Continuous-Time Latent Markov Factor Analysis for Exploring Longitudinal Measurement Invariance

Vogelsmeier, Leonie V.D.E [a], Vermunt, Jeroen K. [a], Böing-Messing, Florian [a,b], De Roover, Kim [a]

[a] Tilburg University, The Netherlands

[b] Jheronimus Academy of Data Science, The Netherlands

Vogelsmeier, Leonie V.D.E. (l.v.d.e.vogelsmeier@uvt.nl), Vermunt, Jeroen K. (j.k.vermunt@tilburguniversity.edu), De Roover, Kim (k.deroover@tilburguniversity.edu)
Department of Methodology and Statistics
PO Box 90153 5000 LE Tilburg, The Netherlands

Böing-Messing, Florian (florian.boeingmessing@gmail.com)
Jheronimus Academy of Data Science
Sint Janssingel 92, 5211 DA 's-Hertogenbosch, The Netherlands &
Department of Methodology and Statistics
PO Box 90153, 5000 LE Tilburg, The Netherlands

The research leading to the results reported in this paper was sponsored by the Netherlands Organization for Scientific Research (NWO) [Research Talent grant 406.17.517; Veni grant 451.16.004]. Correspondence concerning this paper should be addressed to Leonie V. D. E. Vogelsmeier, Department of Methodology and Statistics, PO Box 90153 5000 LE Tilburg. E-mail: l.v.d.e.vogelsmeier@uvt.nl

Online supplementary material for this article: Can be found online at: (announced later)

Acknowledgments: The authors would like to thank Damiano D'Urso for his helpful comments on earlier versions of this manuscript.

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 paper, 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.

Keywords: experience sampling, measurement invariance, factor analysis, latent Markov modeling, continuous-time

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, e.g., a traditional daily or weekly diary study or modern Experience Sampling Methodology (ESM; e.g., Scollon, Kim-Prieto, & Diener, 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 value 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 artefact such as response styles (e.g., an agreeing response style leads to higher loadings; Cheung & Rensvold, 2000) or substantive changes in item interpretation (e.g., an item becomes 'reconceptualized', meaning that it measures another factor; Oort, Visser, & Sprangers, 2005) may affect the MM differently over time. To efficiently evaluate such changes for multiple subjects simultaneously, Vogelsmeier, Vermunt, van Roekel, and De Roover (in press) proposed latent Markov factor analysis (LMFA), which clusters observations depending on the underlying MM into a few latent states, where each state gathers validly comparable observations as will be described in detail in Section 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 and the interval automatically becomes longer.

To accommodate unequally spaced measurement intervals, we extend LMFA in this paper, 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 to 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, Eid, Lischetzke, & Vermunt, 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 paper is organized as follows: Section 2 describes the data structure and explains the differences between CT- and DT-LMFA, but also how the DT approach may approximate CT and furthermore, the general model estimation. Section 3 presents a simulation study comparing the performance of CT-, DT-LMFA, and the DT approximation to CT. Section 4 continues with a CT-LMFA application. Section 5 raises some points for discussion and future research.

2. Method

2.1. Data Structure

The repeated measures observations (with multiple continuous variables) assumed to be 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.

2.2. LMFA

We first give the building blocks of the regular DT-LMFA (2.2.1), then describe the DT approximation of CT, referred to as discrete-time-grid-LMFA (DTG-LMFA) (2.2.2), before we present CT-LMFA (2.2.3).

2.2.1. DT-LMFA

The first building block of LMFA is a latent Markov model (LMM; Bartolucci, Farcomeni, & Pennoni, 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 a certain state *k* ($k = 1, \dots, K$) 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* ($l = 1, \dots, K$) 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 $\sum_{k=1}^{K} p_{lk} = 1$. The transition probabilities one obtains in practice depend on the interval length between measurements (e.g., the probabilities to stay in a state are larger if the interval length 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 a factor analysis (FA; Lawley & Maxwell, 1962) model, which defines the state-specific MMs. The state-specific factor model is

$$\mathbf{y}_{it} = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \mathbf{f}_{it} + \mathbf{e}_{it}, \quad (1)$$

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 employs an exploratory FA (EFA) approach, thus without a priori constraints on the factor loadings as would be the case with confirmatory FA (CFA).

From Equation (1) it becomes apparent that the state-specific MMs cannot only differ regarding their loadings $\mathbf{\Lambda}_k$ as described in Section 1 but also regarding their intercepts $\mathbf{\nu}_k$, unique variances $\mathbf{D}_k$, and factor covariances $\mathbf{\Psi}_k$, implying that LMFA explores all levels of measurement non-invariance (described in detail in, e.g., Meredith, 1993):

(1) Configural invariance (equal number of factors and zero loading pattern),

(2) Weak factorial invariance (equal loading values),

(3) Strong factorial invariance (equal intercepts) and

(4) Strict invariance (equal unique variances).

To identify the model, factor variances in $\mathbf{\Psi}_k$ are restricted to one and, by means of criteria to optimize the between-state agreement or simple structure of the factor loadings (e.g., oblimin, Clarkson & Jennrich, 1988; generalized Procrustes, Kiers, 1997), rotational freedom is dealt with. The multivariate normal distribution with the state-specific covariance matrices $\mathbf{\Sigma}_k = \mathbf{\Lambda}_k\mathbf{\Lambda}_k + \mathbf{D}_k$ 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:

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

where $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_T)$ is the state membership indicator matrix. 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.

2.2.2. DTG-LMFA

One possible 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 (e.g. 8.34 am becomes 9 am). The data matrix then contains a 'grid' with one record for each possible time-point (i.e., for one subject and one day, 24 rows for 24 hours). The observations belonging to the rounded time-points should be at the corresponding places in the data matrix and, if no observation at a particular time-point is available, a record with missing values on the indicators should be inserted (assuming missing data at random; MAR; Vermunt, Tran, & Magidson, 2008). When missing data is part of the data matrix, the response probabilities $p\left(\mathbf{y}_{it}|\mathbf{s}_t\right)$ are changed to $p\left(\mathbf{y}_{it}|\mathbf{s}_t\right)^{\kappa_{it}}$, where $\kappa_{it} = 1$ if subject *i* provides information for occasion *t* and $\kappa_{it} = 0$ otherwise. While for $\kappa_{it} = 1$ nothing changes, for $\kappa_{it} = 0$, $p\left(\mathbf{y}_{it}|\mathbf{s}_t\right)^0 = 1$, implying that responses of this time-point are skipped (Vermunt et al., 2008). The smaller the grid, the better can be accounted for unequal TIs (e.g. a 1-hour grid is better than a 2-hour grid). However, this is only an approximation and transforming the dataset might be burdensome for (e.g., substantive) researchers. A much more natural alternative is to account for the unequal TI in the model specification by using a CT-approach.

2.2.3. 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:

$$q_{lk} = \lim_{\delta \to 0} \frac{p(s_{tk}|s_{t-\delta,l})}{\delta}. \quad (3)$$

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, i.e., $-\sum_{k \neq l} q_{lk}$, on the diagonals (Cox & Miller, 1965). For example, for $K = 3$,

$$\mathbf{Q} = \begin{pmatrix} -(q_{12} + q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21} + q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31} + q_{32}) \end{pmatrix}. \quad (4)$$

There are three important 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, 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[1] of the intensity matrix $\mathbf{Q}$ times the TI δ (Cox & Miller, 1965):

$$\mathbf{P}(\delta) = e^{\mathbf{Q}\delta}. \quad (5)$$

With this equation, we can compute the transition probabilities for arbitrary TIs, which is, as mentioned in the introduction, a distinctive advantage of the CT approach. Thus, while the intensities are constant over time (and for simplicity in this paper also across subjects), 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 1 based on an arbitrary intensity matrix $\mathbf{Q}$.

[Insert Figure 1 about here]

As a final remark, note that the joint probability density of subject *i*'s observations and state-memberships for DT-LMFA in Equation (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.

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 loglikelihood

[1] The matrix exponential $e^{\mathbf{A}}$, where $\mathbf{A}$ can be any 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.

function $\log L$. Apart from the transition probability formulation in DT, 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:

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{i=1}^{I} \log\left(\sum_{\mathbf{s}_{i1}} \dots \sum_{\mathbf{s}_{iT}} p(\mathbf{s}_{i1}) \prod_{t=2}^{T} p_{\delta_{ti}}(\mathbf{s}_t|\mathbf{s}_{t-1}) \prod_{t=1}^{T} p\left(\mathbf{y}_{it}|\mathbf{s}_{it}\right)\right), \tag{6}$$

which is complicated by the latent states. Therefore, to find the maximum likelihood (ML) solution, for both DT- and CT-LMFA, LG utilizes the Expectation Maximization (EM; Dempster, Laird, & Rubin, 1977) algorithm, more specifically the forward-backward algorithm (Baum, Petrie, Soules, & Weiss, 1970), which is described in detail for DT-LMFA in Vogelsmeier et al. (in press). Estimation of the CT-LMFA differs in that the Maximization step (M-step) requires using a Fisher algorithm not only for the updating state-specific covariance matrices (Lee & Jennrich, 1979) but also for updating the log transition intensities (Kalbfleisch and Lawless (1985). A summary is provided in the supplementary material (Supplement A). The E- and the M-step are iterated until convergence (Supplement A.3).

3. Simulation study

3.1. Problem

We employed an ESM design with unequal TIs—because ESM is currently the go-to research design to study daily-life dynamics—to evaluate how the three analyses CT-LMFA, standard DT-LMFA, and DTG-LMFA with a 1-hour grid differ in recovering the model parameters. Generally, we expected CT-LMFA to outperform the DT models because the former takes the unequal time lags better into account, although the performance difference might be small (Crayen et al., 2017). We manipulated three types of conditions that likely affect parameter recovery: (1) model complexity, (2) state separability and (3) amount of information available for estimation. The conditions for (1) model complexity—with (a) number of factors and (b) unique variances—and the conditions for (2) state separability—with (c) between-state loading similarity and (d) between-state intercept difference—were based on a recent simulation study investigating the performance of DT-LMFA (see Vogelsmeier et al., in press). Possible performance differences across the thee methods are expected to be especially visible for a (1) higher model complexity and thus a higher number of factors (a), which

decreases factor overdetermination for a fixed number of items (e.g., Preacher & MacCallum, 2002) and higher unique variances (b), which lowers common variance and therefore also decrease factor overdetermination (e.g., Briggs & MacCallum, 2003) and (2) a lower state separability in terms of loadings (c) and intercept differences (d) (De Roover, Vermunt, Timmerman, & Ceulemans, 2017; Vogelsmeier et al., in press).

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. We expect that possible performance differences are especially visible for a lower amount of information (Crayen et al., 2017). Finally, T_{day} determines the amount of DT violation as well as the transition probabilities, as will be described below. Performance differences are expected to be especially pronounced for lower levels of T_{day}, implying increasing TIs, DT violations, and transition probabilities to other states, where the latter leads to lower dependence of states at two consecutive time-points, making estimation more difficult (Vogelsmeier et al., in press).

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 $2\,(\mathrm{a}) \times 2\,(\mathrm{b}) \times 2\,(\mathrm{c}) \times 2\,(\mathrm{d}) \times 2\,(\mathrm{e}) \times 2\,(\mathrm{f}) \times 3\,(\mathrm{g}) = 192$ conditions.

The number of items *J* was fixed to 20 and the number of states K was fixed to 3. The number of factors *F* was identical across states and either 2 or 4 (a). The unique variances *e* (b) were either .40 or .20.

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 $F = 2$:

$$\mathbf{\Lambda}_{Base} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \quad (7)$$

To induce loading differences between the states, we altered the base matrices for $F = 2$ and $F = 4$ 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

$$\begin{aligned} \mathbf{\Lambda}_1 &= \begin{pmatrix} \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \\ \mathbf{\Lambda}_2 &= \begin{pmatrix} 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \end{aligned} \quad (8)$$

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 (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 $1 - e$.

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:

$$\boldsymbol{\nu}_{Base} = (5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \qquad (9)$$

For the no intercept difference condition, we used $\boldsymbol{\nu}_{Base}$. For the low intercept difference condition, we increased respectively 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.

$$\begin{aligned} \boldsymbol{\nu}_1 &= (\mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \\ \boldsymbol{\nu}_2 &= (5 \quad 5 \quad \mathbf{5.5} \quad \mathbf{5.5} \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5)' \end{aligned} \qquad (10)$$

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. 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 1.18, and therefore, we sample the deviation from $U[-1.8, +1{,}8]$).

Finally, the transition intensities in $\mathbf{Q}$ were fixed across all conditions, subjects, and time. To determine $\mathbf{Q}$, we considered transition probabilities $\mathbf{P}$ 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

$$\mathbf{P} = \begin{pmatrix} .950 & .025 & .025 \\ .025 & .950 & .025 \\ .025 & .025 & .950 \end{pmatrix}, \qquad (11)$$

$\mathbf{Q}$ was derived by taking the matrix logarithm[2]:

$$\mathbf{Q} = \begin{pmatrix} -.05 & .03 & .03 \\ .03 & -.05 & .03 \\ .03 & .03 & -.05 \end{pmatrix}. \quad (12)$$

As a consequence of the design, the switching probabilities across measurement occasions will be larger for $T_{day} = 3$ because of larger intervals δ_{ti} and smaller for $T_{day} = 9$, where intervals are shorter.

3.3. Data Generation

In the open-source program R (R Core Team, 2018) for each subject, we sampled $T_{day} \times D$ time-points as previously described (see 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 (1), assuming orthogonal factors, and concatenated the $\mathbf{Y}_i$'s into one dataset $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_3')'$. Note that, for the DTG-LMFA, we followed the procedure described in 2.2.1. and put the observations on a 1-hour grid. In total, 20 replicates of the 192 conditions and thus 3840 datasets were generated.

3.4. Results

Performances were evaluated based on the results that converged in all three analyses. For the 3840 datasets, this was the case for 71.02 % (with 79.04 % converged analyses in CT-LMFA, 89.82 % in DTG-LMFA, and 99.95 % in DT-LMFA). The differences in convergence will further be described in 3.4.3 [3].

3.4.1. 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 0 (i.e., overlap is at chance) to 1 (i.e., sequences are identical). Second, to obtain the differences in the

[2] Note that the rows do not sum to zero only because of rounding in this representation.
[3] Note that solutions may also end up in a local optimum. More information about local optima is provided in Supplement B.

goodness of loading recovery ($GOSL$), we averaged the Tucker congruence coefficient between the true and the estimated loading matrices across factors and states:

$$GOSL = \frac{\sum_{k=1}^{K}\sum_{f=1}^{F}\varphi(\boldsymbol{\Lambda}_k^f, \widehat{\boldsymbol{\Lambda}}_k^f)}{\sum_{k=1}^{K} F_k}. \quad (13)$$

We used Procrustes rotation (Kiers, 1997)[4] to rotate state-specific loadings $\widehat{\boldsymbol{\Lambda}}_k^f$ to $\boldsymbol{\Lambda}_k^f$. 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[5]. Note that, because of label switching for the transition and initial state parameters, we considered the state permutation as already evaluated as the best for the loading recovery. Furthermore, the transition parameters are probabilities for DT and DTG 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.

3.4.2. Goodness of parameter recovery

As can be seen from the 'average' results in Table 1, the three analyses were very comparable with regard to parameter recovery of all parameters but the transition probabilities, yet with CT-LMFA being slightly superior to DTG- and DT-LMFA regarding the general state-recovery. Furthermore, the transition probabilities were perfectly recovered in CT-LMFA, well in DTG-LMFA, and moderate in DT-LMFA. Moreover, contrary to our expectations, the effects of the manipulated conditions on the present performance differences—thus regarding state recovery (ARI) and transition probability recovery (MAD_{trans})—were rather small or absent. Only the number of measurements per day, T_{day}, had a

[4] We conducted the rotation in R, since rotation is not yet possible in LG.

[5] 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, which is usually diagnostic for underdetermined factors or insufficient sample size (e.g., Van Driel, 1978). Heywood cases did not occur in any of the analyses and are therefore not further discussed.

considerable impact in that (1) the performance differences between CT- and DTG-LMFA were only visible for the lowest manipulated factor, $T_{day} = 3$, and (2) the performance difference between CT- and DT-LMFA increased for lower levels of T_{day}, which was in line with our expectation.

[Insert Table 1 about here]

3.4.3. Encountered estimation problems

As already indicated above, while almost all analyses converged in DT-LMFA[6], 20.95 % of the replication in CT-LMFA and 10.18 % in DTG-LMFA exhibited estimation problems, which we did not anticipate in advance. Closer investigation of the non-convergence problems in CT-LMFA revealed these are caused by the way LG computes the analytic first derivatives of the transition probabilities $p_{\delta_{ti}}(\mathbf{s}_t|\mathbf{s}_{t-1})$ towards the intensities q_{lk} needed in the M-step of the EM algorithm (see Supplement A.2.3 and Kalbfleisch & Lawless, 1985). More specifically, valid derivatives are available only if all eigenvalues of $\mathbf{Q}$ are different, which may not always be the case (Kalbfleisch & Lawless, 1985). As illustrated in Table 1, manipulated factors hardly affected convergence.

In contrast to CT-LMFA, investigation of the estimation problems in DTG-LMFA revealed that they were caused by reaching the maximum number of EM iterations without convergence (despite the high number of 10,000 iterations). As can be seen in Table 1, convergence was remarkably influenced by the number of measurement occasions per day, T_{day}, in that the proportion of estimation problems was especially high for the lowest level (i.e., $T_{day} = 3$), while the analyses perfectly converged for higher levels (i.e., for $T_{day} = 6 \text{ and } T_{day} = 9$). The problem here is that fewer measurement occasions per day increase the amount of rows with missing values in the dataset[7],

[6] Only two analyses did not converge for DT-LMFA because the maximum number of EM iterations was reached without convergence, which can be resolved by running the models again with different start values.

[7] To clarify this, consider the difference between $T_{day} = 3$ and $T_{day} = 9$. For the former, there are only 3 observations on the 1-hour grid for 24 hours and thus $24 - 3 = 21$ empty rows, while for the latter there are only $24 - 9 = 15$ empty rows.

which hampers convergence. Re-estimating the non-converged models with new starting values or increasing the number of iterations likely helps. However, it should be noted that also the computation time is influenced. For instance, with an average of almost 10 minutes, estimation took about three times longer for $T_{day} = 3$ than for $T_{day} = 6$)[8]. Although this estimation time is perfectly feasible, DTG-LMFA can become infeasible for datasets with highly unequal time intervals and very fine grids, which lead to very large numbers of empty rows with missing values only. An example of this will be described in the real-data application in Section 4.

3.5. Conclusion and Recommendations

To sum up, there was a striking similarity in recovering parameters under a wide range of conditions across the three analyses CT-, DTG- and DT-LMFA which generally demonstrates the robustness of DT-LMFA against unequal TIs and the good performance of DTG-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. However, CT-LMFA exhibited the most estimation problems. In cases of non-convergence, researchers are advised to use DTG- or DT-LMFA because they are relatively adequate substitutes. Although transition probabilities are better captured by DTG-LMFA, DT-LMFA should generally be preferred if the grid in DTG-LMFA would contain many missing entries between observations. That is because the estimation would require many iterations and hence, be very time-intensive. This could be especially inconvenient when exploring many different models in terms of the number of states and factors[9]. Moreover, cumbersome data-preparation is not needed in DT-LMFA.

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)

[8] Just to give a reference, the conditions with $T_{day} = 3$ took only about 1 minute in CT- and DT-LMFA.
[9] Note that, in contrast to the simulation study, where the right number of states and factors was given, a model selection procedure is required when working with real data. The Bayesian information criterion proved to be a valid criterion (Vogelsmeier et al., in press).

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, Rush, Shaw, & Emery, 1979; items listed in Table 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 4-week questionnaire were completed on two consecutive days) to very large (e.g., a year when certain follow ups were skipped)[10].

To begin with the data-analysis, model selection with the Bayesian information criterion (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 3 and 2 factors[11]. In order to shed light on the state-specific MMs, we investigated the standardized oblimin rotated loadings (Table 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 e.g., 'mood', 'pessimism', and 'lack of satisfaction'—, (2) 'self-image'—with strong loadings of e.g., 'guilty feeling', 'self-hate', and 'self accusation'—, and (3) 'cognition/body'—with strong loadings of e.g., '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.

[Insert Table 2 and Figure 2 about here]

[10] Note that some additional information about choices made regarding the data is provided in Supplement D.
[11] The syntax for the final model can be found in Supplement C.

Next, we look at the estimated transition intensity matrix $\mathbf{Q} = \begin{pmatrix} -.2 & .2 \\ .2 & -.2 \end{pmatrix}$, from which we can calculate $\mathbf{P}$ for any interval of interest, e.g., for one week $\mathbf{P}_{week} = \begin{pmatrix} .89 & .11 \\ .08 & .92 \end{pmatrix}$ and a year $\mathbf{P}_{year} = \begin{pmatrix} .43 & .57 \\ .43 & .57 \end{pmatrix}$, showing how transitions become more likely over time. Looking at the estimated initial state probabilities $\boldsymbol{\pi} = (.9 \quad .1)$ and Figure 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 with 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 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.

5. Discussion

In this paper, 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 and the CT approximation with discrete-time grid (DTG)-LMFA. Although the recovery of states was only slightly superior in CT-LMFA, we demonstrated why the method should be favored whenever possible (i.e., when the estimation procedure converges). CT-LMFA has a natural match with the assumption that processes evolve at irregular TIs and transition intensities can be transformed to any DT transition

probabilities and therefore allows to compare transition probabilities within and across studies, leading to more freedom in interpreting time-intensive longitudinal data.

Although the CT approach avoids invalid results through unequal TIs, they might still be induced by assuming constant transition intensities over time and thus constant transition probabilities for identical TIs. For example, it could be restrictive to assume that transition intensities are the same for night and day intervals. However, conditioning transition intensities on time-constant or time-varying categorical covariates likely increases estimation problems because transition intensity matrices have to be estimated per level of the covariate and hence the likelihood of obtaining invalid first derivatives using eigen-decompositions increases. Therefore, future research should look at the performance of alternative ways of computing these derivatives, e.g., numerically.

Finally, the simulation study employed a typical ESM design, where the degree of DT violation is generally rather small. For this design, the comparison of all three methods was possible. In contrast, as our simulation study results have shown, computation time would make a simulation study with data such as the application data difficult because the grid for DTG-LMFA would contain too many missing entries (e.g. 365 if a year lies between two consecutive observations). However, especially DT-LMFA is likely affected by the size of DT violations (while, theoretically, a finer grid in DTG-LMFA may compensate for larger violations) and therefore, we cannot generalize our findings purporting that DT-LMFA is an adequate substitute for datasets with large DT violations. This should therefore be investigated in future research.

6. References

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

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.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York, NY: Guilford Press.

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

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

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. *Psychometrica, 53*, 251–259. doi:10.1007/BF02294136

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

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 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

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.

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*. doi:doi:10.18637/jss.v077.i05

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.

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

Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (2014). *Multivariate Data Analysis: Pearson International Edition*. Edinburgh, UK: Pearson.

Jackson, C. H., & Sharples, L. D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Stat Med, 21*, 113–128.

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.

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

Lawley, D. N., & Maxwell, A. E. (1962). *Factor analysis as a statistical method*. London, UK: Butterworth.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi:10.1007/BF02294825

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

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

R Core Team. (2018). A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Scollon, C., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5–34.

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 Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika, 43*, 225–243. doi:10.1007/BF02293865

Vermunt, J. K., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In *Handbook of Longitudinal Research: Design, Measurement, and Analysis* (pp. 373–385). Burlington, MA: Elsevier.

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., Vermunt, J. K., van Roekel, E., & De Roover, K. (in press). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. *Structural Equation Modeling: A Multidisciplinary Journal*.

Figures and Tables

Table 1. Goodness-of-recovery per type of parameter and convergence conditional on the manipulated factors.

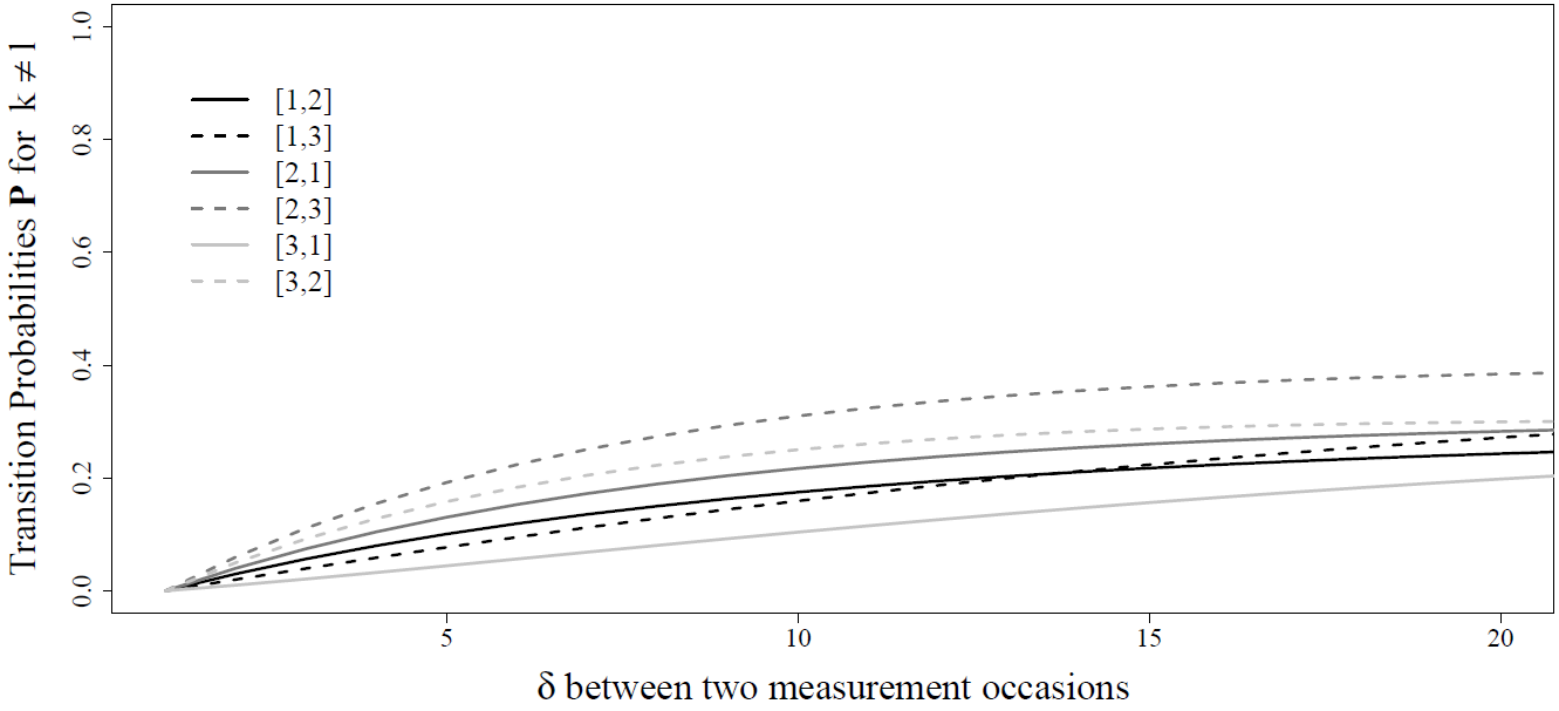
		Goodness of																		
		State Recovery (ARI)			Loading Recovery ($GOSL$)			Transition Parameter Recovery (MAD_{trans})			Intercept Recovery (MAD_{int})			Unique Variance Recovery (MAD_{uniq})			Initial State Recovery ($MAD_{initial}$)			Convergence		
		Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA		
Condition	Factors	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT
Factors per	2	.86	.85	.85	1	1	1	0	.02	.06	.03	.03	.03	.01	.01	.01	.06	.07	.06	.79	.89	1
State F^k	4	.86	.85	.84	1	1	1	0	.02	.06	.02	.02	.02	.01	.01	.01	.06	.07	.06	.79	.91	1
Between-State Loading Difference	medi-um	.84	.83	.83	1	1	1	0	.02	.06	.02	.03	.03	.01	.01	.01	.06	.07	.06	.81	.89	1
	low	.87	.86	.86	1	1	1	0	.02	.06	.02	.02	.02	.01	.01	.01	.06	.07	.06	.77	.91	1
Between-State Intercept difference	no	.81	.80	.79	1	1	1	0	.02	.06	.02	.03	.03	.01	.01	.01	.06	.08	.06	.77	.89	1
	low	.90	.90	.89	1	1	1	0	.02	.06	.02	.03	.02	.01	.01	.01	.06	.07	.06	.81	.91	1
Unique Variance e	.2	.91	.90	.90	1	1	1	0	.02	.06	.02	.02	.02	.01	.01	.01	.06	.07	.06	.80	.92	1
	.4	.80	.79	.79	1	1	1	0	.02	.06	.03	.03	.03	.02	.02	.02	.06	.08	.06	.78	.88	1
Sample Size N	35	.85	.85	.84	1	1	1	.01	.02	.06	.03	.03	.03	.02	.02	.02	.07	.09	.07	.79	.91	1
	75	.86	.85	.85	1	1	1	0	.02	.06	.02	.02	.02	.01	.01	.01	.05	.06	.05	.79	.89	1
Number of Participation Days D	7	.85	.84	.84	1	1	1	.01	.03	.06	.03	.03	.03	.02	.02	.02	.06	.08	.06	.78	.93	1
	30	.87	.86	.85	1	1	1	0	.02	.06	.02	.02	.02	.01	.01	.01	.06	.07	.06	.80	.87	1
Measurements per day T_{day}	3	**.75**	**.72**	**.74**	1	1	1	**.01**	**.08**	**.10**	.04	.04	.04	.02	.02	.02	.06	.08	.06	.78	**.70**	1
	6	**.87**	**.87**	**.85**	1	1	1	**0**	**0**	**.05**	.02	.02	.02	.01	.01	.01	.06	.07	.06	.78	**1**	1
	9	**.92**	**.92**	**.90**	1	1	1	**0**	**0**	**.03**	.02	.02	.02	.01	.01	.01	.06	.07	.06	.80	**1**	1
		Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA			Type LMFA		
All Conditions		CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT	CT	DTG	DT
Average		**.86**	**.85**	**.84**	1	1	1	**0**	**.02**	**.06**	.02	.03	.02	.01	.01	.01	.06	.07	.06	.79	.90	1
SD		.11	.13	.11	0	0	0	0	.08	.03	.01	.01	.01	.01	.01	.01	.03	.04	.03			

Note. LMFA = latent Markov factor analysis; CT = continuous-time; DTG = discrete-time-grid; DT = discrete-time. Most notable analysis effects on the performance and condition effects on the performance differences are in boldface. The perfect loading recoveries are a consequence of the highly similar loading matrices across the states that have been mixed up.

Table 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	**0.44**	0.22	**0.32**	1.20	0.30	**0.73**	-0.03	0.40	0.16
Pessimism	**0.56**	0.29	0.08	1.32	0.31	**0.77**	-0.17	0.38	0.15
Sense of Failure	**0.45**	**0.52**	-0.03	1.29	0.27	**0.79**	-0.25	0.34	0.13
Lack of Satisfaction	**0.55**	-0.09	**0.42**	1.38	0.28	**0.70**	0.29	0.55	0.14
Guilty Feeling	0.13	**0.62**	-0.02	1.20	0.49	**0.49**	-0.1	0.23	0.18
Sense of Punishment	-0.06	**0.43**	-0.01	0.98	0.93	0.00	0.00	0.00	0.00
Self-Hate	0.20	**0.60**	0.17	1.39	0.25	**0.77**	-0.15	0.52	0.16
Self Accusations	0.20	**0.70**	-0.08	1.30	0.25	**0.80**	-0.28	0.50	0.14
Self Punitive Wishes	**0.41**	0.22	0.05	0.65	0.40	0.00	0.00	0.00	0.00
Crying Spells	0.01	0.23	**0.45**	0.97	0.68	**0.43**	0.10	0.19	0.13
Irritability	0.02	0.14	**0.57**	0.97	0.31	**0.55**	0.24	0.53	0.21
Social Withdrawal	0.23	0.04	**0.62**	1.14	0.30	**0.66**	0.25	0.46	0.20
Indecisiveness	0.18	0.18	**0.48**	1.25	0.43	**0.64**	**0.34**	0.50	0.18
Body Image	-0.10	**0.57**	0.18	1.23	0.59	**0.62**	-0.18	0.53	0.42
Work Inhibition	**0.37**	0.14	0.28	1.29	0.31	**0.64**	**0.33**	0.49	0.15
Sleep Disturbance	0.05	-0.07	**0.58**	1.26	0.63	**0.42**	0.24	0.51	0.33
Fatigability	0.26	-0.01	**0.59**	1.32	0.32	**0.63**	0.29	0.61	0.18
Loss of Appetite	-0.01	-0.10	**0.45**	0.75	0.54	0.00	0.00	0.00	0.00
Somatic Preoccupation	-0.15	0.10	**0.40**	0.56	0.38	**0.42**	-0.04	0.29	0.22
Loss of Libido	0.04	0.00	**0.55**	1.08	0.75	**0.44**	0.20	0.49	0.43

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.



Intensities in **Q**

	[,1]	[,2]	[,3]
[1,]	−0.05	0.03	0.02
[2,]	0.04	−0.11	0.06
[3,]	0.01	0.05	−0.06

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 = 20

	[,1]	[,2]	[,3]
[1,]	0.47	0.25	0.28
[2,]	0.29	0.33	0.39
[3,]	0.21	0.3	0.49

Figure 1. Increasing probabilities to switch to another state (state $k \neq l$) for larger time intervals (Tis) δ and three example probability matrices calculated based on **Q** and TI δ. Note that [x,y] indicates the element in the matrices with x referring to the rows and y to the columns and that two rows do not sum to zero/one only because of rounding.

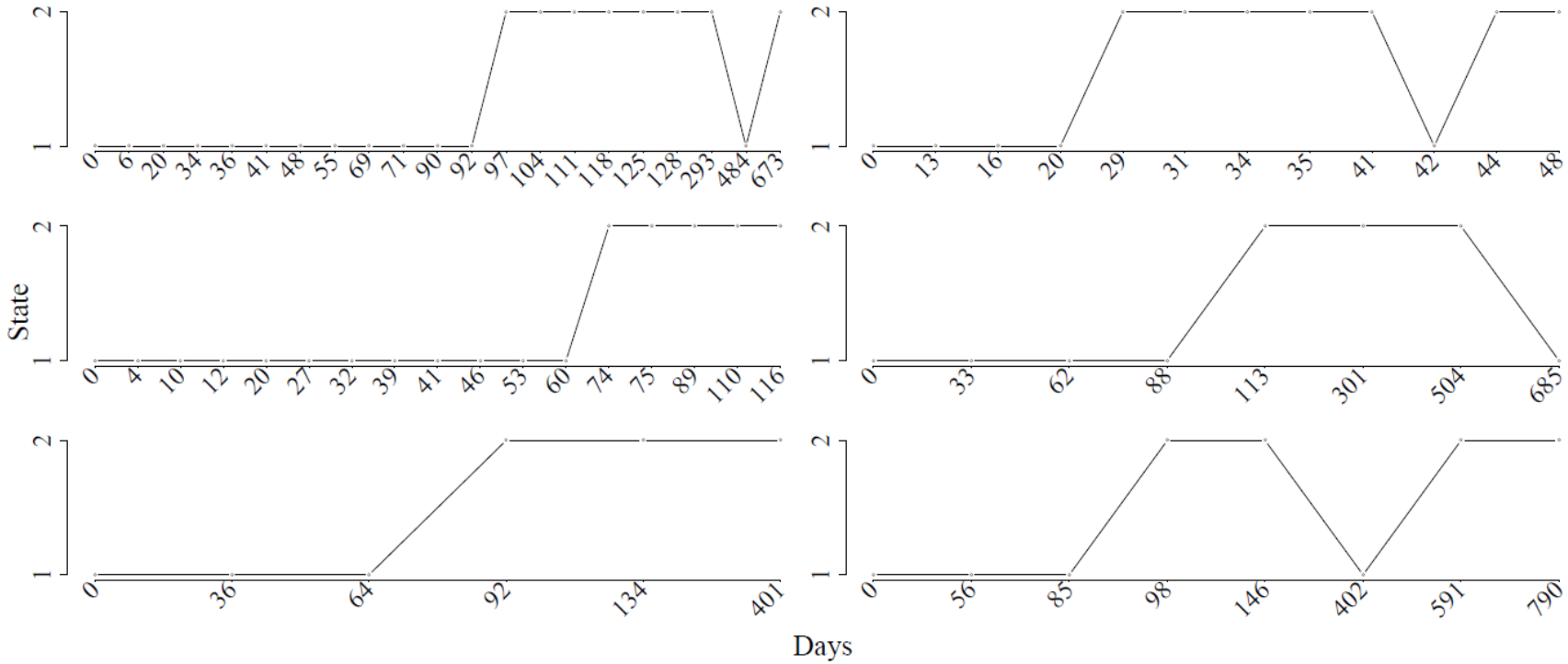


Figure 2. Six representative examples of individual transition plots. Note that the scale of the spacing of the x-axis is not in line with the amount of days elapsed but, to enable the illustration, equal spaces are chosen.

Supplement A

In brief, the Expectation Maximization algorithm (EM algorithm; Dempster, Laird, & Rubin, 1977) employs the so-called complete-data loglikelihood ($\log L_c$), which means that the state assignments of all time-points are assumed to be known. In the Expectation step (E-step; Section 1.A.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, the parameters are determined based on a set of initial values and, for every other iteration, the parameters are determined by estimates from the previous iteration $\widehat{\boldsymbol{\theta}}^{old}$. 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, Petrie, Soules, & Weiss, 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; Section A.2), the parameters $\widehat{\boldsymbol{\theta}}$ get updated so that they maximize $E(\log L_c)$. This procedure is repeated until convergence (Section 1.A.3).

As mentioned in Section 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, and De Roover (in press). However, we describe the M-step to update the transition intensities in more detail (Section A.2.3) because this is the part where CT-LMFA differs from DT-LMFA.

A.1. E-step

The $E(\log L_c)$ is given by

$$E(\log L_c) = \sum_{i=1}^{I}\sum_{k=1}^{K}\gamma(s_{i1k})\log(\pi_k) + \sum_{i=1}^{I}\sum_{t=2}^{T}\sum_{l=1}^{K}\sum_{k=1}^{K}\varepsilon\left(s_{it-1,l}, s_{itk}\right)\log\left(e^{q_{ilk}\delta_{ti}}\right) - \frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{k=1}^{K}\gamma(s_{itk})[J\log(2\pi) + \log(|\boldsymbol{\Sigma}_k|) + (\mathbf{y}_{it} - \boldsymbol{\nu}_k)\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_{it} - \boldsymbol{\nu}_k)'] . \quad \text{(A1)}$$

Here, δ_{ti} refers to the time interval between time-point t and $t-1$ and for subject i. Furthermore, $\gamma(s_{itk})$ are the expected values to belong to each of the states and $\varepsilon(s_{it-1,l}, s_{itk})$ 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

$$\alpha(s_{i1k}) = \pi_k p(\mathbf{y}_{i1}|s_{i1k}) \tag{A2}$$

and for all for all the remaining time-points with

$$\alpha(s_{itk}) = p(\mathbf{y}_{it}|s_{itk}) \sum_{l=1}^{K} \alpha(s_{it-1,l}) e^{q_{ilk}\delta_{ti}}. \tag{A3}$$

The backward probabilities for time-point $t=T$ are computed with

$$\beta(s_{iTk}) = p(\emptyset|s_{iTk}) = 1, \tag{A4}$$

where ∅ refers to 'producing no outcome'. For all the remaining time-points the backward probabilities are computed with

$$\beta(s_{itk}) = \sum_{l=1}^{K} \beta(s_{it+1,l}) p(\mathbf{y}_{it+1}|s_{it+1,l})\, e^{q_{ilk}\delta_{ti}}. \tag{A5}$$

Finally, the expected values to belong to each of the states are calculated with

$$\gamma(s_{itk}) = p(s_{itk}|\,\mathbf{Y}_i) = \frac{\alpha(s_{itk})\,\beta(s_{itk})}{\sum_{k=1}^{K}\alpha(s_{iTk})} \tag{A6}$$

and the expected values to make transitions between the states with

$$\varepsilon(s_{it-1,l}, s_{itk}) = p(s_{it-1,l}, s_{itk}|\mathbf{Y}_i) = \frac{\alpha(s_{it-1,l})\, p(\mathbf{y}_{it}|s_{itk})\, e^{q_{ilk}\delta_{ti}}\, \beta(s_{itk})}{\sum_{k=1}^{K}\alpha(s_{iTk})}. \tag{A7}$$

A.2. M-step

In the M-step, the parameters get updated so that they maximize ($\log L_c$).

A.2.1. Update initial state probabilities and intercepts

The initial state probabilities and intercepts are updated as follows:

$$\pi_k^{new} = \frac{\sum_{i=1}^{I} \gamma(s_{i1k})}{\sum_{k=1}^{K} \sum_{i=1}^{I} \gamma(s_{i1k})}, \tag{A8}$$

$$\boldsymbol{\nu}_k^{new} = \frac{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk}) \mathbf{y}_{it}}{\sum_{i=1}^{I} \sum_{t=1}^{T} \gamma(s_{itk})}. \tag{A9}$$

A.2.2. Update state-specific covariance matrices

In order to find the maximum likelihood estimates for updating the state-specific covariance matrices $\sum_k^{new} = \boldsymbol{\Lambda}_k^{new} \boldsymbol{\Lambda}_k^{new} + \mathbf{D}_k^{new}$, the observations are weighted by the corresponding $\gamma(s_{itk})$-values and these K weigthed datasets $\mathbf{Y}_k$ are in turn factor analyzed by means of Fisher scoring (Lee & Jennrich, 1979).

A.2.3. 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 $\boldsymbol{\theta}^{\mathbf{Q}}$. In the following, we describe the algorithm and illustrate some elements by means of a $K = 3$ states example with the following 3×3 matrix $\mathbf{Q}$:

$$\mathbf{Q} = \begin{pmatrix} -(q_{12} + q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21} + q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31} + q_{32}) \end{pmatrix}, \tag{A10}$$

where the parameters are thus $\boldsymbol{\theta}^{\mathbf{Q}} = (q_{12}, q_{13}, q_{21}, q_{23}, q_{31}, q_{32})$. 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. However, for simplicity of this illustration, we will continue without this re-parameterization.

1.2.3.1. Partial derivatives

We first calculate the partial derivatives of $\mathbf{P}(\delta_{ti})$ with respect to the parameters $\theta_1^{\mathbf{Q}}$ to $\theta_b^{\mathbf{Q}}$ in $\boldsymbol{\theta}^{\mathbf{Q}}$:

$$\frac{\partial \mathbf{P}(\delta_{ti})}{\partial \theta_u^{\mathbf{Q}}} = \mathbf{A}\mathbf{V}_u(\delta_{ti})\mathbf{A}^{-1}, \tag{A11}$$

for $u = 1, \ldots, b$. The $k \times k$ matrix $\mathbf{A}$ contains the three eigenvectors of $\mathbf{Q}$ and $\mathbf{V}_u(\delta_{ti})$ is the time and subject-specific $k \times k$ matrix $\mathbf{V}_u(\delta_{ti})$ with entries

$$\begin{gathered} {g_{lk}}^{(u)} \delta_{ti} e^{d_l \delta_{ti}} \text{ for } k \neq l \text{ and} \\ \frac{{g_{lk}}^{(u)} (e^{d_l \delta_{ti}} - e^{d_k \delta_{ti}})}{(d_l - d_k)} \text{ for } k = l, \end{gathered} \tag{A12}$$

where $d_l, d_k = d_1, \ldots, d_K$ are the eigenvalues of $\mathbf{Q}$ and ${g_{lk}}^{(u)}$ are the entries in

$$\mathbf{G}^{(u)} = \mathbf{A}^{-1} \frac{\partial \mathbf{Q}}{\partial \theta_u^{\mathbf{Q}}} \mathbf{A}. \tag{A13}$$

Hence, in order to calculate the partial derivatives $\partial \mathbf{P}(\delta_{ti}) / \partial \theta_u^{\mathbf{Q}}$, we first have to calculate the partial derivatives of $\mathbf{Q}$ with respect to the parameters $\theta_1^{\mathbf{Q}}, \ldots, \theta_b^{\mathbf{Q}}$, then calculate the $\mathbf{G}^{(u)}$ matrices for each parameter $\theta_1^{\mathbf{Q}}$ and finally compute the $\mathbf{V}_u(\delta_{ti})$ matrices. For the 3-state example, there are the following six $\mathbf{G}^{(u)}$ matrices:

$$\mathbf{G}^{(1)} = \mathbf{A}^{-1} \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{A}, \qquad \mathbf{G}^{(2)} = \mathbf{A}^{-1} \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{A}, \qquad \mathbf{G}^{(3)} = \mathbf{A}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{A},$$

$$\mathbf{G}^{(4)} = \mathbf{A}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{A}, \qquad \mathbf{G}^{(5)} = \mathbf{A}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \mathbf{A}, \qquad \mathbf{G}^{(6)} = \mathbf{A}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \mathbf{A}.$$

and the general $\mathbf{V}_u(\delta_{ti})$ matrix looks as follows

$$\mathbf{V}_u(\delta_{ti}) = \begin{pmatrix} {g_{11}}^{(u)} \delta_{ti} e^{d_1 \delta_{ti}} & \frac{{g_{12}}^{(u)} (e^{d_1 \delta_{ti}} - e^{d_2 \delta_{ti}})}{(d_1 - d_2)} & \frac{{g_{13}}^{(u)} (e^{d_1 \delta_{ti}} - e^{d_3 \delta_{ti}})}{(d_1 - d_3)} \\ \frac{{g_{21}}^{(u)} (e^{d_2 \delta_{ti}} - e^{d_1 \delta_{ti}})}{(d_2 - d_1)} & {g_{22}}^{(u)} \delta_{ti} e^{d_2 \delta_{ti}} & \frac{{g_{23}}^{(u)} (e^{d_2 \delta_t} - e^{d_3 \delta_{ti}})}{(d_2 - d_3)} \\ \frac{{g_{31}}^{(u)} (e^{d_3 \delta_{ti}} - e^{d_1 \delta_{ti}})}{(d_3 - d_1)} & \frac{{g_{32}}^{(u)} (e^{d_3 \delta_{ti}} - e^{d_2 \delta_{ti}})}{(d_3 - d_2)} & {g_{22}}^{(u)} \delta_{ti} e^{d_3 \delta_{ti}} \end{pmatrix}. \tag{A14}$$

1.2.3.2. Scoring procedure

Once we have obtained the partial derivatives of $\mathbf{P}(\delta_{ti})$, 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

$$S(\theta_u^{\mathbf{Q}}) = \frac{\partial logL}{\partial \theta_u^{\mathbf{Q}}} = \sum_{i=1}^{I}\sum_{t=2}^{T}\sum_{k,l=1}^{K} \frac{\varepsilon(s_{it-1,l}, s_{itk})}{p_{lk}(\delta_{ti})} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^{\mathbf{Q}}}. \quad \text{(A15)}$$

Here, $\varepsilon(s_{it-1,l}, s_{itk})$ are the expected state-membership probabilities obtained from the E-step (Equation (A7)). Next, we calculate the $b \times b$ matrix $\mathbf{M}(\boldsymbol{\theta}^{\mathbf{Q}})$ with entries

$$M(\theta_u^{\mathbf{Q}}, \theta_v^{\mathbf{Q}}) = \sum_{i=1}^{I}\sum_{t=2}^{T}\sum_{k,l=1}^{K} \frac{\gamma(s_{it-1,k})}{p_{lk}(\delta_{ti})} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^{\mathbf{Q}}} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_v^{\mathbf{Q}}}. \quad \text{(A16)}$$

Finally, we put all the elements together to compute the update $\boldsymbol{\theta}_{new}^{\mathbf{Q}}$:

$$\boldsymbol{\theta}_{new}^{\mathbf{Q}} = \boldsymbol{\theta}_{old}^{\mathbf{Q}} + \mathbf{M}(\boldsymbol{\theta}_{old}^{\mathbf{Q}})^{-1}\mathbf{S}(\boldsymbol{\theta}_{old}^{\mathbf{Q}}), \quad \text{(A17)}$$

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 Section A.3.

A.3. Convergence

Convergence is 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, i.e., $\omega = \sum_{r=1}^{R} \left| \frac{\hat{\theta}_r^{new} - \hat{\theta}_r^{old}}{\hat{\theta}_r^{old}} \right|$, with $r = 1, \ldots, R$ 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.

A.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 a global maximum. The start sets generally consist of random start values but, for loadings and residual varinces, they are based on PCA (Jolliffe, 1986) 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, Vermunt, Timmerman, and Ceulemans (2017) and Vermunt and Magidson (2016).

References Supplement A

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.

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

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.

Jolliffe, I. T. (1986). *Principal Component Analysis*. NewYork, NY: Springer.

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.

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

Vermunt, J. K., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations.

Vogelsmeier, L. V. D. E., Vermunt, J. K., van Roekel, E., & De Roover, K. (in press). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. *Structural Equation Modeling: A Multidisciplinary Journal*.

Supplement B

It may 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 an approximation ('proxi') 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. Comparing the ML solution of the simulated datasets that converged in all analyses, a local maximum was found for only 0.27 % of the datasets analyzed with CT-LMFA and for 0.23 % of the datasets analyzed with DT-LMFA. However, the percentage was higher for datasets analyzed with DTG-LMFA and amounted to 6.51 %. For CT- and DT-LMFA, there was no considerable influence of the manipulated factors while for DTG-LMFA, the local maxima especially occurred for the lowest level of the number of measurement occasions per day, $T_{day} = 3$ and hence the level with the most empty rows in a dataset (which was also the condition level with the most convergence problems in DTG-LMFA). As briefly outlined in Supplement A.4, the multistart procedure reduces the probability of retaining local ML solutions, where more random starts are better. Therefore, when many empty rows are present in the dataset, it is advised to increasing the number of 25 random starts (which was used in the simulation study) to 50 or 100.

Supplement C

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
```

Supplement D

In the following, we provide some additional information about the treatment and the Becks Depression Inventory (BDI; Beck, Rush, Shaw, & Emery, 1979) used in the presented application (Section 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 FA. Since desired weight loss is not part of depression, we deemed it important to remove the item from our analyses.

References Supplement D

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York, NY: Guilford Press.

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.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal Psychtherapy of Depression*. New York, NY: Basic Books Inc Publishers.