

Continuous-Time Latent Markov Factor Analysis for Exploring Measurement Model Changes Across Time

Vogelsmeier, Leonie V.D.E.^a, Vermunt, Jeroen K.^a, Böing-Messing, Florian^{b,a}, De Roover, Kim^a

^aTilburg University, The Netherlands

^bJheronimus Academy of Data Science, The Netherlands

Vogelsmeier, Leonie V.D.E. (l.v.d.e.vogelsmeier@tilburguniversity.edu), 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@tilburguniversity.edu

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, for instance, 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 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 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, Voelke, 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 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 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, the differences between CT- and DT-LMFA, how the DT approach may approximate CT, and the general model estimation. Section 3 presents a simulation study comparing the performance of CT- and DT-LMFA. Section 4 illustrated CT-LMFA with an application. Section 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.

2. Method

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, ..., 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) and then present CT-LMFA (2.2.2).

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 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 $\sum_{k=1}^K p_{lk} = 1$. 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 a factor analysis (FA; Lawley & Maxwell, 1962) model, which defines the state-specific MMs. The state-specific factor model is

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

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 \mathbf{v}_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 FA (EFA) approach, thus without a priori constraints on the factor loadings. If desired, however, confirmatory FA (CFA) could also be used by imposing zero loadings.

From Equation (1) it becomes apparent that the state-specific MMs can differ regarding their loadings Λ_k , intercepts \mathbf{v}_k , unique variances \mathbf{D}_k , and factor covariances Ψ_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 invariance (equal unique variances).

To identify the model, factor variances in Ψ_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 $\Sigma_k = \Lambda_k \Lambda_k' + \mathbf{D}_k$ 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 $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T)$ 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, 1984) 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 2.3 is required to achieve convergence, causing long computation times (for more information on this see Supplement A). Therefore, we only consider the CT-approach, which is a much more natural alternative to account for the unequal TI.

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:

$$q_{lk} = \lim_{\delta \rightarrow 0} \frac{p(s_{tk} = 1 | s_{t-\delta, l} = 1)}{\delta}. \quad (3)$$

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. 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 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 time¹, 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 exponential² of the intensity matrix \mathbf{Q} times the TI δ (Cox & Miller, 1965):

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

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 (5), we can compute the transition probabilities for arbitrary TIs, which is, as mentioned in the introduction, 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 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 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 paper.

² The matrix exponential $e^{\mathbf{A}}$, where \mathbf{A} can be any square matrix, is equal to $\sum_{a=0}^{\infty} \frac{\mathbf{A}^a}{a!} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}\mathbf{A}}{2!} + \frac{\mathbf{A}\mathbf{A}\mathbf{A}}{3!} + \dots$, 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:

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

which is complicated by the latent states. Therefore, to find the maximum likelihood (ML) solution, 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. (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 the appendix (Section A). 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 et al., 2019).

3. Simulation study

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 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 Section A.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 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 3.2. 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 estimation more difficult (Vogelsmeier et al., 2019).

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, Λ_{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 Λ_{Base} for the example of $F = 2$:

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

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 overdetermination 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

$$\Lambda_1 = \begin{pmatrix} \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}' \quad (8)$$

$$\Lambda_2 = \begin{pmatrix} 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}'$$

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 Λ_1 and Λ_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 $1 - e$, 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:

$$\mathbf{v}_{Base} = (5 \ 5)' \quad (9)$$

For the no intercept difference condition, we used \mathbf{v}_{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.

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

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 \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

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

\mathbf{Q} was derived by taking the matrix logarithm³:

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

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, 2018) for each subject, we sampled $T_{day} \times D$ time-points as previously described (see Section 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 $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \dots, \mathbf{Y}_N')$. In total, 20 replicates of the 192 conditions and thus 3840 datasets were generated.

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)⁴.

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 0

³ Note that the rows do not sum to zero only because of rounding in this representation.

⁴ 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 an approximation ('proxy') can be obtained by using the population parameters as starting values and comparing the multistart solution to the proxy solution: When $\log L_{multistart} < \log L_{proxy}$ (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.

(i.e., overlap is at chance) to 1 (i.e., perfect overlap). Second, to obtain the differences in the goodness of loading recovery (*GOSL*), we averaged the Tucker congruence coefficient between the true and the estimated loading matrices across factors and states:

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

We used Procrustes rotation (Kiers, 1997)⁵ to rotate state-specific loadings $\hat{\Lambda}_k^f$ to Λ_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 $\varphi(\Lambda_k^f, \hat{\Lambda}_k^f)$. 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 parameters⁶. 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.

3.3.2. Goodness of parameter recovery

As can be seen from the ‘average’ results in Table 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 performance difference between CT- and DT-LMFA, which was in line with our expectations.

⁵ We conducted the rotation in R, since factor rotation was just added to LG after the study was conducted.

⁶ 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.

[Insert Table 1 about here]

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.

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, 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)⁷.

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 state⁸. Hence,

⁷ Note that some additional information about choices made regarding the data is provided in Supplement C.

⁸ The syntax for the final model can be found in Supplement B.

configural invariance is clearly violated. 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, 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.

[Insert Table 2 and Figure 2 about here]

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 $\boldsymbol{\pi} = (.9 \ .1)$ 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 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 such as a replication study.

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. 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-SEM 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.

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. *Psychometrika*, *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. (2019). On the exploratory road to unraveling factor loading non-invariance: A new multigroup rotation approach. *Structural Equation Modeling: A Multidisciplinary Journal*. 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
- 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.
- 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
- 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.
- Jolliffe, I. T. (1986). *Principal Component Analysis*. New York, 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.
- 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.
- 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
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. doi:10.1007/BF02294825
- 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.
- 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.
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*, 37–47. doi:10.1007/BF02294204
- 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.

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. (2019). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. *Structural Equation Modeling: A Multidisciplinary Journal*. doi:10.1080/10705511.2018.1554445

Figures and Tables

Table 1

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

		Goodness of Recovery for											
		States (<i>ARI</i>)		Loadings (<i>GOSL</i>)		Transition Parameters (<i>MAD_{trans}</i>)		Intercepts (<i>MAD_{int}</i>)		Unique Variances (<i>MAD_{uniq}</i>)		Initial States (<i>MAD_{initial}</i>)	
		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

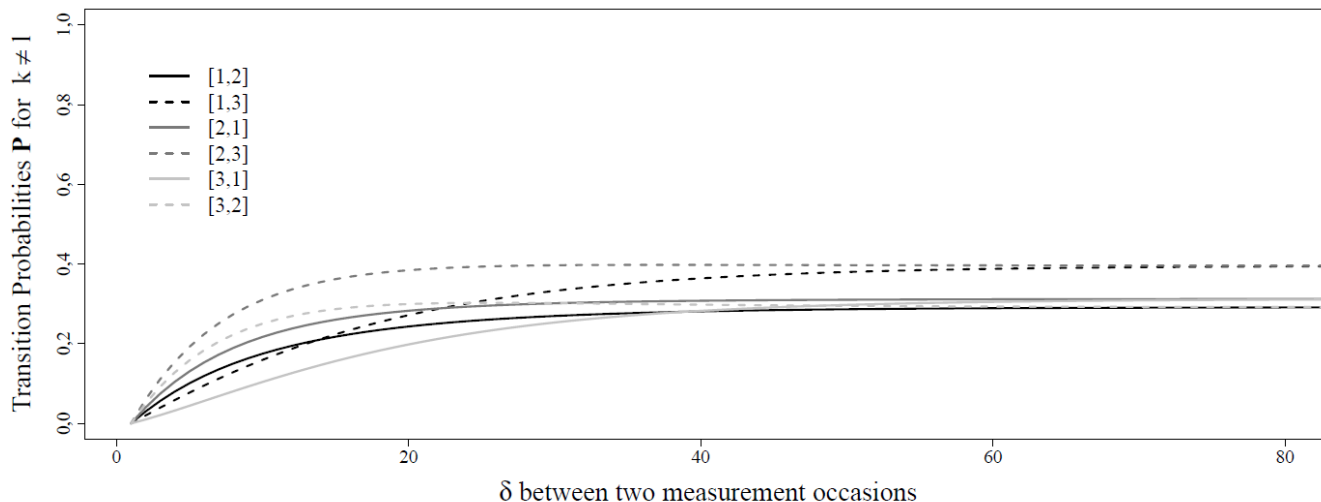
Note. LMFA = latent Markov factor analysis; CT = continuous-time; DT = discrete-time. The perfect loading recoveries are a consequence of the highly similar loading matrices across the states that have been mixed up.

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

Items	State 1			Int.	Unique V.	State 2			
	Factors					Factors	Int.	Unique V.	
	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, $\text{Cor}(\text{Despair}, \text{Self-image}) = .57$, $\text{Cor}(\text{Despair}, \text{Cognition/Body}) = -.28$, and $\text{Cor}(\text{Self-image}, \text{Cognition/Body}) = -.25$; In state 2, $\text{Cor}(\text{Depression}, \text{Cognition/Body}) = -.77$.



Intensities in $Q = \log_m(P \text{ for } \delta = 1)$

	[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 $\delta = 1$

	[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 $\delta = 10$

	[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 $\delta = 80$

	[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 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 $\delta = 1$.

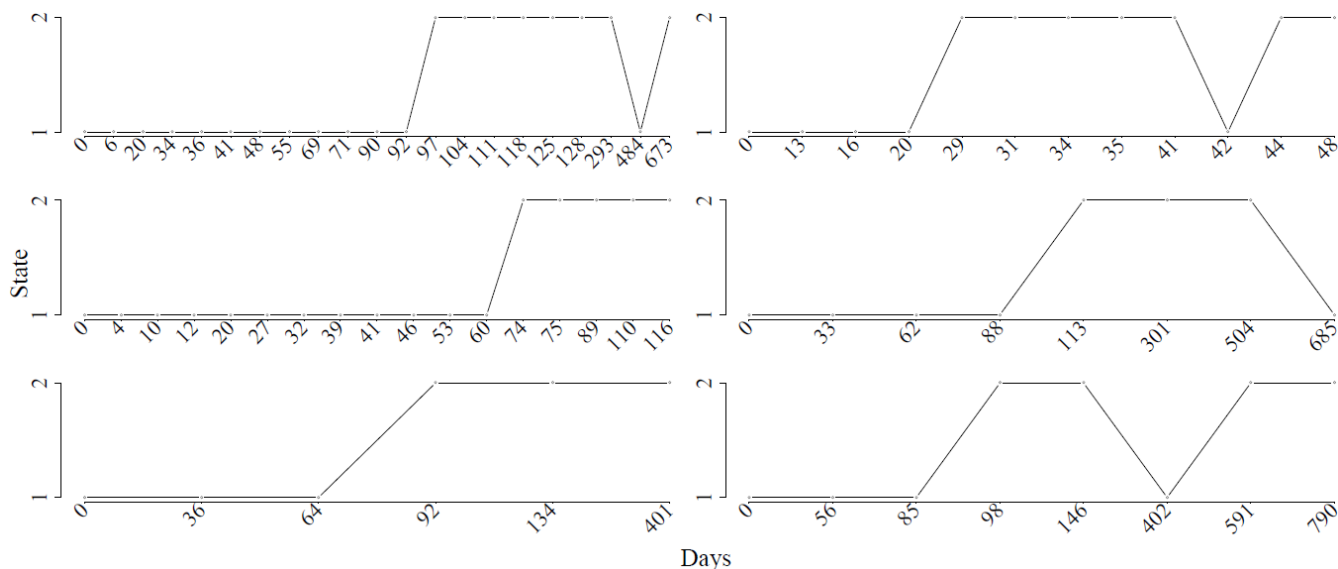


Figure 2. Six representative examples of individual transition plots. Note that the scale of the spacing of the x-axis is not in line with the amount of days elapsed but, to enable the illustration, equal spaces are chosen.

Appendix A

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; Section A.1), the parameters of interest, $\hat{\theta}$, (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 $\hat{\theta}^{old}$ 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; Section A.2), the parameters $\hat{\theta}$ get updated so that they maximize $E(\log L_c)$. This procedure is repeated until convergence (Section 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 et al. (2019). 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

$$\begin{aligned}
E(\log L_c) = & \sum_{i=1}^I \sum_{k=1}^K \gamma(s_{i1k}) \log(\pi_k) + \sum_{i=1}^I \sum_{t=2}^T \sum_{l=1}^K \sum_{k=1}^K \varepsilon(s_{it-1,l}, s_{itk}) \log(e^{q_{ilk}\delta_{ti}}) \\
& - \frac{1}{2} \sum_{i=1}^I \sum_{t=1}^T \sum_{k=1}^K \gamma(s_{itk}) [J \log(2\pi) + \log(|\Sigma_k|) + (\mathbf{y}_{it} - \mathbf{v}_k) \Sigma_k^{-1} (\mathbf{y}_{it} - \mathbf{v}_k)'].
\end{aligned} \tag{A1}$$

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 $\varepsilon(s_{it-1,l}, s_{itk})$ 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 \emptyset refers to ‘producing no outcome’. For all the remaining time-points the backward probabilities are computed with

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

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

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

and the expected bivariate values to make transitions between the states with

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

Note that, upon convergence (see Section A.3), observations are assigned to the state they most likely belong to (i.e., to the state with the largest probability $\gamma(s_{itk})$).

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 state-specific intercepts are updated as follows:

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

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

A.2.2. Update state-specific covariance matrices

In order to find the maximum likelihood estimates for updating the state-specific covariance matrices $\Sigma_k^{new} = \Lambda_k^{new} \Lambda_k^{new'} + \mathbf{D}_k^{new}$, 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).

A.2.3. Update transition intensities

In order to calculate the updates for the intensities, we also have to 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 $\boldsymbol{\theta}^Q$. For the example of $K = 3$ states, the parameters would be $\boldsymbol{\theta}^Q = (q_{12}, q_{13}, q_{21}, q_{23}, q_{31}, q_{32})$. Note that Kalbfleisch and Lawless (1985) suggest to re-parameterize the parameters to $\boldsymbol{\theta}^Q = (\log(q_{12}), \log(q_{13}), \log(q_{21}), \log(q_{23}), \log(q_{31}), \log(q_{32}))$ 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 θ_1^Q to θ_b^Q in $\boldsymbol{\theta}^Q$ 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 θ^Q . This implies that we first calculate the $b \times 1$ vector $\mathbf{S}(\theta^Q)$ with entries

$$S(\theta_u^Q) = \frac{\partial \log L}{\partial \theta_u^Q} = \sum_{i=1}^I \sum_{t=2}^T \sum_{k,l=1}^K \frac{\varepsilon(s_{it-1,l}, s_{itk})}{p_{lk}(\delta_{ti})} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^Q}, \quad (\text{A10})$$

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 (A7)). Next, we calculate the $b \times b$ matrix $\mathbf{M}(\theta^Q)$ with entries

$$M(\theta_u^Q, \theta_v^Q) = \sum_{i=1}^I \sum_{t=2}^T \sum_{k,l=1}^K \frac{\gamma(s_{it-1,k})}{p_{lk}(\delta_{ti})} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_u^Q} \frac{\partial p_{lk}(\delta_{ti})}{\partial \theta_v^Q}, \quad (\text{A11})$$

where $v = 1, \dots, b$, just as u . Finally, we put all the elements together to compute the update θ_{new}^Q :

$$\theta_{new}^Q = \theta_{old}^Q + \mathbf{M}(\theta_{old}^Q)^{-1} \mathbf{S}(\theta_{old}^Q), \quad (\text{A12})$$

where θ_{old}^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 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 $r = 1, \dots, R$ 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.

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

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, Vermunt, Timmerman, and Ceulemans (2017) and Vermunt and Magidson (2016).

Supplementary Material for the Manuscript:
Continuous-Time Latent Markov Factor Analysis for Exploring Measurement Model
Changes Across Time

Vogelsmeier, Leonie V.D.E ^a, Vermunt, Jeroen K. ^a, Böing-Messing, Florian ^{b,a}, De Roover, Kim ^a

^a Tilburg University, The Netherlands

^b Jheronimus Academy of Data Science, The Netherlands

Supplement A

This supplement 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, Tran, & Magidson, 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 infeasible for datasets with

highly unequal time intervals and very fine grids (such as the application that was described in Section 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 Section A.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.

References Supplement A

- 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.

Supplement B

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 C

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 factor analysis. Since desired weight loss is not part of depression, we deemed it important to remove the item from our analyses.

References Supplement C

- 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 Psychotherapy of Depression*. New York, NY: Basic Books Inc Publishers.