y}'_{iT})'$. The covariate scores are denoted by z_{itu} (with $u = 1, \dots, U$ referring to the covariates) and are collected in the $U \times 1$ vectors $\mathbf{z}_{it} = (z_{it1}, z_{it2}, \dots, z_{itU})'$, which are themselves collected in the $T \times J$ matrices $\mathbf{Z}_i = (\mathbf{z}'_{i1}, \mathbf{z}'_{i2}, \dots, \mathbf{z}'_{iT})$.

N.2. LMFA

The LMFA model consists of a transition model (i.e., the continuous-time- (CT-)LMM) and state-specific measurement models (MMs) (i.e., the FA models). The conceptual ideas behind the models were described in Section 6.2. In the following, we describe the technical details.

N.2.1. CT-LMM

The CT-LMM makes two assumptions. First, the first-order Markov assumption states that the probability to be in state k (with $k = 1, \dots, K$) at time-point t depends only on the state membership l (with $l = 1, \dots, K$) at time-point $t - 1$. Second, the local independence assumption states that the responses $\mathbf{y}_{it}$ at time-point t depend only on the state membership k at this time-point. The CT-LMM, for subject i, is defined as follows:

$$\begin{aligned} p(\mathbf{Y}_i, \mathbf{S}_i | \mathbf{Z}_i) &= p\left(\mathbf{y}_{i1}, \dots, \mathbf{y}_{iT,} \mathbf{s}_{i1}, \dots, \mathbf{s}_{iT} | \mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}\right) \\ &= \overbrace{p(\mathbf{s}_{i1}|\mathbf{z}_{i1})}^{\substack{initial\ state \\ probabilities}} \prod_{t=2}^{T} \overbrace{p_{\delta_{ti}}(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it})}^{\substack{transition \\ probabilities}} \prod_{t=1}^{T} \overbrace{p(\mathbf{y}_{it}|\mathbf{s}_{it})}^{\substack{response \\ probabilities}} . \end{aligned} \tag{N.1}$$

[109] Note that the number of time-points T may differ across subjects but we mostly omit the index i for the sake of simplicity.

The $K \times 1$ vectors $\mathbf{s}_{it} = (s_{it1}, \ldots, s_{itK})'$ contain binary indicators that determine the state memberships at time-point t, with $s_{itk} = 1$ for one state k and $s_{itk} = 0$ for all other states. As can be seen in Equation (N.1), the model contains three types of probabilities:

1. $p(\mathbf{s}_{i1}|\mathbf{z}_{i1})$ is a $K \times 1$ vector with the initial state probabilities that define the probabilities to start in a certain state at the first time-point and thus sum to one.

2. $p_{\delta_{ti}}(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it})$ is a $K \times K$ matrix $\mathbf{P}_{\delta_{ti}}$ with transition probabilities that define the probabilities to stay in a state or transition to another state at two consecutive measurement occasions with rowsums equal to 1. As indicated by the index δ_{ti}, the transition probabilities are a function of the interval δ_{ti} between two consecutive observations and the $K \times K$ transition intensity matrix $\mathbf{Q}$ with the off-diagonal elements $q_{lk} = \lim_{\delta \to 0} \frac{p(s_{itk}=1|s_{it-\delta,l}=1,\mathbf{z}_{it})}{\delta}$, which define transitions between the origin state l and the destination state k for a very small time unit. The diagonal elements are equal to the negative row sums (Cox & Miller, 1965). The transition probability matrix $\mathbf{P}_{\delta_{ti}}$ is obtained with $\mathbf{Exp}(\mathbf{Q} \times \delta_{ti})$, where $\mathbf{Exp}(\cdot)$ denotes the matrix exponential.

3. $p(\mathbf{y}_{it}|\mathbf{s}_{it})$ determine the state-specific response probabilities and thus the probabilities of having a certain response pattern depending on the state membership at time-point t. In LMFA, the probabilities depend on the state-specific MMs.

As can be seen from Equation (N.1), the initial state and transition intensities may depend on the covariates $\mathbf{z}_{it}$. Note that the covariates and the observations at time-point t are assumed to be conditionally independent given the state membership at that time-point (i.e., the covariates affect only the state membership and not the indicators directly). The covariates are included by means of regression (Bartolucci et al., 2014; Vermunt et al., 1999). Note that a logit model is used for the initial state probabilities and a log-linear model is used for the transition intensities. Specifically, the initial state probabilities are modelled as:

$$\log \frac{p(s_{i1k} = 1|\mathbf{z}_{i1})}{p(s_{i11} = 1|\mathbf{z}_{i1})} = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{z}_{it=1}; \text{ for } k = 2, \ldots, K, \qquad \text{(N.2)}$$

with β_{0k} indicating the initial state intercepts and $\boldsymbol{\beta}'_k = \left(\beta_{k,z_{i11}}, \ldots, \beta_{k,z_{i1U}}\right)'$ indicating the slopes (i.e., the covariate effects). For the transition intensities, the model is

$$\log q_{lk} = \gamma_{0lk} + \boldsymbol{\gamma}'_{lk}\mathbf{z}_{it}; \text{ for } k \neq \text{l}, \tag{N.3}$$

with γ_{0lk} as transition intercepts and $\boldsymbol{\gamma}'_{lk} = \left(\gamma_{lk,z_{it1}}, \ldots, \gamma_{lk,z_{itU}}\right)$ as slopes.

N.2.2. State-Specific FA Models

The state-specific FA models determine what the MMs look like. The state-specific FA model for subject i being in state k at time-point t (i.e., $s_{itk} = 1$) is:

$$[\mathbf{y}_{it}|s_{itk} = 1] = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \mathbf{f}_{it} + \mathbf{e}_{it}, \tag{N.4}$$

where $\boldsymbol{\nu}_{\text{k}}$ is a state-specific $J \times 1$ intercept vector; $\boldsymbol{\Lambda}_k$ is a state-specific $J \times F_k$ loading matrix, where F_k indicates the number of factors for state k; $\mathbf{f}_{it} \sim MVN(\mathbf{0}, \boldsymbol{\Phi}_k)$ is a subject-specific and time-point-specific $F_k \times 1$ vector of factor scores with $\boldsymbol{\Phi}_k$ being the state-specific factor covariance matrix. Furthermore, $\mathbf{e}_{it} \sim MVN(\mathbf{0}, \boldsymbol{\Psi}_k)$ is a subject- and time-point-specific $J \times 1$ vector of residuals, where $\boldsymbol{\Psi}_k$ contains the unique variances ψ_{kj} on the diagonal and zeros on the off-diagonal. This implies that the response probabilities, $p(\mathbf{y}_{it}|\mathbf{s}_{it})$ in Equation (N.1) are determined by state-specific multivariate normal distributions with covariance matrices $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}'_k + \boldsymbol{\Psi}_k$ and mean vectors $\boldsymbol{\nu}_k$.

N.3. Step 1 in *lmfa*

The first step of the 3S procedure consists of obtaining the maximum likelihood (ML) estimates for the state-specific MMs (and thus the FA models) by means of mixture factor analysis (McLachlan & Peel, 2000; McNicholas, 2016). In *lmfa*, this is done by means of the `step1()` function. Details about the arguments and the output are provided in Appendix M.1. In the following, we first show the loglikelihood ($\log L$) function that has to be optimized. Then, we explain how the model is estimated in the *lmfa* package, followed by the description of the inherent convergence criteria, the implemented algorithm, and, finally, the multistart procedure.

N.3.1. Likelihood Function

In order to obtain the ML estimates, the following $\log L$ function has to be maximized:

$$\log L_{STEP1} = \sum_{i=1}^{I}\sum_{t=1}^{T}\log\left(\sum_{k=1}^{K} p(s_{itk}=1)\, p(\mathbf{y}_{it}|s_{itk}=1)\right), \tag{N.5}$$

where $p(s_{itk}=1)$ are the state proportions, $p(\mathbf{y}_{it}|s_{itk}=1) = MVN(\mathbf{y}_{it}|\boldsymbol{\nu}_k, \boldsymbol{\Sigma}_k)$ are the response probabilities for a specific state, and $\boldsymbol{\Sigma}_k$ was defined before (N.2.2). In `step1()`, the proportions and response probabilities are estimated by means of an expectation maximization (EM) algorithm described next.

N.3.2. Estimation

The observed-data log$L$ in Equation (N.5) is complicated by the latent-state memberships and the latent factor scores. The EM algorithm solves this problem by iterating though the following steps. First, in the expectation- (E-)step, the model parameters are assumed to be given and the posterior state-membership probabilities are calculated accordingly (i.e., under current estimates of the model parameters). Subsequently, in the maximization- (M-)step, the posterior state-membership probabilities are treated as observed and the model parameters are updated (i.e., optimized) one by one. In fact, in the M-step, another EM algorithm with a limited number of iterations is used to update the factor parameters for each state. The algorithm iterates through the E- and M-steps until convergence. In the following, we describe the specific steps of the EM algorithm of *lmfa* (N.3.4), including convergence criteria (N.3.3), and the multistart procedure (N.3.5).

N.3.3. Convergence

The EM algorithm (N.3.4) stops when reaching a convergence criterion. In the `step1()` function, the convergence is evaluated with respect to both the logL_{STEP1}:

$$\Delta_{logL} = \log L^{v}_{STEP1} - \log L^{v-1}_{STEP1}, \tag{N.6}$$

where v refers to the iteration number, and with respect to the sum of the absolute changes in the parameter estimates:

$$\Delta_{\hat{\theta}_{EM}} = \sum_{r=1}^{R}\left|\frac{\hat{\theta}^{v}_{r} - \hat{\theta}^{v-1}_{r}}{\hat{\theta}^{v-1}_{r}}\right|, \tag{N.7}$$

with $r = 1, \ldots, R$ referring to the separate parameters. More specifically, the algorithm stops when one of the two criteria (or the maximum number of specified iterations) is reached. Within the EM algorithm, the M-step is either repeated as long as the parameters still change a lot, which is evaluated with

$$\Delta_{\hat{\theta}_{M-step}} = \sum_{r=1}^{R} \left| \frac{\hat{\theta}_r^b - \hat{\theta}_r^{b-1}}{\hat{\theta}_r^{b-1}} \right|, \tag{N.8}$$

where b refers to the M-step iteration number, or until a certain number of M-step iterations is reached.

N.3.4. EM algorithm

In the following description of the EM algorithm, we use specific values to specify, for instance, the convergence tolerance and the maximum number of iterations. These are just the default values, which may be changed by the user.

START

Set the iteration number: $v = 0$ (or equal to the final iteration number of the chosen start set as described in N.3.5). **While** Δ_{logL} and $\Delta_{\hat{\theta}_{EM}} > 1e-06$ and $v < 1000$:

1. Update the iteration number: $v = v + 1$.
2. Update posterior state-membership probabilities $p(s_{itk} = 1|\mathbf{y}_{it})$ as follows:

$$\begin{aligned} p(s_{itk} = 1|\mathbf{y}_{it}) &= \frac{p(s_{itk} = 1)p(\mathbf{y}_{it}|s_{itk} = 1)}{p(\mathbf{y}_{it})} \\ &= \frac{p(s_{itk} = 1)p(\mathbf{y}_{itk}|s_{itk} = 1)}{\sum_{k=1}^{K} p(s_{itk} = 1)p(\mathbf{y}_{itk}|s_{itk} = 1)}. \end{aligned} \tag{N.9}$$

3. Update the state-specific expected sample size N_k, state proportions $p(s_{itk} = 1)$, and $J \times 1$ state-specific intercept vectors $\boldsymbol{\nu}_k$ with

$$N_k = \sum_{i=1}^{I} \sum_{t=1}^{T} p(s_{itk} = 1|\mathbf{y}_{it}), \tag{N.10}$$

$$p(s_{itk} = 1) = \frac{N_k}{I \times T}, \text{and} \tag{N.11}$$

$$\boldsymbol{\nu}_k = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}[p(s_{itk}=1|\mathbf{y}_{it})\mathbf{y}_{it}]}{N_k}. \quad \text{(N.12)}$$

4. Compute the $J \times J$ state-specific expected observed covariance matrices $\mathbf{C}_k$ with

$$\mathbf{C}_k = \frac{\sum_{i=1}^{I}\sum_{t=1}^{T}[p(s_{itk}=1|\mathbf{y}_{it})(\mathbf{y}_{it}-\boldsymbol{\nu}_k)(\mathbf{y}_{it}-\boldsymbol{\nu}_k)']}{N_k}. \quad \text{(N.13)}$$

5. Set the M-step iteration number: $b = 0$.
6. While $\Delta_{\hat{\theta}_{M-step}} > 1e-03$ and $b < 10$:
    a. Update the M-step iteration number: $b = b + 1$.
    b. Compute the $F_k \times J$ regression-weight vectors $\boldsymbol{\beta}_k$ (from regressions of the items on the latent factors) and the $F_k \times F_k$ matrices $\boldsymbol{\Theta}_k$ containing the expectations of the factor covariances based on the current parameters and the observed data with

$$\boldsymbol{\beta}_k = \boldsymbol{\Lambda}_k'(\boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k' + \mathbf{D}_k)^{-1} \text{ and} \quad \text{(N.14)}$$

$$\boldsymbol{\Theta}_k = \mathbf{I}_{F_k} - \boldsymbol{\beta}_k\boldsymbol{\Lambda}_k + \boldsymbol{\beta}_k\mathbf{C}_k\boldsymbol{\beta}_k'. \quad \text{(N.15)}$$

    c. Update loadings $\boldsymbol{\Lambda}_k$ and unique variances $\boldsymbol{\Psi}_k$ with

$$\boldsymbol{\Lambda}_k = \mathbf{C}_k\boldsymbol{\beta}_k'\boldsymbol{\Theta}_k^{-1} \text{ and} \quad \text{(N.16)}$$

$$\boldsymbol{\Psi}_k = diag(\mathbf{C}_k - \boldsymbol{\Lambda}_k\boldsymbol{\beta}_k\mathbf{C}_k). \quad \text{(N.17)}$$

    Updating the loadings and unique variances basically comes down to calculating the regression parameters and residual variance in a series of simple linear regressions for items with the factors as predictors. Note that small unique variances can lead to numerical problems. Therefore, if the variances fall below a certain threshold value, they are fixed to this value. The item-specific threshold values are constant across states and determined by multiplying the observed variance of an item with 1e-06.[110]
    d. Compute the change in parameters $\Delta_{\hat{\theta}_{M-step}}$ (Equation (N.8)).

[110] Note that this value cannot be changed by the user.

7. Compute the logL_{STEP1} value (Equation (N.5)).
8. Compute the change in parameters $\Delta_{\hat{\theta}_{EM}}$ and logL_{STEP1} Δ_{logL} (Equations (N.7) and (N.6)).

END

N.3.5. Multistart Procedure

In order to increase the chances of finding the global maximum, the following multistart procedure with multiple start sets is used where the number of starts is equal to the specified number of model-based clusterings plus 10 times the specified number of start sets (i.e., by default, $5 + 10 \times 25 = 255$). Specifically, the procedure starts with five model-based clusterings by means of the *mclust* package (Scrucca, Fop, Murphy, & Raftery, 2016) to assign observations to states. Then, parameters are initialized for each of the five start sets as follows. The state-specific sample size N_k, state proportions $p(s_{itk} = 1)$, and state-specific intercepts $\mathbf{v}_k$, as well as the state-specific weighted sample covariance matrices $\mathbf{C}_k$ are computed as in Equations (N.12) and (N.13). Subsequently, probabilistic principal component analysis (Tipping & Bishop, 1999) is used in order to obtain the state-specific loading matrices $\mathbf{\Lambda}_k$ and unique variances $\mathbf{\Psi}_k$. To this end, the following eigendecomposition is conducted:

$$\widetilde{\mathbf{C}_k} = \mathbf{M}_k \mathbf{v}_k {\mathbf{M}_k}^{-1}, \tag{N.18}$$

where the state-specific $J \times J$ matrix $\mathbf{M}_k$ is the matrix of eigenvectors and the state-specific $J \times 1$ vector $\mathbf{v}_k$ the eigenvalues. Moreover, the state-specific $J \times F_k$ matrix $\mathbf{M}_{k,F_k}$ contains the first F_k eigenvectors, the state-specific $J \times (J - F_k)$ matrix $\mathbf{M}_{k,-F_k}$ contains the disregarded eigenvectors, and the state-specific $F_k \times F_k$ diagonal matrix $\mathbf{V}_{k,F_k}$ contains the first F_k eigenvalues of $\mathbf{v}_k$ on its diagonal. The loadings and unique variances are then obtained with

$$\mathbf{\Lambda}_k = \mathbf{M}_{k,F_k} \sqrt{\mathbf{V}_{k,F_k} - \frac{\sum \mathbf{M}_{k,-F_k}}{J - F_k} \mathbf{I}_{F_k}} \text{ and with} \tag{N.19}$$

$$\mathbf{\Psi}_k = \frac{\sum \mathbf{M}_{k,-F_k}}{J - F_k} \mathbf{I}_J, \tag{N.20}$$

where $\sum \mathbf{M}_{k,-F_k}$ denotes the sum of the disregarded eigenvalues and $\mathbf{I}_{F_k}$ and $\mathbf{I}_J$ denote $F_k \times F_k$ and $J \times J$ identity matrices, respectively. Next, the loadings $\mathbf{\Lambda}_k$ and unique variances $\mathbf{\Psi}_k$ are updated once as in Equations (N.16) and (N.17). Subsequently, the value of $\log L_{STEP1}$ is obtained (Equation (N.5)). Then, the partitions are ranked according to their $\log L_{STEP1}$ values.

From the *mclust* start set with the largest $\log L_{STEP1}$ value, 250 random start sets are generated by iteratively reassigning 30 percent of the assignments, that is, for start set 1, 30 percent of the *mclust* assignments are reassigned, for start set 2, 30 percent of the assignments from start set 1 are reassigned, and so on. Then, for each of the start sets, parameters are again initialized and the $\log L_{STEP1}$ values are computed as described above. Next, the partitions of the random start sets and the best *mclust* set are ranked according to their $\log L_{STEP1}$ values and the best 25 start sets (i.e., the number of specified start sets) are selected as start partitions. For each start set, 15 iterations are performed by the EM algorithm (N.3.4). Subsequently, the testing strategy selects the start set with the highest $\log L_{STEP1}$ and saves the parameter estimates, $\hat{\theta}_r^{best}$, which serve as the initial values in the EM algorithm. Note that, instead of setting the number of iterations $v = 0$ (as at the beginning of the first iterations through the start sets), the algorithm continues with the number of iterations that have already been performed; that is $v = 15$.

N.4. Step 2 in *lmfa*

In step 2, the subject- and time-point-specific observations are classified into the states $\mathbf{w}_{it} = (w_{it1}, \ldots, w_{itK})$ based on the largest estimated posterior probability to belong to a state (i.e., based on a so-called "modal" state assignment). Thus, the observations are assigned to the MM that is most likely underlying the item responses. This can be expressed as $p(w_{itm} = 1|\mathbf{y}_{it}) = 1$ for state k with the largest $p(s_{itm} = 1|\mathbf{y}_{it})$. It is important to understand that the assignment of almost any observation includes some amount of uncertainty or "classification error" and can be calculated by conditioning the assigned state membership on the true state membership; that is $p(\mathbf{w}_{it} \mid \mathbf{s}_{itk})$. For details, see N.6.5. As stated before, the amount of classification error is related to the degree of state separation, which is quantified by the entropy-based R-squared measure $R^2_{entropy}$. The larger the state separation, the smaller the classification error. In *lmfa*, the state assignments, classification errors, and the $R^2_{entropy}$ can be obtained with the `step2()` function. For details about the arguments and the output, see the Appendix M.2.

N.5. Step 3 in *lmfa*

The third step of the 3S procedure consists of estimating the transition model by means of a single indicator CT-LMM (with covariates), which automatically corrects for the classification uncertainty from step 2. More specifically, the single indicator model is:

$$p(\mathbf{W}_i|\mathbf{Z}_i) = \sum_{\mathbf{s}_{i1}} \cdots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}|\mathbf{z}_{i1}) \prod_{t=2}^{T} p(\mathbf{s}_{it}|\mathbf{s}_{it-1}, \mathbf{z}_{it}) \prod_{t=1}^{T} p(\mathbf{w}_{it}|\mathbf{s}_{it}), \qquad \text{(N.21)}$$

where the response probabilities $p(\mathbf{w}_{it}|\mathbf{s}_{it})$ are fixed to the classification errors of step 2 and $\mathbf{W}_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots, \mathbf{w}_{iT})$ are manifest single indicators (containing error) of the latent (error free) latent states $\mathbf{S}_i$ (Di Mari et al., 2016; Vogelsmeier et al., 2020). Note that $\mathbf{W}_i$ and $\mathbf{S}_i$ may differ, which is increasingly more likely for larger classification errors. In *lmfa*, the CT-LMM can be estimated with the `step3()` function. Details about the arguments and the output are given in Appendix M.3. In the following, we first show the log$L$ function that has to be maximized to obtain the ML estimates. Afterwards, we describe how the estimation is performed in the *lmfa* package, including the inherent convergence criteria and a multistart procedure.

N.5.1. Likelihood Function

The following log$L$ function has to be maximized to obtain the ML parameter estimates.

$$\log L_{STEP3} = \sum_{i=1}^{I} \log\big(p(\mathbf{W}_i|\mathbf{Z}_i)\big). \qquad \text{(N.22)}$$

Note that all parameters have been defined before.

N.5.2. Estimation

The model is estimated by means of an optimization routine from the *msm* package (Jackson, 2011). The corresponding function `msm()` can be used to estimate various types of CT-LMMs. In order to obtain the estimates, `msm()` itself uses the `optim()` function, which performs "general-purpose optimization" (R Core Team, 2020). In brief, instead of maximizing the log$L$, `optim()` minimizes a loss function that is equal to -2 times the log$L$. For details about the estimation procedure, we refer to the function documentations, that can be called with the commands `?msm` and `?optim`. The `step3()` function in *lmfa* is tailored to the type of model that needs to be estimated in the third step of LMFA (i.e., a

single indicator CT-LMM with response probabilities fixed to the classification errors that results from the modal state assignment in step 2). Thus, `step3()` can be seen as a wrapper that facilitates the usage of `msm()` by providing proper parameter specifications and constraints and showing the desired parameter estimates, including significance tests for covariate effects by means of Wald tests.

N.5.3. Convergence

The optimization of the loss function (N.5.2) stops when the convergence criterion (say, 1e-10, the default) is reached. More specifically, the optimization stops when the loss function (i.e., $-2\log L$) can no further be reduced by a factor equal to the specified tolerance times the sum of the absolute value of the loss function and the tolerance. Thus, when defining the reduction in the loss function as

$$\Delta_{-2logL} = -2\log L_{STEP3}^{v-1} - (-2\log L_{STEP3}^{v}), \tag{N.23}$$

the estimation stopes when $\Delta_{-2logL} < 1e-10$ ($|-2\log L_{STEP3}^{v}| + 1e-10$).

N.5.4. Multistart Procedure

The results of the CT-LMM are very sensitive to the start values of the log transition intensities. More specifically, as previously stated, the intensities are directly related to the size of the time unit in the dataset. If the unit of the intensities from which the algorithm starts are too far from the actual unit, the model estimation will likely end up in a local maximum. Therefore, the following multistart procedure has been implemented: First, (by default) 25 random diagonal transition probability matrices **P** are sampled with staying probabilities on the diagonal that lie between 0.5 and 1. Per row, the off-diagonal probabilities are set equal (considering the constraint that rows must sum to 1). Subsequently, the transition intensity matrices **Q** are obtained by taking the matrix logarithm of the transition probability matrices **P**. Then, the **Q** matrix is rescaled by dividing it by the average length of the time intervals. Next, for each start set, 10 initial iterations of the CT-LMM analysis are performed. For each set, the log$L$ values are obtained and ranked and the solution with the best log$L$ value is used for the final analysis. Note that the estimation is not sensitive to the start values of the initial state logits and the covariate effects, which are, therefore, simply initialized to zero (i.e., covariate effects are absent and probabilities to start in a state are equally likely).

N.6. Statistics

N.6.1. Number of Free Parameters

The number of free parameters, fp, is obtained as follows:

$$fp = \overbrace{K-1}^{state\ proportions} + \overbrace{K \times J}^{intercepts} + \overbrace{K \times J}^{unique\ variances} + \overbrace{K \times J \times F_k}^{loadings}. \quad \text{(N.24)}$$

Note that the number of activated constraints for small variances (see N.3.4) would be subtracted from fp.

N.6.2. BIC

The BIC considers complexity and parsimony by penalizing models with more parameters (fp) and larger sample size (N) as follows:

$$\text{BIC} = -2 \times \log L_{STEP1} + fp \times log(N), \quad \text{(N.25)}$$

with $\log L_{STEP1}$ as in Equation (N.5), fp as in Equation (N.24), and $N = \sum_{k=1}^{K} N_k$ with N_k as in Equation (N.10).

N.6.3. CHull Scree-Test Value

For all models (but the least and most complex model) on the upper boundary of the convex hull (CHull), the following scree-test value, st, is obtained:[111]

$$st_n = \frac{\left(\frac{\log L_{STEP1,n} - \log L_{STEP1,n-1}}{fp_n - fp_{n-1}}\right)}{\left(\frac{\log L_{STEP1,n+1} - \log L_{STEP1,n}}{fp_{n+1} - fp_n}\right)}, \quad \text{(N.26)}$$

where the index n denotes the nth hull model. Note that the numerator and the denominator pertain to the slopes of two consecutive parts of the upper boundary of the

[111] For the procedure to obtain the models on the upper boundary, see Bulteel et al. (2013) and Vervloet, Wilderjans, Durieux, and Ceulemans (2017).

CHull. Large values of st_n indicate that model n fits clearly better than model $n-1$, while model $n+1$ only leads to a small increase in model fit (Bulteel et al., 2013).

N.6.4. Explained Variance

The amount of explained variance, EV, is calculated by taking the sum of squares of the standardized loadings per state, SS_k, dividing them by the number of items, J, weighting them by the state proportions, $p(s_{itk}=1)$, and , finally, adding them up across states K. Thus,

$$SS_k = \sum_{j=1}^{J}\sum_{f=1}^{F_k}\left(\lambda_{kjf}\right)^2, \tag{N.27}$$

$$EV = \sum_{k=1}^{K} SS_k/J \times\ p(s_{itk}=1). \tag{N.28}$$

N.6.5. Classification Error

The classification error is obtained by conditioning the assigned state $w_{itm}=1$ on the true, latent state $s_{itk}=1$ for all states $m, k = 1, \ldots, K$. That is:

$$p(w_{itm}=1 \mid s_{itk}=1) = \frac{\frac{1}{I\times T}\sum_{i=1}^{I}\sum_{t=1}^{T} p(w_{itm}=1|\mathbf{y}_{it})p(s_{itk}=1|\mathbf{y}_{it})}{p(s_{itk}=1)}. \tag{N.29}$$

For a derivation, see Vogelsmeier et al. (2020). Note that $p(s_{itk}=1)$ are the state proportions (Equation (N.11) and $p(w_{itm}=1|\mathbf{y}_{it})$ are the modal state assignments based on the posterior state-membership probabilities $p(s_{itk}=1|\mathbf{y}_{it})$ (Equation (N.9).

N.6.6. R-squared Entropy

The state separation in terms of the R-square measure $R^2_{entropy}$ can be calculated as follows:

$$R^2_{entropy} = \frac{Entropy(\mathbf{S}) - Entropy(\mathbf{S}|\mathbf{Y})}{Entropy(\mathbf{S})} = 1 - \frac{Entropy(\mathbf{S}|\mathbf{Y})}{Entropy(\mathbf{S})} \tag{N.30}$$

with

$$Entropy(\mathbf{S}) = \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{k=1}^{K} -p(s_{itk} = 1)\log p(s_{itk} = 1) \tag{N.31}$$

and

$$Entropy(\mathbf{S}|\mathbf{Y}) = \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{k=1}^{K} -p(s_{itk} = 1|\mathbf{y}_{it})\log p(s_{itk} = 1|\mathbf{y}_{it}). \tag{N.32}$$

Thus, Equation (N.30) shows that the $R^2_{entropy}$ measure determines the relative improvement in predicting the state memberships given the observations (numerator) versus predicting the state memberships without the observations (denominator).

N.6.7. Wald-test Statistic

The Wald statistic $W^{\mathbf{2}}$ is computed as follows:

$$W^{\mathbf{2}} = \theta_u' \widehat{\Sigma}\,(\theta_u)^{-1}\theta_u \tag{N.33}$$

with θ_u and $\widehat{\Sigma}\,(\theta_u)^{-1}$ indicating the slopes and the estimated variance covariance matrix for covariate u, respectively. Note that the degrees of freedom (df) are equal to the number of constrained parameters (e.g., df = 6 when testing the significance of a covariate with six slopes).

N.6.8. Factor Scores

The state-specific factor score estimates, $\hat{\mathbf{F}}_{ik}$, are obtained by means of the regression method (Thomson, 1934; Thurstone, 1935):

$$\hat{\mathbf{F}}_{ik} = \mathbf{Y}_i\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-\mathbf{1}}\boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k, \tag{N.34}$$

with $\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k\boldsymbol{\Lambda}_k'$.

Bibliography

Abramowitz, M., & Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Washington, D.C.: Dover Publications.

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: a distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology, 5*, 1–14. doi:10.3389/fpsyg.2014.00883

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573. doi:10.1007/BF02293814

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2016). Dynamic latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 257–269. doi:10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic structural equation models. Technical Report. doi:10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 329–341. doi:10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2014b). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 495–508. doi:10.1080/10705511.2014.919210

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the Association between Latent Class Membership and External Variables Using Bias-adjusted Three-step Approaches. *Sociological Methodology, 43*, 272–311. doi:10.1177/0081175012470644

Baltagi, B. H. (2011). *Econometrics* (5th ed.). Berlin, Germany: Springer.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion, 12*, 579–599. doi:10.1080/026999398379574

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion, 15*, 713–724. doi:10.1080/02699930143000239

Bartolucci, F. (2006). Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society, series B, 68*, 155–178.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). Comments on: Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. *Test, 23*, 473–477. doi:10.1007/s11749-014-0387-1

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2015). *Latent Markov Models for Longitudinal Data*. Hoboken, NJ: CRC Press.

Bartolucci, F., Montanari, G. E., & Pandolfi, S. (2015). Three-step estimation of latent Markov models with covariates. *Computational Statistics & Data Analysis, 83*, 287–301. doi:10.1016/j.csda.2014.10.017

Bartolucci, F., & Solis-Trapala, I. L. (2010). Multidimensional latent Markov models in a developmental study of inhibitory control and attentional flexibility in early childhood. *Psychometrika, 75*, 725–743. doi:10.1007/s11336-010-9177-1

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., . . . Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research, 137*, 1–14. doi:10.1016/j.jpsychores.2020.110211

Bastiaansen, J. A., Ornee, D. A., Meurs, M., & Oldehinkel, A. J. (2020). An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). *Psychological Medicine*, 1–10. doi:10.1017/S0033291720004845

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757–786. doi:10.1080/00273170701710338

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22*, 507–526. doi:10.1037/met0000077

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics, 41*, 164–171. doi:10.1214/aoms/1177697196

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York, NY: Guilford Press.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9*, 65–73. doi:10.1016/j.coph.2008.12.014

Billiet, J. B., & McClendon, M. J. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608-628. doi:10.1207/s15328007sem0704_5

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (5th ed.). New York: Springer.

Böckenholt, U. (2005). A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychological Methods, 10*, 65–83. doi:10.1037/1082-989X.10.1.65

Bokhorst, C. L., Sumter, S. R., & Westenberg, P. M. (2010). Social support from parents, friends, classmates, and teachers in children and adolescents aged 9 to 18 years: Who Is perceived as most supportive? *Social Development, 19*, 417–426. doi:10.1111/j.1467-9507.2009.00540.x

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*, 3–27. doi:10.1093/pan/mph001

Bolt, D. M., Cohen, A., & Wollack, J. (2001). A mixture item response model for multiple-choice data. *Journal of Educational and Behavioral Statistics, 26*, 381–409. doi:10.3102/10769986026004381

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement, 33*, 335–352. doi:10.1177/0146621608329891

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*, 203–219. doi:10.1037/0033-295X.110.2.203

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38*(1), 25-56. doi:10.1207/S15327906MBR3801_2

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., . . . Tuerlinckx, F. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One, 8*, e60188. doi:10.1371/journal.pone.0060188

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150. doi:10.1207/S15327906MBR3601_05

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods, 45*, 782–791. doi:10.3758/s13428-012-0293-y

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466. doi:10.1037/0033-2909.105.3.456

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods, 49*, 988–1005. doi:10.3758/s13428-016-0754-9

Carvalho, C. M., & Lopes, H. F. (2007). Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis, 51*, 4526–4542. doi:10.1016/j.csda.2006.07.019

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276. doi:10.1207/s15327906mbr0102_10

Ceulemans, E., & Kiers, H. A. (2006). Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology, 59*, 133–150. doi:10.1348/000711005X64817

Ceulemans, E., & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: interrelations and model selection. *Psychometrika, 70*, 461–480. doi:10.1007/s11336-003-1067-3

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005–1018. doi:10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*, 187–212. doi:10.1177/0022022100031002003

Clarkson, D. B., & Jennrich, R. I. (1988). Quartic rotation criteria and algorithms. *Psychometrika, 53*, 251–259. doi:10.1007/BF02294136

Cohen, A., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement, 42*, 133–148. doi:10.1111/j.1745-3984.2005.00007

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New York: Wiley.

Cox, D. R., & Miller, H. D. (1965). *The Theory of Stochastic Process*. London: Chapman & Hall.

Crayen, C., Eid, M., Lischetzke, T., & Vermunt, J. K. (2017). A continuous-time mixture latent-state-trait Markov model for experience sampling data. *European Journal of Psychological Assessment, 33*, 296–311. doi:10.1027/1015-5759/a000418

Csikszentmihalyi, M., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology* (pp. 35–54). Dordrecht: Springer.

D'Urso, E. D., Tijmstra, J., Vermunt, J. K., & De Roover, K. (2021). *Awareness is bliss: How acquiescence affects exploratory factor analysis*. Preprint. Retrieved from https://psyarxiv.com/pdvym/

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology, 8*, 1–19. doi:10.3389/fpsyg.2017.01849

de Jong, V. M. T., Eijkemans, M. J. C., van Calster, B., Timmerman, D., Moons, K. G. M., Steyerberg, E. W., & van Smeden, M. (2019). Sample size considerations and predictive performance of multinomial logistic prediction models. *Statistics in Medicine, 38*, 1601–1619. doi:10.1002/sim.8063

De Roover, K., Timmerman, M. E., De Leersnyder, J., Mesquita, B., & Ceulemans, E. (2014). What's hampering measurement invariance: detecting non-invariant items using clusterwise simultaneous component analysis. *Frontiers in Psychology, 5*, 1–11. doi:10.3389/fpsyg.2014.00604

De Roover, K., Timmerman, M. E., Van Diest, I., Onghena, P., & Ceulemans, E. (2014). Switching principal component analysis for modeling means and covariance changes over time. *Psychological Methods, 19*, 113–132. doi:10.1037/a0034525

De Roover, K., & Vermunt, J. K. (2019). On the exploratory road to unraveling factor loading non-invariance: A new multigroup rotation approach. *Structural Equation Modeling: A Multidisciplinary Journal, 26*, 905–923. doi:10.1080/10705511.2019.1590778

De Roover, K., Vermunt, J. K., Timmerman, M. E., & Ceulemans, E. (2017). Mixture simultaneous factor analysis for capturing differences in latent variables between higher level units of multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 1–18. doi:10.1080/10705511.2017.1278604

de Winter, J. C., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*, 147–181. doi:10.1080/00273170902794206

Dejonckheere, E., Kalokerinos, E. K., Bastian, B., & Kuppens, P. (2019). Poor emotion regulation ability mediates the link between depressive symptoms and affective bipolarity. *Cognition and Emotion, 33*, 1076–1083. doi:10.1080/02699931.2018.1524747

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., . . . Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology, 114*, 323–341. doi:10.1037/pspp0000186

Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., & Kalokerinos, E. K. (2019). The relation between positive and negative affect becomes more negative in response to personally relevant events. *Emotion*. doi:10.1037/emo0000697

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*, 1–38.

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement, 76*, 741–770. doi:10.1177/0013164415607618

Di Mari, R., Oberski, D. L., & Vermunt, J. K. (2016). Bias-adjusted three-step latent Markov modeling with covariates. *Structural Equation Modeling: A Multidisciplinary Journal, 23*, 649–660. doi:10.1080/10705511.2016.1191015

Dias, J. G., Vermunt, J. K., & Ramos, S. (2008). *Heterogeneous hidden Markov models: COMPSTAT 2008 Proceedings*, Porto, Portugal: Physica.

Dietvorst, E., Hiemstra, M., Maciejewski, D., van Roekel, E., ter Bogt, T. F. M., Hillegers, M. H. J., & Keijsers, L. (2021). Grumpy or depressed? Disentangling typically developing adolescent mood from prodromal depression using experience sampling methods. *Journal of Adolescence, 88*, 25–35. doi:10.1016/j.adolescence.2021.01.009

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309–326.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 295–314. doi:10.1080/10705510902751416

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R Package ctsem. *Journal of Statistical Software, 77*, 1–35. doi:doi:10.18637/jss.v077.i05

Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-Item positive and negative affect schedule for children, child and parent shortened versions: Application of item response theory for more efficient assessment. *Journal of Psychopathology and Behavioral Assessment, 34*, 191–203. doi:10.1007/s10862-011-9273-2

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*. doi:10.1177/1073191120957102

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry, 46*, 971–982. doi:10.1001/archpsyc.1989.01810110013002

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika, 85*, 206–231. doi:10.1007/s11336-020-09697-3

Erbas, Y., Ceulemans, E., Kalokerinos, E., Houben, M., Koval, P., Pe, M. L., & Kuppens, P. (2018). Why I don't always know what I'm feeling: The role of stress in within-

person fluctuations in emotion differentiation. *Journal of Personality and Social Psychology, 115*, 179–191.

Erbas, Y., Ceulemans, E., Koval, P., & Kuppens, P. (2015). The role of valence focus and appraisal overlap in emotion differentiation. *Emotion, 15*, 373–382. doi:10.1037/emo0000039

Erbas, Y., Kalokerinos, E., Kuppens, P., van Halem, S., & Ceulemans, E. (2020). Momentary Emotion Differentiation: The derivation and validation of a framework to study within-person fluctuations in emotion differentiation. *Assessment*. doi:10.31234/osf.io/v6u7t

Ethica Data Services Inc. (2018). Ethica Data. Retrieved from https://ethicadata.com/

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology, 69*, 153–166. doi:10.1037/0022-3514.69.1.153

Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology, 63*, 427–448. doi:10.1348/000711009X470740

Flynn, M., & Rudolph, K. D. (2010). The contribution of deficits in emotional clarity to stress responses and depression. *Journal of Applied Developmental Psychology, 31*, 291–297. doi:10.1016/j.appdev.2010.04.004

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: an illustration of longitudinal measurement invariance. *Psychol Assessment, 25*, 520–531. doi:10.1037/a0031669

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time . . . Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28*, 1354–1367. doi:10.1037/pas0000275

Galinha, I. C., Pereira, C. R., & Esteves, F. G. (2013). Confirmatory factor analysis and temporal invariance of the Positive and Negative Affect Schedule (PANAS). *Psicologia: Reflexão e Crítica, 26*, 671–679.

Geminiani, E., Ceulemans, E., & De Roover, K. (2020). Testing for factor loading differences in mixture simultaneous factor analysis: a Monte Carlo simulation-based perspective. *Structural Equation Modeling: A Multidisciplinary Journal*. doi:10.1080/10705511.2020.1807351

Goodman, L. A. (2007). On the assignment of individuals to latent classes. *Sociological Methodology, 37*, 1–22. doi:10.1111/j.1467-9531.2007.00184.x

Green, B. F. (1976). On the factor score controversy. *Psychometrica, 41*, 263–266. doi:10.1007/BF02291843

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*, 430–450. doi:10.1037/1082-989X.6.4.430

Grühn, D., Lumley, M. A., Diehl, M., & Labouvie-Vief, G. (2013). Time-based indicators of emotional complexity: interrelations and correlates. *Emotion, 13*, 226–237. doi:10.1037/a0030363

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5*, 1–16. doi:10.3389/fpsyg.2014.00980

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. New York: Cambridge University Press.

Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (2014). *Multivariate Data Analysis: Pearson International Edition*. Edinburgh, UK: Pearson.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Meehl & T. S. Conner (Eds.), (pp. 43–61). New York, NY: Guilford.

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review, 7*, 316–322. doi:10.1177/1754073915590619

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2010). Regime-switching models to study psychological processes. In P. C. M. Molenaar & K. M. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (pp. 155–168). Washington, DC: American Psychological Association.

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present. *Current Directions in Psychological Science, 26*, 10–15. doi:10.1177/0963721416666518

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.

Heininga, V. E., & Kuppens, P. (2020). Psychopathology and positive emotions in daily life. doi:10.31234/osf.io/yse56

Heininga, V. E., Van Roekel, E., Ahles, J. J., Oldehinkel, A. J., & Mezulis, A. H. (2017). Positive affective functioning in anhedonic individuals' daily life: Anything but flat and blunted. *Journal of Affective Disorders, 218*, 437–445. doi:10.1016/j.jad.2017.04.029

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection - A review and recommendations for the practicing statistician. *Biometrical Journal, 60*, 431–449. doi:10.1002/bimj.201700067

Hofmann, B. U., & Meyer, T. D. (2006). Mood fluctuations in people putatively at risk for bipolar disorders. *British Journal of Clinical Psychology, 45*, 105–110. doi:10.1348/014466505X35317

Horstmann, K. T., & Ziegler, M. (2020). Assessing Personality States: What to Consider when Constructing Personality State Measures. *European Journal of Personality*. doi:10.1002/per.2266

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141*, 901–930. doi:10.1037/a0038822

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193–218.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software, 38*, 1–29.

Jackson, C. H., & Sharples, L. D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine, 21*, 113–128. doi:10.1002/sim.886

Janssen, L. H. C., Elzinga, B. M., Verkuil, B., Hillegers, M. H. J., & Keijsers, L. (2020). The link between parental support and adolescent negative mood in daily life: between-person heterogeneity in within-person processes. *Journal of Youth and Adolescence*. doi:10.1007/s10964-020-01323-w

Jennrich, R. I., & Sampson, P. F. (1976). Newton–Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics, 18*, 11–17. doi:10.2307/1267911

Johnson, T. R., & Bolt, D. M. (2010). On the Use of Factor-Analytic Multinomial Logit Item Response Models to Account for Individual Differences in Response Style. *Journal of Educational and Behavioral Statistics, 35*, 92–114. doi:10.3102/1076998609340529

Jolliffe, I. T. (1986). *Principal Component Analysis*. NewYork, NY: Springer.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books.

Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200. doi:10.1007/BF02289233

Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association, 80*, 863–871. doi:10.2307/2288545

Kang, S. M., & Shaver, P. R. (2004). Individual differences in emotional complexity: their psychological implications. *Journal of Personality, 72*, 687–726. doi:10.1111/j.0022-3506.2004.00277.x

Kang, X., & Thakor, N. V. (2012). *Factor analyzed hidden Markov models for estimating prosthetic limb motions using premotor cortical ensembles*. Paper presented at the IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, Roma, Italy.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research, 40*, 279–310. doi:10.1177/0049124111405301

Kappenburg -ten Holt, J. (2014). *A comparison between factor analysis and item response theory modeling in scale analysis.* (Doctoral dissertation), Retrieved from http://www.rug.nl/research/portal

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation. *Current Directions in Psychological Science, 24*, 10–16. doi:10.1177/0963721414550708

Kauer, S. D., Reid, S. C., Crooke, A. H., Khor, A., Hearps, S. J., Jorm, A. F., . . . Patton, G. (2012). Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *Journal of Medical Internet Research, 14*, e67. doi:10.2196/jmir.1858

Keijsers, L., Boele, S., & Bülow, A. (2017). *ADAPT: Assessing the Dynamics between Parenting and Adaptation in Teens*. Retrieved from: osf.io/svyau

Keijsers, L., Boele, S., & Bülow, A. (2019). Dutch Postivite and Negative Affect Schedule for Children (PANAS-C-NL) retrieved from the ADAPT project: Assessing the Dynamics between Parenting and Adaptation in Teens. osf.io/svyau

Kelava, A., & Brandt, H. (2019). A nonlinear dynamic latent class structural equation model. *Structural Equation Modeling: A Multidisciplinary Journal, 26*, 509–528. doi:10.1080/10705511.2018.1555692

Kendall, A. D., Wilt, J., Walls, C. E., Scherer, E. A., Beardslee, W. R., Revelle, W., & Shrier, L. A. (2014). The social context of positive and negative affective states in depressed youth. *Journal of Social and Clinical Psychology, 33*, 805–830. doi:10.1521/jscp.2014.33.9.805

Kiers, H. A. (1997). Techniques for rotating two or more loading matrices to optimal agreement and simple structure: A comparison and some technical details. *Psychometrika, 62*, 545–568. doi:10.1007/BF02294642

Kim, E. S., & Wang, Y. (2017). Class enumeration and parameter recovery of growth mixture modeling and second-order growth mixture modeling in the presence of measurement noninvariance between latent classes. *Frontiers in Psychology, 8*, 1–18. doi:10.3389/fpsyg.2017.01499

Kim, M., Vermunt, J. K., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*, 601–614. doi:10.1080/10705511.2016.1158655

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis* (3 ed.). New York, NY: Springer.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal Psychtherapy of Depression*. New York, NY: Basic Books Inc Publishers.

Kovacs, M. (1992). *Children's Depression Inventory*. New York, NY: Multi-Health Systems.

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., . . . Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry, 13*, 68–77. doi:10.1002/wps.20090

Kunkels, Y. K., Riese, H., Knapen, S. E., Riemersma-van der Lek, R. F., George, S. V., van Roon, A. M., . . . Wichers, M. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry, 11*. doi:10.1038/s41398-021-01465-w

Larson, R. W. (1983). Adolescents' daily experience with family and friends: Contrasting opportunity systems. *Journal of Marriage and the Family, 45*, 739–750. doi:10.2307/351787

Larson, R. W., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology* (pp. 21–34). Dordrecht, The Netherlands: Springer.

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician, 12*, 209–229.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. New York: Houghton Mifflin.

Lee, S. Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika, 44*, 99–113. doi:10.1007/BF02293789

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics, 50*, 325–335.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*, 357–365. doi:10.1177/0165025407077757

Liu, M., Harbaugh, A. G., Harring, J. R., & Hancock, G. R. (2017). The effect of extreme response and non-extreme response styles on testing measurement invariance. *Frontiers in Psychology, 8*, 1–15. doi:10.3389/fpsyg.2017.00726

Lommen, M. J., van de Schoot, R., & Engelhard, I. M. (2014). The experience of traumatic events disrupts the measurement invariance of a posttraumatic stress scale. *Frontiers in Psychology, 5*, 1–7. doi:10.3389/fpsyg.2014.01304

Long, S. J. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.

Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica, 14*, 41–67.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research, 34*, 347–365. doi:10.1207/S15327906MBR3403_3

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*, 21–39. doi:10.1037/1082-989X.10.1.21

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology, 40*, 247–283. doi:10.1111/j.1467-9531.2010.01231.x

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin, 11*, 490–504. doi:10.1037/0033-2909.111.3.490

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*, 611–637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99.

Maciejewski, D. F., Keijsers, L., van Lier, P. A. C., Branje, S. J. T., Meeus, W. H. J., & Koot, H. M. (2019). Most fare well-But some do not: Distinct profiles of mood variability development and their association with adjustment during adolescence. *Developmental Psychology, 55*, 434–448. doi:10.1037/dev0000650

Mackinnon, A., Jorm, A., Christensen, H., Korten, A., Jacomb, P., & Rodgers, B. (1999). A short form of the positive and negative affect schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences, 27*, 405–416.

Magidson, J., & Vermunt, J. K. (2007). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. *Sawtooth Software Conference Proceedings*, 139–154.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., . . . Maris, G. (2018). An introduction to network psychometrics: Relating ising network models to item response theory models. *Multivariate Behavioral Research, 53*, 15–35. doi:10.1080/00273171.2017.1379379

Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicators multiple cause modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 180–197. doi:10.1080/10705511.2016.1254049

McDonald, R. P., & Krane, W. R. (1979). A Monte Carlo study of local identifiability and degrees of freedom in the asymptotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology, 32*, 121–132. doi:10.1111/j.2044-8317.1979.tb00757.x

McLachlan, G. J., & Peel, D. (2000). Mixtures of factor analyzers. In P. Langley (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning.* (pp. 599–606). San Francisco: Morgan Kaufmann.

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods, 25*, 610–635. doi:10.1037/met0000250

McNeish, D., & Harring, J. R. (2017). The effect of model misspecification on growth mixture model class enumeration. *Journal of Classification, 34*, 223–248. doi:10.1007/s00357-017-92

McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in Intensive Longitudinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–16. doi:10.1080/10705511.2021.1915788

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods, 52*, 2287–2305. doi:10.3758/s13428-020-01398-0

McNicholas, P. D. (2016). *Mixture Model-Based Classification*. Boca Raton: Chapman and Hall/CRC Press.

McNicholas, P. D., ElSherbiny, A., McDaid, A. F., & Murphy, T. B. (2019). pgmm: Parsimonious Gaussian Mixture Models. Retrieved from https://CRAN.R-project.org/package=pgmm

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437–455. doi:10.1037/a0028085

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, 69–77. doi:10.1097/01.mlr.0000245438.73837.89

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Molenaar, P. C., & Campbell, G. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112–117.

Moler, C., & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev, 45*, 3–49. doi:10.1137/S00361445024180

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*, 277–302. doi:10.1023/A:1024472110002

Morren, M., Gelissen, J. P., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: a restricted latent class factor analysis approach. *Sociological Methodology, 41*, 13–47. doi:10.1111/j.1467-9531.2011.01238.x

Muraki, E. (1992). A generalized partial credit model: Application of an em algorithm. *Applied Psychological Measurement, 16*, 159–176. doi: 10.1002/j.2333-8504.1992.tb01436.x

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 81–117. doi:10.2333/bhmk.29.81

Muthén, L. K., & Muthén, B. (2005). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles, CA: Muthén & Muthén.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry, 17*, 123–132. doi:10.1002/wps.20513

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90-100. doi:10.1016/s0022-2496(02)00028-7

Nagelkerke, E., Oberski, D. L., & Vermunt, J. K. (2016). Power and type I error of local fit statistics in multilevel latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 216-229. doi:10.1080/10705511.2016.1250639

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill,* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A Latent Transition Mixture Model Using the Three-Step Specification. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 439–454. doi:10.1080/10705511.2014.915375

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal, 23*, 782–797. doi:10.1080/10705511.2016.1221313

Oberski, D. L. (2017). Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis, 22*, 45–60. doi:10.1093/pan/mpt014

Oberski, D. L., van Kollenburg, G. H., & Vermunt, J. K. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Data Analysis and Classification, 7*(3), 267-279. doi:10.1007/s11634-013-0146-2

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika, 65*, 167–185. doi:10.1007/BF02294372

Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14*, 481–500. doi:10.1207/s15327906mbr1404_7

Oort, F. J., Visser, M. R., & Sprangers, M. A. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research, 14*, 599–609. doi:10.1007/s11136-004-0831-x

Paulhus, D. L. (1991). Measures of personality and social psychological attitudes. In J. P. Robinson & R. P. Shaver (Eds.), *Measures of Social Psychological Attitudes Series* (Vol. 1, pp. 17–59). San Diego, CA: Academic Press.

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics, 32*, 153–161. doi:10.1023/A:1015210025234

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. doi:10.1016/j.dr.2016.06.004

R Core Team. (2002). A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

R Core Team. (2020). A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ram, N., Brose, A., & Molenaar, P. C. M. (2012). Dynamic factor analysis: Modeling person-specific process. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 2, pp. 441–457). New York, NY: Oxford University Press.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373. doi:10.1037/a0029315

Rindskopf, D. (1984a). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research, 13*, 109–119. doi:10.1177/0049124184013001004

Rindskopf, D. (1984b). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*, 37–47. doi:10.1007/BF02294204

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.

Rost, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271–282. doi:10.1177/014662169001400305

Rosti, A. I., & Gales, M. J. (2002). *Factor analysed hidden Markov models: ICASSP 2002 Proceedings*, Florida, USA: IEEE.

Russell, J. A. (1980). A circumplex model of affect. *Journal ofPersonality and Social Psychology, 39*, 1161–1178. doi:10.1037/h00 77714

Ryan, O. (2020). *Dynamic systems and causal structures in psychology: Connecting data and theory.* (Doctoral dissertation, Utrecht University, Utrecht, The Netherlands), Retrieved from file:///C:/Users/vogelsme/Desktop/oryandissertation%20-%205f685e37d78e8.pdf#page=221&zoom=100,94,666

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika-Monograph-Supplement, 34*.

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal Invariance of Adult Psychometric Ability Factor Structures Across 7 years. *Psychology and Aging, 13*, 8–20. doi:10.1080/13825580490511134

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*, 95–113. doi:10.1177/0013164410387348

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Scollon, C., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5–34. doi:10.1023/A:1023605205115

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal, 8*, 289–317.

Silk, J. S., Forbes, E. E., Whalen, D. J., Jakubcak, J. L., Thompson, W. K., Ryan, N. D., . . . Dahl, R. E. (2011). Daily emotional dynamics in depressed youth: a cell phone ecological momentary assessment study. *Journal of Experimental Child Psychology, 110*, 241–257. doi:10.1016/j.jecp.2010.10.007

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods, 46*, 41–54. doi:10.3758/s13428-013-0353-y

Skrondal, A., & Rabe-Hesketh, S. (2011). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* New York, NY: Chapman & Hall/CRC.

Smit, J. A., Kelderman, H., & van der Flier, H. (2000). The mixed Birnbaum model: Estimation using collateral information. *Methods of Psychological Research Online, 5*, 31–43.

Snippe, E., Simons, C. J., Hartmann, J. A., Menne-Lothmann, C., Kramer, I., Booij, S. H., . . . Wichers, M. (2016). Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology, 35*, 433–441. doi:10.1037/hea0000312

Soenens, B., Deci, E. L., & Vansteenkiste, M. (2017). How parents contribute to children's psychological health: The critical role of psychological need support. *Development of self-determination through the life-course, 13*, 171–187. doi:10.1007/978-94-024-1042-6_13

Song, X., Xia, Y., & Zhu, H. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics, 73*, 313–323. doi:10.1111/biom.12536

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.

Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371–384.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–107.

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods, 9*, 386–396. doi:10.1037/1082-989X.9.3.386

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: recommendations and cautions. *Psychological Methods, 16*, 63–79. doi:10.1037/a0022673

Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thomson, G. H. (1934). The meaning of i in the estimate of g. *British Journal of Psychology, 25*, 92–99. doi:10.1111/j.2044-8295.1934.tb00728.x

Thurstone, L. L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: The University of Chicago Press.

Tijmstra, J., Bolsinova, M., & Jeon, M. (2018). General mixture item response models with different item response structures: Exposition with an application to Likert scales. *Behavior Research Methods, 50*(6), 2325–2344. doi:10.3758/s13428-017-0997-0

Timbremont, B., Braet, C., & Roelofs, J. (2008). *Handleiding Children's Depression Inventory (herziene versie)*. Amsterdam, The Netherlands: Pearson Assessment and Information B.V.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society, 61*, 611–622. doi:10.1111/1467-9868.00196

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience & Biobehavioral Reviews, 35*, 537–555. doi:10.1016/j.neubiorev.2010.06.006

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies. Personnel Research Section Report No. 984*. Washington, DC: Department of the Army.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1–67.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology, 6*, 1–4. doi:10.3389/fpsyg.2015.01064

van der Krieke, L., Emerencia, A. C., Bos, E. H., Rosmalen, J. G., Riese, H., Aiello, M., . . . de Jonge, P. (2015). Ecological momentary assessments and automated time series analysis to promote tailored health care: a proof-of-principle study. *JMIR research protocols, 4*, e100. doi:10.2196/resprot.4000

Van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika, 43*, 225–243. doi:10.1007/BF02293865

van Roekel, E., Bennik, E. C., Bastiaansen, J. A., Verhagen, M., Ormel, J., Engels, R. C., & Oldehinkel, A. J. (2015). Depressive Symptoms and the Experience of Pleasure in Daily Life: An Exploration of Associations in Early and Late Adolescence. *J Abnorm Child Psychol, 44*, 999–1009. doi:10.1007/s10802-015-0090-z

van Roekel, E., Goossens, L., Verhagen, M., Wouters, S., Engels, R. C. M. E., & Scholte, R. H. J. (2013). Loneliness, affect, and adolescents' appraisals of company: An Experience Sampling Method study. *Journal of Research on Adolescence, 24*, 350–363. doi:10.1111/jora.12061

van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical recommendations. *Journal of Research on Adolescence, 29*, 560–577. doi:10.1111/jora.12471

van Roekel, E., Scholte, R. H. J., Engels, R. C. M. E., Goossens, L., & Verhagen, M. (2014). Loneliness in the Daily Lives of Adolescents. *The Journal of Early Adolescence, 35*, 905–930. doi:10.1177/0272431614547049

Van Roekel, E., Vrijen, C., Heininga, V. E., Masselink, M., Bos, E. H., & Oldehinkel, A. J. (2017). An exploratory randomized controlled trial of personalized lifestyle advice and tandem skydives as a means to reduce anhedonia. *Behavior Therapy, 48*, 76–96. doi:10.1016/j.beth.2016.09.009

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*, 231–251. doi:10.1037/1082-989X.3.2.231

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*, 1–28. doi:10.1207/s15327906mbr2501_1

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research, 17*, 371–388. doi:10.1207/s15327906mbr1703_5

Vermunt, J. K. (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis, 18*, 450–469. doi:10.1093/pan/mpq025

Vermunt, J. K., Langeheine, R., & Böckenholt, U. (1999). Discrete-time discrete-state latent markov models with time constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24*, 179–207. doi:10.2307/1165200

Vermunt, J. K., & Magidson, J. (2005). Factor Analysis with Categorical Indicators: A Comparison Between Traditional and Latent Class Approaches. In A. v. d. Ark, M. A. Croon, & K. Sijtsma (Eds.), *New Developments in Categorical data Analysis for the Social and Behavioral Sciences* (pp. 41–62). Mahwah: Erlbaum.

Vermunt, J. K., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2020). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–9. doi:10.1080/10705511.2020.1818084

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of Longitudinal Research: Design, Measurement, and Analysis* (pp. 373–385). Burlington, MA: Elsevier.

Vervloet, M., Wilderjans, T. F., Durieux, J., & Ceulemans, E. (2017). *multichull: A Generic Convex-Hull-Based Model Selection Method*.

Visser, I. (2007). depmix: An R-package for fitting mixture models on mixed multivariate data with Markov dependencies. *R-package manual*.

Visser, I., Raijmakers, M. E. J., & van der Maas, H. L. J. (2009). Hidden Markov Models for Individual Time Series. In J. Valsiner, P.C.M. Molenaar, M.C.D.P. Lyra, & N. Chaudhary (Eds.), *Dynamic Process Methodology in the Social and Developmental Sciences* (pp. 269-289). New York: Springer-Verlag.

Voelkle, M. C., & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *Br. J. Math. Stat. Psychol., 66*, 103–126. doi:10.1111/j.2044-8317.2012.02043.x

Vogelsmeier, L. V. D. E., & De Roover, K. (2021). 'lmfa': An R-package for exploring measurement invariance in intensive longitudinal data with continuous-time latent Markov factor analysis. Retrieved from https://github.com/LeonieVm/lmfa

Vogelsmeier, L. V. D. E., Vermunt, J. K., Böing-Messing, F., & De Roover, K. (2019). Continuous-time latent Markov factor analysis for exploring measurement model changes across time. *Methodology, 15*, 29–42. doi:10.1027/1614-2241/a000176

Vogelsmeier, L. V. D. E., Vermunt, J. K., Bülow, A., & De Roover, K. (2020). *Evaluating covariate effects on ESM measurement model changes with latent Markov factor analysis: A three-step approach*. Preprint. Retrieved from https://doi.org/10.31234/osf.io/6ufrc

Vogelsmeier, L. V. D. E., Vermunt, J. K., Keijsers, L., & De Roover, K. (2021). Latent Markov latent trait analysis for exploring measurement model changes in intensive longitudinal data. *Evaluation & the Health Professions, 44*, 61–76. doi:10.1177/0163278720976762

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*, 557–575. doi:10.1080/10705511.2018.1554445

Wang, W.-L., Castro, L. M., Lachos, V. H., & Lin, T.-I. (2019). Model-based clustering of censored data via mixtures of factor analyzers. *Computational Statistics & Data Analysis, 140*, 104–121. doi:10.1016/j.csda.2019.06.001

Watson, D. (1992). Correcting for acquiescent response bias in the absence of a balanced scale. *Sociological Methods & Research, 21*, 52–88. doi:10.1177/0049124192021001003

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–235. doi:10.1037/0033-2909.98.2.219

Wedel, M., & Kamakura, W. A. (1998). *Market Segmentation: Concepts and Methodological Foundations*. Boston: Kluwer Academic Publishers.

Wichers, M., Groot, P. C., & Psychosystems, E. S. M. G. E. W. S. G. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics, 85*, 114–116. doi:10.1159/000441458

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: a generic convex-hull-based model selection method. *Behavior Research Methods*, 1–15. doi:10.3758/s13428-012-0238-5

Xia, Y.-M., Tang, N.-S., & Gou, J.-W. (2016). Generalized linear latent models for multivariate longitudinal measurements mixed with hidden Markov models. *Journal of Multivariate Analysis, 152*, 259–275. doi:10.1016/j.jmva.2016.09.001

Ximenez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification. *Behavior Research Methods, 41*, 1038–1052. doi:10.3758/BRM.41.4.1038

Ximénez, C. (2006). A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling, 13*, 587–614. doi:10.1207/s15328007sem1304_5

Zautra, A. J., Berkhof, J., & Nicolson, N. A. (2002). Changes in affect interrelations as a function of stressful events. *Cognition and Emotion, 16*, 309–318. doi:10.1080/02699930143000257

Zhang, G. (2006). *Bootstrap Procedures for Dynamic Factor Analysis*. Unpublished doctoral dissertation. The Ohio State University, Columbus, OH.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R (2nd Edition)*. Boca Raton: Chapman & Hall/CRC.

Summary

Studying dynamics in psychological constructs in intensive longitudinal data (ILD) becomes increasingly popular among researchers in the social and behavioral sciences. Technological advances facilitate gathering such ILD, for instance, with experience sampling methodology. However, before beginning with their analyses, researchers have to verify that the measured constructs are equivalent across subjects and time. To this end, the "measurement model" (MM) – indicating how items measure the constructs – needs to be invariant across subjects and time (i.e., "measurement invariance" (MI) must hold). If violations of MI are undetected or ignored, conclusions about (between-person differences in) within-person dynamics in the constructs may be invalid.

MI is often violated in ILD because response styles or item interpretation may not only differ across subjects, but subjects may also change in the way they respond to the questionnaire items over time. For instance, how subjects interpret an item could change depending on the context in which the questionnaire is completed, or subjects may start agreeing to all items once they are no longer motivated to complete the questionnaires.

Existing methods could only test a priori assumptions about MI violations. However, these assumptions are usually absent or incomplete, leaving the social and behavioral scientists with no efficient approach to explore the tenability of the MI assumption in their ILD. This dissertation aims to solve this gap by presenting latent Markov factor analysis (LMFA) for unraveling MM differences/changes for many subjects and time-points simultaneously. In LMFA, a latent Markov model (i.e., a latent class model that allows subjects to transition between dynamic latent classes or "states" over time) classifies observations based on their underlying MM into a few states and exploratory factor analysis per state evaluates the structure of the state-specific MMs. Observations within the same state have the same underlying MM and are thus validly comparable.

In this dissertation, we first introduce LMFA and show that the method performs well in recovering the state memberships and the MMs per state under a wide range of conditions (Chapter 2). Then, we extend LMFA by means of a continuous-time latent Markov model to adequately handle typically encountered unequally-spaced observations (Chapter 3). Next, we propose a three-step maximum likelihood estimation as an alternative to the originally employed one-step full information maximum likelihood estimation because this facilitates the inclusion of explanatory variables and thus offers researchers the possibility to understand why MMs differ/change (Chapter 4). Thereafter, we extend LMFA into latent Markov latent trait analysis that adequately handles categorical data, which allows researchers to investigate MI in their ILD also if the data contain responses with only a few categories and skewed distributions (Chapter 5). Finally, we provide a tutorial on how to investigate MI with our R package *lmfa*, which is openly available to all researchers in the social and behavioral sciences (Chapter 6).

Acknowledgments

First, I would like to thank my daily supervisor, Kim De Roover. Kim, you gave me the chance to start my PhD, and I could not be more grateful for your intense mentoring, supervision, your detailed and ever constructive feedback, for your empathy and support whenever I needed it, and for your immense trust in me, which taught me to also trust myself and become the researcher and teacher I am today. You walked the “philosopher's path” with me, both in academia and in Japan, during a vacation in connection with my first international conference. It was an experience I will never forget. In some ways, the path in academia and the journey in Japan were very similar. Sometimes the way to the goal was difficult and took a bit longer than planned (looking for errors in a simulation study or algorithm is a bit like looking for the right entrance in Tokyo's subway stations). From time to time, we encountered disappointing events (driving all the way to the Kegon Falls and not seeing the waterfall because it was so foggy is a bit like getting Reviewer 2’s comments after working so hard on a paper). Usually, though, the effort was totally worth it (I'm not sure if we smiled more when we were in the middle of a gigantic crowd watching the most amazing fireworks we've ever seen or when we received the NWO talent grant). Thank you, Kim, you were an awesome supervisor!

Next, I would like to express my sincere gratitude to my promotor, Jeroen Vermunt. Jeroen, on one of my first days in the department, you told me that your door was always open when I needed help. And indeed, whenever I had a question, you immediately took the time needed to answer it and get me back on track. I learned an incredible amount from you. You also granted me a lot of autonomy and you had great confidence in my own judgment and work, which helped me develop confidence in my own decisions. Additionally, you taught me that a PhD is not just about writing papers but also about personal development, enjoying conferences and vacations, and attending MTO borrels. Thank you for having been such an inspiring promotor, Jeroen!

I would also like to thank the members of my committee for their time to read and evaluate my dissertation and for their valuable suggestions. Eva Ceulemans, Ellen Hamaker, Jelte Wicherts, and Zsuzsa Bakk, it is an honor that you are in my committee. Eva, I would also like to thank you for making an online research visit possible in times of Covid-19 and for welcoming me so kindly to Leuven together with Rosi.

Moreover, I am grateful to my co-authors for their valuable contributions to the chapters and our insightful discussions about real data applications. I really enjoyed collaborating with and learning from you.

I would also like to thank all my current and former colleagues at the Department of Methodology and Statistics. I was always looking forward to coming to work, and I am very grateful for all the fun discussions we had and for all the things I learned from you.

Specifically, I would like to thank all the colleagues who provided their helpful comments for the talent grant proposal and helped me prepare for the NWO interview. Thanks also to all the members of the Vici and Psychometrics lab groups for reading my chapters and giving me constructive feedback, and for the possibility of learning about your projects. Moreover, I am grateful to my colleagues whose doors were always open to ask for help when I did not know how to solve a problem I encountered while consulting at the Elisabeth Hospital in Tilburg. Additionally, I would like to thank my colleagues who supported me in coordinating my first course. Furthermore, thanks for all the breaks in the form of "gezellige borrels" at the Esplanade, walks through the forest, and mindful yoga sessions in the evening. And last but not least, Anne-Marie, thank you for all your support, whether it was planning conference travel or an entire symposium. I really appreciated your help!

A special thanks goes to my great paranymphs, Esther and Damiano. I am so grateful that you stand behind me. Thank you both for countless hours ranting about measurement in psychology, for laughing at things that went wrong in research (or, more often, in life), for your unfiltered, honest opinions (as well as for accepting mine), and for the drinks and walks when it was time for some distraction from research. Thanks to you, I stayed motivated and had a lot of fun at work. Esther, thank you for being so inspiring and caring. Damiano, thank you for being such a supportive PhD-bro.

Before and during my time at Tilburg University, I met the most amazing friends both in Germany and in the Netherlands. While I was working on my dissertation, I was often very busy, but you all made sure that there was always something fun to distract me from work and enjoy life. Thank you for all the laughter, for giving me many great memories and experiences, and for coming all the way to Tilburg from Hemer, Iserlohn, Dortmund, Bocholt, Düsseldorf, Berlin, Hamburg, Munich, Amsterdam, or Utrecht. I am lucky and grateful to have you in my life.

On my first day at work, Jeroen told me that he had reserved a desk next to you, Florian, saying that he thought we would get along well. He couldn't have been more right. I am immensely grateful for the past five years with you. There hasn't been a single year that we haven't had a great vacation, not a single month that you haven't tried to answer one of my many statistics questions, not a single week that you haven't made sure we had fun and relaxing distractions by hiking through nature or seeing friends and family, and not a single day that you haven't made me smile. Thank you for being my best friend, for being patient with me and supporting me when I feel anxious or stressed, and for genuinely being happy for me in moments of joy.

I would furthermore like to thank the Böing-Messings for welcoming me so warmly into their family. Thank you for all the long nights when we shared not only lots

of wine and delicious food but also lots of happiness, joy, and emotional support. I am grateful to be part of your family.

I am also deeply attached and thankful to my brothers, Sebastian and Benjamin Vogelsmeier. Sebastian, I am so grateful for your honesty and down-to-earth attitude and for the fact that we have the same sense of humor, which often made me laugh and forget about problems during my PhD. Benjamin, I greatly appreciate your ease of mind, and I am so grateful that you always share your positivity and optimism with me because this gave me a lot of positive energy while working on this dissertation.

My very special thanks go to my parents, Gabriele and Eberhard Vogelsmeier, who supported me every day in the last 31 years of my life and to whom I dedicate this dissertation. Mama, you have always been my role model, and Papa, I have always looked up to you. Thank you both for letting me follow my own path, for constantly believing in me, for your unconditional love, and for your never-ending enthusiasm for trying to understand what exactly it is I worked on during my PhD. I love you.

Tilburg, November 16, 2021 *Leonie V. D. E. Vogelsmeier*