A continuous-time mixture latent state-trait Markov model for experience sampling data: Application and Evaluation.

Claudia Crayen and Michael Eid

Freie Universität Berlin

Tanja Lischetzke

University of Koblenz-Landau

Jeroen K. Vermunt

Tilburg University

Author Note

Claudia Crayen, Department of Education and Psychology, Freie Universität Berlin; Michael Eid, Department of Education and Psychology, Freie Universität Berlin; Tanja Lischetzke, Department of Psychology, University of Koblenz-Landau; Jeroen K. Vermunt, Department of Methodology and Statistics, Tilburg University.

This research was supported by a grant to Tanja Lischetzke and Michael Eid from the German Research Foundation (grant LI-1827/1-2). The authors would like to thank two anonymous reviewers for their valuable comments.

Correspondence concerning this article should be addressed to Claudia Crayen, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany. Email: claudia.crayen@fu-berlin.de

Word count: 8518; [Multistudy Report]

Abstract

In psychological research, statistical models of latent state-trait theory are popular for the analysis of longitudinal data. We identify several limitations of available models when applied to intensive longitudinal data with categorical observed and latent variables and inter- and intraindividually varying time intervals. As an extension of available LST models for categorical data, we describe a general mixed continuous-time LST model that is suitable for intensive longitudinal data with unobserved heterogeneity and individually varying time intervals. This model is illustrated by an application to momentary mood data that were collected in an experience sampling study (*N*=164). In addition, the results of a simulation study are reported that was conducted to find out (a) the minimal data requirements with respect to sample size and number of occasions, and (b) how strong the bias is if the continuous-time structure is ignored. The empirical application revealed two classes for which the transition pattern and effects of time-varying covariates differ. In the simulation study, only small differences between the continuous-time model and its discrete-time counterpart emerged. Sample sizes *N*=100 and larger in combination with six or more occasions of measurement tended to produce reliable estimation results. Implications of the models for future research are discussed.

Keywords: Intensive longitudinal data, mixture distribution models, latent Markov models, continuous time, multilevel latent class models

Introduction

Models of latent state-trait (LST) theory have been widely applied in psychological research to separate stable from occasion-specific influences in longitudinal data analysis (Steyer, Ferring, & Schmitt, 1992; Steyer, Schmitt, & Eid, 1999; Steyer, Mayer, Geiser, & Cole, 2015). In most applications, LST models (a) refer to continuous observed and continuous latent variables, (b) suppose that the population is homogeneous with respect to the parameters of the models, and (c) assume that the time lag between two adjacent occasions does not change over time and is the same for all individuals considered. In recent years, LST models have been developed that can be applied when these assumptions of traditional LST models are not fulfilled. We will briefly review these extensions and other approaches developed to meet these challenges. We will go on to integrate these approaches into a general mixture distribution LST model for continuous-time data. In an application of the model to momentary mood data, we go beyond existing applications by including information on varying time intervals and by including time-varying covariates.

For example, Eid and Hoffmann (1998) have developed an LST model for polytomous observed and continuous latent variables, and Eid and Langeheine (1999) and Eid (2002, 2007) have shown how LST models for categorical observed and categorical latent variables can be defined in the framework of latent class analysis. Here, latent classes represent the occasion-specific latent state, which is not continuous but limited to a few state types (e.g., good mood vs. bad mood). In the same way, latent classes represent stable interindividual differences in the trait by some trait types (e.g., cheerful person vs. melancholic). The stable trait influence is reflected, for example, in the higher probability of being in a good mood for cheerful persons. The latent states are mapped to the manifest responses via conditional response probabilities, which reflect measurement error when the mapping is not perfect (a person in a good mood state does not necessarily check the good mood response). Concerning the homogeneity of the population, mixture distribution LST models can be applied when the population is thought to consist of different subpopulations differing in the parameters of a LST model (with both continuous or categorical traits and indicators). Eid and Langeheine (Eid & Langeheine, 2003, 2007; Langeheine & Eid, 2003) have shown how mixture distribution LST models for categorical variables can be defined and applied to separate stable from variable individuals in longitudinal data analysis. Rijmen, Vansteelandt and de Boeck (2008) as well as Vermunt (2010) have developed hierarchical mixture distribution latent Markov models for categorical variables (hierarchical MLM). These take into account the nesting of multiple measurement occasions not only within individuals but, for example, also within different days. In these models, there are day-specific classes that differ in the within-day change process. These day classes represent “day-specific traits”, that is, interindividual differences that are stable within a day. For example, individuals’ mood course could differ between weekdays and the weekend. Moreover, changes on the between-day level are considered by a second Markov process on the level of day-classes. Crayen, Eid, Lischetzke, Courvoisier, and Vermunt (2012) applied the hierarchical MLM to data from an experience sampling study to assess interindividual differences in mood regulation. They were able to predict latent classes of different mood fluctuation patterns with trait measures of perceived effectiveness in mood regulation. Courvoisier, Eid, and Nussbeck (2007) developed a similar mixture distribution LST model for continuous variables.

All models described so far assume that time lags do not differ between individuals or across time. A violation of this assumption is particularly troublesome if there is an autoregressive structure in an LST model, because the size of an autoregressive parameter depends on the time lag (Eid, Courvoisier & Lischetzke, 2012). With respect to varying time intervals, modern continuous-time approaches which incorporate trait variables have been developed for autoregressive models with continuous observed and continuous latent variables (Oud & Jansen, 2000; Oud, 2002; Voelkle, Oud, Davidov, and Schmidt, 2012) and recently implemented into an R package (ctsem; Driver, Oud, and Voelkle, in press). However, continuous-time approaches have not been integrated explicitly into the LST framework. With respect to categorical latent variables, Böckenholt (2005) presented a continuous-time Markov model for categorical observed and latent variables. Yet this model does not allow for population heterogeneity with respect to the continuous change process. The aim of the current paper is to describe a general mixture distribution LST model for continuous-time data that has not been considered so far. It integrates the discrete-time hierarchical mixture latent Markov model and continuous-time approaches. We will refer to it as hierarchical continuous-time mixture latent Markov model (CT-MLM) and illustrate it with an empirical example. To be appropriate for experience sampling method (ESM) data, the model should be flexible enough to integrate different types of nesting (occasions of measurement nested within days and days nested within individuals).

For a long time, the use of such longitudinal models for categorical data was limited to a small number of measurement occasions. Hence, these models could not be applied to studies with many occasions of measurement such as ESM studies. Vermunt, Tran, and Magidson (2008) described an efficient algorithm that allows the estimation of longitudinal models for data with many measurement occasions and relatively small sample sizes. These developments opened up the field of possible applications to ESM data. This efficient algorithm was used to estimate the parameters in our empirical example of the hierarchical CT-MLM. However, little is known about the conditions under which this estimator leads to valid results. Therefore, we conducted a Monte Carlo simulation study to scrutinize the minimal conditions with respect to sample size and number of occasions for validly applying this model.

As Eid (2002) has shown, LST models for categorical variables are special cases of mixture distribution latent Markov models (or the latent mixed Markov model; Langeheine & van de Pol, 1990). In these models, occasion-specific categorical latent state variables are linked by an autoregressive (Markov) structure. The latent state of an occasion depends on the latent state of the previous occasion. Variability is expressed in probabilities of transitioning from one latent state to another. The trait aspect is represented by a latent class variable, which allows for unobserved subgroups of variability patterns (several Markov chains with different parameters). Mixture distribution latent Markov models are more general than mixed LST models for categorical observed and categorical latent variables, because some restrictions on response and transition parameters can be relaxed. For example, mixed LST models contain no interaction effects of traits and states, so that the state effects are the same for different trait types within the same chain (see Eid & Langeheine, 2003). Therefore, we will develop the hierarchical CT-MLM as a special mixed latent Markov model for continuous time, and we will define the model by referring to the terminology of Markov models.

The paper is organized as follows. First, we will show how the hierarchical CT-MLM can be defined. We will then present an application of this model to ESM data. Finally, we will evaluate the model performance and explore minimal data requirements for the application of the model by conducting a Monte Carlo simulation study.

The Mixture Latent Markov Model

We start our description of the discrete-time hierarchical mixture latent Markov model with the measurement part, progressing to the within-day structural part and finally to the between-day structure. Subsequently, the continuous-time approach is introduced. Let y_{idtj} denote the observed response from subject i ($i = 1, \dots, N$), measured on study day d ($d = 0, \dots, D$) at occasion t ($t = 0, \dots, T$) on one of a set of J manifest indicators ($j = 1, \dots, J$). The observed response can take on one of M categories $(m = 1, \ldots, M)$. While number of categories could vary across indicators and number of occasions could vary across days and subjects, we refrain from additional indexing (e.g., M_j) for better readability. For each measurement occasion *dt*, there is a latent state-class variable SC_{dt} with *K* possible values $(k = 1, \ldots, K)$. These latent categories are related to the categories of the manifest indicators via conditional response probabilities $P\left(y_{idtj} = m \middle| SC_{dt} = k\right)$. Unless restricted, there are $(M - 1) * K$ free conditional response probabilities per indicator. The local independence assumption from standard latent class analysis (Goodman, 1974) is made here, stating that there is no association between the indicators of the same occasion conditional on the common latent state.

The latent state categories can be labeled according to the association pattern eminent from the conditional response probabilities. The conditional response probabilities are usually restricted to be the same over time to preserve the interpretation of the latent states. For our mood example, we assume that the number and meaning of manifest categories directly carries over to the latent state categories (i.e., *unpleasant* - *pleasant* - *very pleasant*). If the manifest indicators measured the latent construct perfectly (without measurement error), the conditional response probabilities given the “true” state would equal 1, and conditional response probabilities given all other states would equal 0.

Therefore, any value lower than 1 and greater than 0 indicate some degree of measurement error. The within-day fluctuation process we are interested in is now regarded on the level of the occasion-specific latent state variable SC_{dt} (we consider only the within-day process for now but keep the index for the day). The basic temporal structure of the latent states adopted here is that of a simple first-order stationary Markov process. This means that each state is only dependent on the previous one and that the strength of this dependency is the same across occasions. These assumptions can be relaxed and tested to some degree, but we will focus on this most parsimonious case that is well-suited as a starting point. This process can be conveniently described by a set of K initial state probabilities $P(SC_{d0} = k)$ and (in the discrete-time case) a set of $K * K$ transition probabilities $P(SC_{dt} = k|\ SC_{dt-1} = k')$ with $k \neq k'$.

It is of course very restrictive to assume that every person in the sample is well described by the same process. Instead, the typological approach used here assumes that there is more than one typical pattern, and that individuals can be grouped together according to their transition pattern (i.e., it is not necessary to assume individual patterns). Such a latent categorical variable that classifies individuals according to their within-day transition pattern is situated on the day level and will be called day-class variable (*DC*) here. The number of categories $(l = 1, \dots, L\ )$ has to be determined from theory or from the data in an exploratory manner. Because there are multiple consecutive days (*D*), there is one *DC* for each day. Such a day-specific class can be considered a “day-specific trait” that represents stable day-specific differences between individuals in their within-day variability and change process. However, individuals can switch membership in a day-specific class across days.

In our example, there are two day-classes that differ with regard to their typical initial mood and their mood fluctuation pattern. Conditioning the probabilities introduced so far on the day-class of the particular day yields the conditional response probabilities $P(y_{idtj} = m|SC_{dt} = k, DC_d = l)$, the initial state probabilities $P(SC_{d0} = k|DC_d = l)$, and state transition probabilities $P(SC_{dt} = k\ |SC_{dt-1} = k', DC_d = l)$. For the Markov process on the between-day level, there are initial day-class probabilities $P(DC_0 = l)$ – also known as the class sizes – and day-class transition probabilities $P(DC_d = l\ |DC_{d-1} = l')$ with $l \neq\ l'$.

Again, relaxing the assumption that these probabilities hold for the total sample, a stable latent trait-class (TC) variable with a number of categories $(w = 1, \dots, W)$ is introduced with class size $P(TC = w)$. Just like the state-class process is conditioned on the day-class, the day-class process is now conditioned on the trait-class. The probability for a certain day-class now depends on the trait class, $P(DC_0 = l \mid TC = w)$, as well as the transition probabilities on the between-day level: $P(DC_d = l \mid DC_{d-1} = l', TC = w)$.

Now we define the model for the probability of subject *i*'s responses $\mathbf{y}_i$ as a combination of the probability types introduced. For the most general case, we also condition the response probabilities, the initial state probabilities, and the state transition probabilities on the trait-class. This way, there could be trait classes with completely different parameters. From a substantive point of view, it is often preferable to have some degree of invariance, at least in the measurement part of the model. The between-day sub-model is given by

$$P(\mathbf{y}_i|\mathbf{z}_i) = \sum_{w=1}^{W}\sum_{l_0=1}^{L}\sum_{l_1=1}^{L}\dots\sum_{l_{D_i}=1}^{L} P(TC = w \mid \mathbf{z}_i)\, P(DC_0 = l \mid TC = w, \mathbf{z}_{i0})$$
$$\left[\prod_{d=1}^{D} \begin{matrix} P(DC_d = l \mid DC_{d-1} = l', \\ TC = w, \mathbf{z}_{id}) \end{matrix}\right]\left[\prod_{d=0}^{D} P(\mathbf{y}_{id}|DC_d = l\,, TC = w, \mathbf{z}_{id})\right], \tag{1}$$

and the within-day part by

$$P(\mathbf{y}_{id}|TC = w, DC_d = l, \mathbf{z}_{id})$$
$$= \sum_{k_0=1}^{K}\sum_{k_1=1}^{K}\dots\sum_{k_{T_d}=1}^{K} P(SC_{d0} = k|DC_d = l, TC = w, \mathbf{z}_{id0})$$
$$\left[\prod_{t=1}^{T_d} P(SC_{dt} = k \mid SC_{dt-1} = k'\,, DC_d = l, TC = w, \mathbf{z}_{idt})\right] \tag{2}$$
$$\left[\prod_{t=1}^{T_d}\prod_{j=1}^{J} P\left(y_{idtj} = m\middle|SC_{dt} = k, DC_d = l, TC = w, \mathbf{z}_{idt}\right)\right],$$

with **z** denoting covariates on the various levels.

Continuous time

So far, we have treated the within-day measurement occasions as if they were equally spaced (with a constant time interval). However, in ESM studies, intervals between measurement occasions usually vary within and between individuals. The set of transition probabilities from a discrete-time model is most correctly interpreted as being representative for a time interval equal to the overall mean interval. Still, the estimated transition process will not be precise, because the effect of the previous state is assumed to be constant across time. In the data, more closely spaced observations will have stronger associations (higher stability) than more distant observations. Another assumption in discrete-time models concerns the timing of transitions: These can only occur when a state is observed. A continuous-time process with transitions occurring at any point in time is often more in line with psychological theory. For example, continuous trajectories of emotions can provide information on evaluative processes operating in the extended process model of emotion regulation (Gross, 2015; Kuppens & Verduyn, 2015).

The continuous-time approach has a long history in Markov model theory (Singer & Spilerman, 1976; Kalbfleisch & Lawless, 1985; Böckenholt, 2005). The main difference between the discrete-time (DT in the following) and the continuous-time (CT) approach is that instead of transition probabilities, transition intensities are defined (Coleman, 1981; Kalbfleisch & Lawless, 1985). Transition intensities (or rates) can be thought of as probability per very small time unit. For a transition from state *a* to state *b*, they can be written as

$$q_{ab} = \lim_{\Delta t \to 0} \frac{P(SC_{dt} = b \mid SC_{d(t-\Delta t)} = a)}{\Delta t} \text{ and} \tag{3}$$

$$q_{ab} * \Delta t = P(SC_{dt} = b \mid SC_{d(t-\Delta t)} = a) \quad a \neq b. \tag{4}$$

When two states are observed at nearly the same time point ($\Delta t \to 0$), it is highly unlikely to observe a transition. The longer the time interval, the more likely it becomes to observe a transition. The transition intensity is the parameter that defines this rate of change in transition probabilities. Just like the transition probabilities in each row of a transition probability matrix sum to 1, the corresponding transition intensities sum to 0, thereby making it convenient to obtain the non-transition rate

$$q_{aa} = -\sum_{a \neq b} q_{ab}. \tag{5}$$

Note that while the specific transition probabilities are a function of the time interval, the transition rates themselves are constant over time (stationary process). Just like the probabilities, they are assumed to be constant for all subjects within the same day-class. In order to obtain a set of transition probabilities from the intensities for a specific time interval, the following relationship is crucial:

$$\mathbf{P}(\Delta t) = e^{\mathbf{Q}_i \Delta t}, \tag{6}$$

where $\mathbf{P}(\Delta t)$ is the transition probability matrix for the specific time interval Δt, $e^{\mathbf{Q}_i \Delta t}$ is the matrix exponential and $\mathbf{Q}_i$ is the transition intensity matrix. For three states, the transition intensity matrix is

$$\mathbf{Q}_i = \begin{pmatrix} -(q_{ab} + q_{ac}) & q_{ab} & q_{ac} \\ q_{ba} & -(q_{ba} + q_{bc}) & q_{bc} \\ q_{ca} & q_{cb} & -(q_{ca} + q_{cb}) \end{pmatrix}. \tag{7}$$

To give an example, the set of transition intensities leading to the transition probabilities reported in Table 3 for Day Class 1 in the applied continuous-time model is

$$\mathbf{Q}_i = \begin{pmatrix} -.70 & .70 & .00 \\ .13 & -.28 & .15 \\ .06 & .66 & -.72 \end{pmatrix}. \tag{8}$$

As would be expected, stability decreases and transition probabilities increase exponentially with a longer time interval. This is illustrated in Figure 1 with the transition probability from an unpleasant to a pleasant mood state as a function of the time interval.

The hierarchical CT-MLM model presented here perfectly matches the initially stated typical characteristics of ESM data: Through its transition process on the level of categorical latent variables, the chronological order of the multivariate longitudinal categorical indicators is taken into account. Depending on the nesting of measures, two processes at different time units are defined. And finally, heterogeneity and varying time intervals are incorporated.

Application

Here, we briefly review the application of the discrete-time hierarchical Mixture latent Markov model reported in Crayen et al. (2012). There, the best fitting DT model that emerged according to information criteria (BIC, AIC3) contained a single stable trait class (W=1), L=2 classes on the day level and K=3 classes on the within-day state-level. This DT model with 33 parameters is a special case of the model developed (e.g., no mixture distribution for day-class transitions) and will serve us as a baseline model. In the current application, we will fit the extended models to the same data as used by Crayen et al. (2012), but we will go beyond Crayen et al. (2012) by extending this baseline model to include (a) information on varying time intervals between within-day measurement occasions and (b) time-varying covariates.

Sample and measures

We use data from the original ESM study described in Crayen et al. (2012). The sample consisted of N=164 students (88 women; mean age = 23.7, SD = 3.3). The ESM period lasted seven days and included eight signals per day. The signaling schedule differed both over time and between individuals. Of the 56 possible individual measurement occasions, 51 were answered on average (SD = 6.1). The length of the time interval between signals was roughly normally distributed (scaled to units corresponding to the mean of 100 minutes: M=1, SD=.2, min=.6, max=1.7, excess kurtosis=-.38).

Two bipolar items (*well–unwell* and *good–bad*) with three categories each (e.g., *very well*, *rather well*, *not well*) of an adapted short version of the Multidimensional Mood

Questionnaire (Steyer, Schwenkmezger, Notz, & Eid, 1997) were used as indicators for momentary pleasant mood. Reports of positive and/or negative events served as time-varying covariates. Subjects were prompted on each measurement occasion with items asking whether there had been any positive (resp. negative) event since the last measurement occasion. Possible answers were no (=0) and yes (=1). Subjects had been instructed during the initial training session to report even ordinary events similar to the ones identified in the daily hassles and uplifts scale (DeLongis, Folkman, & Lazarus, 1988). Subjects reported a positive event on 28% of the occasions and a negative event on 16% of the occasions. There was neither a positive or negative event reported on 62% of the occasions, and both a positive and a negative event on 6% of the occasions.

Data Analysis

The baseline model comprised three latent momentary mood states labeled *unpleasant*, *pleasant*, and *very pleasant*. There were two latent classes on the day level, which differed in the initial state probabilities and the transition pattern of the momentary mood states. Most measurement model parameters were held equal across classes.[1] Day classes were not perfectly stable, so that it was possible (albeit unlikely) for people to change day classes between days. By including information on the length of the time interval (scaled to the mean of 100 minutes), a CT model was estimated which did not differ otherwise from the original DT model. In a third model, the dummy coded positive and negative event variables were included in the CT model to assess the relation between reported events and the momentary states. All models were estimated using Latent Gold 5.0 (Vermunt & Magidson, 2013).

Results

[1] There was a constant difference between classes in the log odds of adjacent categories (parameters 8 and 15 in Table 1). This reflected a tendency of the second day class to respond with higher manifest categories compared to the first day class given the same latent state.

A complete list of estimated parameters common to all applied models is provided in Table 1. Model fit information can be found in Table 2. For the CT model, the estimated transition intensity parameters can be used to calculate transition probabilities for any given time interval (R script provided in Electronic Supplementary Material 1). As an example, we obtained two sets for intervals that represent the extremes of the sampled intervals (one hour, *d*=.6 and three hours, *d*=1.8). Resulting transition probabilities for all models are given in Table 3. Transition probabilities obtained from the CT model for the mean interval of 100 minutes are congruent with the DT values and therefore omitted. In each of the three sets of transition probabilities, the distinct character of the day classes became apparent: The larger Day Class 1 that comprised about two thirds of the sample was very stable in the pleasant mood state and was likely to return to this state. For the smaller Day Class 2, the unpleasant and pleasant mood states were less stable, but the very pleasant mood state was the most stable. This pattern appeared refined in the results from the CT covariate model, in which transition probabilities differed according to the report of a positive and/or a negative event: When no event was reported, both day classes were more likely to stay in and to return to the pleasant mood state compared to estimates from the model without covariates. However, there was a stronger tendency to stay in a positive mood class or to move into it in the second day class. The second class was better able to maintain positive mood and to repair negative mood.

In order to evaluate the strength of the relationship between reported events and momentary states within each class, it is useful to consult Table 4 and the lower half of Table 3. Compared to situations without any reported event, transitioning into or staying in an unpleasant mood state was much more likely in both classes when a negative event was reported. Even in these situations, Day Class 2 still exhibited a substantial probability of staying in the very pleasant mood state. For situations in which a positive event was reported, the pattern differed between classes. Particularly interesting is the high probability of a direct transition from an unpleasant to a very pleasant mood state (.60) in Day Class 2. A more fine-grained study design would be necessary to shed light on the underlying mechanisms of mood-event interaction.

Simulation study

Given that few studies to date applied or simulated the hierarchical CT-MLM model, little is known about the general performance and data requirements of the model. Our goal was to explore the minimal data requirements needed for trustworthy results by means of a Monte Carlo simulation study. Monte Carlo simulations are a tool to evaluate the performance of statistical models under various conditions (Harwell, Stone, Hsu, & Kirisci, 1996). Multiple data sets are sampled based on a population model with known parameters. The data sets may differ with regard to sample size, or, in our case, number of time points. Each sampled data set is then analyzed along the lines of real empirical data and all results from one condition are aggregated to yield sampling distributions for the estimated parameters. These can then be compared to known population parameters to detect systematic deviations (bias). In a Monte Carlo simulation study, the sample and model characteristics are the independent factors and indicators of model performance are the dependent measures. Our general expectations were that additional empirical information would lead to more precise parameter estimation.

Method

Conditions. Apart from comparing the performance of the two model types (CT vs DT), the amount of empirical information was varied according to three design aspects of an ambulatory assessment study: The number of subjects in the sample (N), the length of the study period (or number of days, D), and the number of measurement occasions within each day (T). Based on a literature review of ambulatory assessment studies that examined affective processes (e.g., Hedeker, Mermelstein, Berbaum, & Campbell, 2009; Miller, Vachon, & Lynam, 2009), five values for the sample size were chosen with an emphasis on small samples: 35, 50, 75, 100, and 150. For the number of days, a short period (D = 3) and the very common period of one week (D = 7) were selected. The number of measurement occasions within a day (T) varied between 4, 6, and 8. The simulation design was fully crossed and resulted in 30 conditions per model type (CT vs. DT).

Data generation and estimation. The population model for the simulation study was a simplified version of the CT model reported in the application.[2] Note that the data generating model was always a CT model, because even in most DT applications, the underlying psychological process is assumed to be continuous in nature. Population parameters for the DT model were adjusted depending on the mean time interval in the generated data. All population parameters can be found in Table 1 (CT column in the simulation section). For each CT condition, 500 replications were generated. First, time points were randomly drawn from a uniform distribution within segments of a presumed observation day lasting from 8 am to 10 pm (840 minutes), with the additional condition of a minimum distance of 30 minutes. The resulting time intervals for the different number of within-day measurement occasions were approximately normally distributed within their possible value range. Based on the time intervals and the parameters of the population model, the manifest responses on the two items with three categories were generated by Latent Gold 5 (Vermunt & Magidson, 2013). Within the same software, both CT and DT models were applied to each of the 15,000 data sets with the population values serving as starting values, the number of iterations for the EM algorithm set to 300 and the number of iterations for the Newton-

[2] Four parameters associated with the conditional response probabilities linking the extreme manifest categories to the opposite latent state-class were fixed to low values (implying probabilities near 0). In addition, four transition parameters for the direct transition between the extreme state-classes (unpleasant to very pleasant) were also fixed to extremely low values. In CT models, this still allows for indirect transitions through the pleasant state. The estimated DT models, however, had to include these additional parameters. They are colored grey in Table 1 and did not enter the analysis.

Raphson algorithm set to 200. The estimated parameters, standard errors and statistics were exported and further processed in R (R CoreTeam, 2015).

Measures. In order to quantify estimation problems, the proportion of non-converged replications and replications with estimation problems was recorded for each condition. As a measure of parameter estimation quality, the root median squared error (*RMdSE*) was calculated. It is calculated as the square root of the median of the squared differences between parameter estimate and population value. This median-based measure is less sensitive to few extreme estimates than measures commonly used in the context of factor analysis (mostly standardized deviations, see, for example, Bandalos, 2006). Extreme estimates occur more frequently when the loglinear and logit parametrizations are used. As a robust measure for the quality of standard error estimation, the median width of the 95% CI (*CIMd*) was used. The standardized standard error bias measure often used as an outcome in factor analysis simulations derives its population value from the simulation estimates, which are themselves susceptible to bias in small samples.

Statistical Analysis. For the calculation of the performance measures, non-converged replications and replications with other estimation problems (mainly rank deficiency) were excluded. In addition, single parameter estimates with an absolute value of 15 or larger and standard errors with values above 100 were counted as boundary values and, together with the corresponding standard error or parameter estimate, excluded from further analysis (but remaining parameters of the replication were kept). Inspection of day-class specific parameters did not show any indication for label-switching.

Performance measures were aggregated as the median across parameters of the same type. Parameters were divided into five types: (1) parameters pertaining to the measurement part of the model, (2) initial state parameter, (3) state-class transition parameters, (4) the (initial) day-class size parameter, and (5) day-class transition parameters. The parametrization for parameter types 2, 3, and 5 differed between CT and DT models. The bias measures for the DT model were based upon the converted parameters (which differed for different *T*; see Table 1). The time interval between days was set to a constant corresponding to 8 hours (*d*=4.8). The evaluation of estimation quality is based on descriptive information and graphical representation. In addition, a table with effect sizes is provided in Electronic Supplementary Material 2.

Results

Estimation problems. Overall, 20% of replications in the CT condition and 14% in the DT condition exhibited estimation problems. As can be seen in Figure 2, the rate of estimation problems was very high in conditions of the short study period ($D = 3$) in combination with few measurement occasions ($T = 4$) and small sample sizes ($N < 100$), but much lower for other conditions. There were hardly any estimation problems in conditions with many measurement occasions (T=8). Non-convergence was more frequent for the CT model (6% of all replications) than for the DT model (1%). The remaining estimation problems were almost uniquely due to rank deficiency and equally frequent in both model types. Because replications with estimation problems were excluded from further analysis, five conditions were left with less than 100 valid replications.[3] These conditions have been omitted in the evaluation of estimation quality. All remaining conditions had at least 230 valid replications.

Estimation Quality. As expected, parameter recovery and standard error size improved with larger sample sizes, a longer study period and more measurement occasions. More surprisingly, there were only minor differences between the model types. Because the estimation of the different types of transition parameters in the different models is of core interest here, we will focus on the evaluation of state-class and day-class transition parameters

[3] The five conditions are (1) DT model, *D*=3, *N*=35, *T*=4; (2) DT model, *D*=3, *N*=50, *T*=4; (3) CT model, *D*=3, *N*=35, *T*=4; (4) CT model, *D*=3, *N*=50, *T*=4; and (5) CT model, *D*=3, *N*=35, *T*=6.

(Figures 3 to 6). Figures for the remaining parameter types are provided in Electronic Supplementary Material 3. There was a striking similarity in the estimation pattern of state-class transition parameters between the different model types, which can be observed in Figure 3 (*RMdSE*) and Figure 5 (*CIMd*). Here, the underlying parameters differed (logs of transition rates vs. logits of transition probabilities). For few measurement occasions and a short study period, the CT model exhibited larger standard errors. There was a considerable drop in both measures when the number of measurement occasions was increased from 4 to 6 in short study period conditions. The effect was less pronounced in long study period conditions. Interestingly, an additional increase in sample size above N=75 on in the long study period with many measurement occasions had little effect on the parameter estimation quality, but still on the size of the standard error.

Parameter recovery of the day-class transition parameters was much worse than of the state-class transition parameters (see Figure 4). This is not surprising given the shorter chain of days compared to repeated signal-chains, and the rather small transition probabilities used as population values. Consequently, increasing the length of the study period had a large effect. For a short study period, only conditions with the largest sample size and many measurement occasions seemed to stabilize estimation. Many measurement occasions were also necessary to compensate for smaller sample sizes in the long study period conditions. There was an unexpected peak in parameter bias for D=7, T=4, N=75 for both model types not mirrored in the standard error measure. This may have been due to the fact that these measures were aggregated across only two day-class transition parameters (compared to eight state-class transition parameters). For measurement model parameters, the effect of T and N was especially pronounced for a short study period. For the long study period, *RMdSE* values converged for sample sizes of 75 and larger. However, the effects of T and N on the *CIMd* remained in these conditions. The class size parameter was not well estimated for few measurement occasions. The quality was much better for six or eight occasions. In these conditions, the study period and the sample size had only small effects on the *RMdSE*, while there was a clear effect of the sample size and the number of occasions (but not study period) on the magnitude of the *CIMd*. Notable differences between model types were only found in the estimation of the initial state parameters. Here, the CT model performed better with the overall mean *RMdSE* for the CT model only about half the size (.153) of the corresponding DT value (.286). The *CIMd* reflected the same, with smaller values for the CT model.

Discussion

In the empirical example, the hierarchical CT-MLM model was successfully applied to ESM data. It was also extended to incorporate time-varying covariates—that is, information on situational influences on mood such as negative and positive daily events. The relation between positive and negative events and the mood fluctuation pattern was substantial. This may suggest either that the latent day-classes differ with regard to their reactivity to events, or the likelihood of events varies depending on day class, or both. While the model matches the structure of ESM mood data very well, another potential field of application are burst designs, in which it is common to collect daily measures for several days during bursts that are placed weeks, months, or even years apart (Nesselroade, 1991). For each burst, there could be latent classes of fluctuation patterns, and transitions between these burst-classes could depend on the time lag. This example shows two noteworthy points: First, varying time intervals may also come into play on the higher-level Markov process once the stable day structure is left, and second, the pattern of transition between burst-classes might hold important information and exhibit heterogeneity. In studies on cognitive aging, for example, it could be interesting to separate individuals that are stable in their performance pattern across some time-span from declining or flourishing individuals by including stable trait classes in the hierarchical CT-MLM model.

Concerning minimal data requirements for the simulated data conditions, reasonable precision in estimation was accomplished for three days/bursts only if the design included eight measurement occasions and a sample size of at least N=100. For a standard ESM design with a measurement period of one week, the same precision was accomplished with fewer measurement occasions (6 or even 4) and a smaller sample size (N=75). The simulation study was designed to closely match the properties of EMA studies in mood regulation research. While the resulting data requirements are credible for this line of research, the study is limited to its specific set up with a certain number of indicators, state-classes, day-classes, a single-trait class and a high degree of measurement invariance. Several modeling issues have not been addressed by this study and would benefit from future research. (1) Model performance when the model does not match the population (misspecification); (2) the ability of model fit indices to recover the “true” model, and (3) the performance of competing models that represent population heterogeneity differently, for example with continuous random effects. In general, the decision on the type of latent variables (continuous vs. categorical) cannot be decided by model fit criteria alone. While the categorical nature of the measures in EMA study designs makes state classes seem intuitive, the categorical nature of day-classes and trait classes of fluctuation patterns should be justifiable in theory. In addition, the level at which the latent class is modeled (day or trait) has to be at least partly determined from theory. For example, the day classes in the application were very stable, but a model with day-classes was preferred because (slightly) different mood fluctuation patterns across days can be expected on theoretical grounds. If the construct had been one for which fluctuations across days/bursts would not make sense, variability across days would have to be accounted for differently. Because there is still little experience with the performance of these models, we recommend conducting specific simulations depending on the study planned, taking into account theoretical expectations.

The proposed hierarchical continuous-time mixture latent Markov model has several advantages compared to previously described LST models, especially for the analysis of ESM data. First, because it is a model for categorical observed variables, the measurement model can be tested on the level of the actual response process. The model would allow identifying unknown subgroups of response styles (mixture on the level of measurement occasions). Second, it separates the observed variability into measurement error, true occasion-specific variability and true stable interindividual differences on different time levels (day, week, etc.). Third, the model allows for population heterogeneity in the change process on the various time levels. Fourth, the model takes into account varying time intervals between measurement occasions, so that the parameters describing the change process are independent of time intervals in the specific study design and allow for a better comparison of stability and change scores across studies. This is one of the reasons why the CT model should always be preferred over the DT model – even though the simulation study revealed only small differences in parameter estimation qualities. From a theoretical point of view, the CT model also mirrors the underlying process more accurately if this is construed to be continuous in nature. Combined with the possibility to include covariates on the different time levels, the hierarchical CT-MLM represents a powerful tool that satisfies many challenges posed by complex data resulting from complex study designs.

Electronic Supplementary Material

ESM 1. R script (ESM1_Rates2Probs.pdf). In this script, calculation of transition probabilities from parameters of the continuous-time model is demonstrated.

ESM 2. Table 1 (ESM2_Table.pdf). Table of ANOVA effect sizes for manipulated factors in the simulation study.

ESM 3. Figures 1-6 (ESM3_Figures.pdf). Figures showing the *RMdSE* and *CIMd* pattern for the three remaining parameter types (measurement part, initial state-class, day-class size) in the simulation study.

References

Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. E. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age.

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

Coleman, J. S. (1981). *Longitudinal data analysis*. New York, NY: Basic Books.

Courvoisier, D. S., Eid, M., & Nussbeck, F. W. (2007). Mixture distribution state-trait analysis, basic ideas and applications. *Psychological Methods, 12*, 80-104. doi:10.1037/1082-989x.12.1.80

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D. S., & Vermunt, J. K. (2012). Exploring dynamics in mood regulation-mixture latent Markov modeling of ambulatory assessment data. *Psychosomatic Medicine, 74*, 366-376. doi:10.1097/psy.0b013e31825474cb

DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology, 54*, 486-495. doi:10.1037/0022-3514.54.3.486

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (in press). Continuous Time Structural Equation Modeling with R Package ctsem. *Journal of Statistical Software*.

Eid, M. (2002). A closer look at the measurement of change: Integrating latent state-trait models into the general framework of latent mixed Markov modeling. *Methods of Psychological Research Online*, *7*, 33-52.

Eid, M. (2007). Latent class models for analyzing variability and change. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 591-607). Oxford, UK: Oxford University Press.

Eid, M., Courvoisier, D. S., & Lischetzke, T. (2012). Structural equation modeling of ambulatory assessment data. In M. Mehl & T. S. Conner (Eds.), *Handbook of research methods* (pp. 384-406). New York, NY: Guilford Press.

Eid, M., & Hoffman, L. (1998). Measuring variability and change with an item response model for polytomous variables. *Journal of Educational and Behavioral Statistics*, *23,* 171-193. doi:10.2307/1165244

Eid, M., & Langeheine, R. (1999). The measurement of consistency and occasion specificity with latent class models: A new model and its application to the measurement of affect. *Psychological Methods, 4*, 100-116. doi:10.1037/1082-989x.4.1.100

Eid, M., & Langeheine, R. (2003). Separating stable from variable individuals in longitudinal studies by mixture distribution models. *Measurement: Interdisciplinary Research and Perspective*, *1*, 179-206. doi:10.1207/s15366359mea0103_01

Eid, M., & Langeheine, R. (2007). Detecting population heterogeneity in stability and change of subjective well-being. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in Positive Psychology* (pp. 608-632). Oxford, UK: Oxford University Press.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable Models. *Biometrika, 61*, 215-231. doi:10.2307/2334349

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. Psychological Inquiry, 26 (1), 1-26. doi:10.1080/1047840X.2014.940781

Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20,* 101-125. doi:10.1177/014662169602000201

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction, 104*, 297-307. doi:10.1111/j.1360-0443.2008.02435.x

Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. Journal of the American Statistical Association, 80, 863-871. doi:10.1080/01621459.1985.10478195

Kuppens, P., & Verduynm P. (2015). Looking at emotion regulation through the window of emotion dynamics. Psychological Inquiry, 26 (1), 72-79. doi:10.1080/1047840X.2015.960505

Langeheine, R., & Eid, M. (2003). Mixture distribution state-trait models, current limitations and future directions. *Measurement: Interdisciplinary Research and Perspective, 1*, 232-240. doi:10.1207/S15366359MEA0103_01

Langeheine, R. & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods and Research*, 18, 416-441.

Miller, D. J., Vachon, D. D., & Lynam, D. R. (2009). Neuroticism, negative affect, and negative affect instability: Establishing convergent and discriminant validity using ecological momentary assessment. *Personality and Individual Differences, 47*, 873–877. doi:10.1016/j.paid.2009.07.007

Nesselroade, J. R. (2001). Intraindividual variability in the development within and between individuals. *European Psychologist, 6*, 187-193. doi:10.1027/1016-9040.6.3.187

R Core Team, 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65, 199–215.

Oud, J. H. L. (2002). Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden*, 23 (69), 1-26.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika, 73*, 167-182. doi:10.1007/s11336-007-9001-8

Singer, B., & Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology, 82*, 1-54. doi:10.1086/226269

Steyer, R., Ferring, D., & Schmitt, M. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *2,* 79-98.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits-revised. *Annual Review of Clinical Psychology, 11*, 71-98. doi:10.1146/annurev-clinpsy-032813-153719

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*, 389-408.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)* [The multidimensional Mood Questionnaire]. Göttingen, Germany: Hogrefe.

Vermunt, J. K. (2010). Longitudinal research using mixture models. In K. van Montfort, J. H. L. Oud, and A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 119-152). Heidelberg, Germany: Springer.

Vermunt, J. K., & Magidson, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of Longitudinal Research: Design, Measurement, and Analysis*. Amsterdam, the Netherlands: Elsevier.

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176-192. doi:10.1037/a0027543

Tables

Table 1

Overview of parameter values for the three models in the application and for the simulation.

		Application			Simulation			
		DT	CT	CT cov	CT	DT		
Index	Parametrization (DT)					T=4	T=6	T=8
Initital day-class size								
1	$\ln\left(\frac{P(DC_0=2)}{P(DC_0=1)}\right)$	-0.8	-0.7	-0.7	-0.7			
Measurement part								
2	$\ln\left(\frac{P(y_{dt1}=2\|SC_{dt}=1,DC_d=1)}{P(y_{dt1}=1\|SC_{dt}=1,DC_d=1)}\right)$	-2.3	-2.3	-2.2	-2.4			
3	$\ln\left(\frac{P(y_{dt1}=3\|SC_{dt}=1,DC_d=1)}{P(y_{dt1}=1\|SC_{dt}=1,DC_d=1)}\right)$	-5.6	-5.6	-5.6	-15*			
4	$\ln\left(\frac{P(y_{dt1}=1\|SC_{dt}=2,DC_d=1)}{P(y_{dt1}=2\|SC_{dt}=2,DC_d=1)}\right)$	-3.0	-3.0	-3.1	-2.9			
5	$\ln\left(\frac{P(y_{dt1}=3\|SC_{dt}=2,DC_d=1)}{P(y_{dt1}=2\|SC_{dt}=2,DC_d=1)}\right)$	-4.1	-4.0	-3.9	-3.6			
6	$\ln\left(\frac{P(y_{dt1}=1\|SC_{dt}=3,DC_d=1)}{P(y_{dt1}=3\|SC_{dt}=3,DC_d=1)}\right)$	-8.2	-8.3	-8.5	-15*			
7	$\ln\left(\frac{P(y_{dt1}=2\|SC_{dt}=3,DC_d=1)}{P(y_{dt1}=3\|SC_{dt}=3,DC_d=1)}\right)$	-0.5	-0.4	-0.6	-0.6			
8	$\ln\left(\frac{\frac{P(y_{dt1}=2\|SC_{dt}=1,DC_d=2)}{P(y_{dt1}=1\|SC_{dt}=1,DC_d=2)}}{\frac{P(y_{dt1}=2\|SC_{dt}=1,DC_d=1)}{P(y_{dt1}=1\|SC_{dt}=1,DC_d=1)}}\right)$ etc.	1.2	1.2	1.2	1.2			
9	$\ln\left(\frac{P(y_{dt2}=2\|SC_{dt}=1,DC_d=1)}{P(y_{dt2}=1\|SC_{dt}=1,DC_d=1)}\right)$	-2.7	-2.7	-2.4	-2.7			
10	$\ln\left(\frac{P(y_{dt2}=3\|SC_{dt}=1,DC_d=1)}{P(y_{dt2}=1\|SC_{dt}=1,DC_d=1)}\right)$	-7.4	-7.4	-7.3	-15*			
11	$\ln\left(\frac{P(y_{dt2}=1\|SC_{dt}=2,DC_d=1)}{P(y_{dt2}=2\|SC_{dt}=2,DC_d=1)}\right)$	-3.7	-3.7	-3.8	-3.6			

		Application			Simulation			
		DT	CT	CT cov	CT	DT		
Index	Parametrization (DT)					T=4	T=6	T=8
12	$\ln\left(\frac{P(y_{dt2}=3 \mid SC_{dt}=2, DC_d=1)}{P(y_{dt2}=2 \mid SC_{dt}=2, DC_d=1)}\right)$	-4.3	-4.4	-4.2	-3.6			
13	$\ln\left(\frac{P(y_{dt2}=1 \mid SC_{dt}=3, DC_d=1)}{P(y_{dt2}=3 \mid SC_{dt}=3, DC_d=1)}\right)$	-6.5	-6.7	-6.6	-15*			
14	$\ln\left(\frac{P(y_{dt2}=2 \mid SC_{dt}=3, DC_d=1)}{P(y_{dt2}=3 \mid SC_{dt}=3, DC_d=1)}\right)$	-0.1	-0.1	-0.1	-0.3			
15	$\ln\left(\frac{\frac{P(y_{dt2}=2 \mid SC_{dt}=1, DC_d=2)}{P(y_{dt2}=1 \mid SC_{dt}=1, DC_d=2)}}{\frac{P(y_{dt2}=2 \mid SC_{dt}=1, DC_d=1)}{P(y_{dt2}=1 \mid SC_{dt}=1, DC_d=1)}}\right)$ etc.	2.0	2.0	1.9	1.9			
Day-class transition								
16	$\ln\left(\frac{P(DC_d=2 \mid DC_{d-1}=1)}{P(DC_d=1 \mid DC_{d-1}=1)}\right)$	-4.1	-4.1	-3.7	-2.2	-0.7		
17	$\ln\left(\frac{P(DC_d=1 \mid DC_{d-1}=2)}{P(DC_d=2 \mid DC_{d-1}=2)}\right)$	-2.2	-2.3	-2.6	-2.3	-0.8		
Initial state-class size								
18	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d0}=1, DC_d=1)}{P(SC_{dt}=1 \mid SC_{d0}=1, DC_d=1)}\right)$	1.5	0.5	0.5	0.5	1.5		
19	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d0}=1, DC_d=1)}{P(SC_{dt}=1 \mid SC_{d0}=1, DC_d=1)}\right)$	-1.0	-2.0	-2.4	-2.1	-1.1		
20	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d0}=1, DC_d=2)}{P(SC_{dt}=1 \mid SC_{d0}=1, DC_d=2)}\right)$	1.3	0.1	0.4	0.2	1.4		
21	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d0}=1, DC_d=2)}{P(SC_{dt}=1 \mid SC_{d0}=1, DC_d=2)}\right)$	1.0	-0.1	-0.3	-0.1	1.1		
State-class transition								
22	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d(t-1)}=1, DC_d=1)}{P(SC_{dt}=1 \mid SC_{d(t-1)}=1, DC_d=1)}\right)$	-0.2	-0.4	-0.4	-0.4	0.8	0.3	-0.1
23	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d(t-1)}=1, DC_d=1)}{P(SC_{dt}=1 \mid SC_{d(t-1)}=1, DC_d=1)}\right)$	-2.9	-5.8	-28.4	-15*	-1.4	-2.3	-2.8
24	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d(t-1)}=1, DC_d=2)}{P(SC_{dt}=1 \mid SC_{d(t-1)}=1, DC_d=2)}\right)$	0.3	0.0	-0.1	0.5	1.1	0.9	0.7

		Application			Simulation			
		DT	CT	CT cov	CT	DT		
Index	Parametrization (DT)					*T*=4	*T*=6	*T*=8
25	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d(t-1)}=1, DC_d=2)}{P(SC_{dt}=1 \mid SC_{d(t-1)}=1, DC_d=2)}\right)$	-0.3	-1.2	-1.7	-15*	0.8	0.1	-0.5
26	$\ln\left(\frac{P(SC_{dt}=1 \mid SC_{d(t-1)}=2, DC_d=1)}{P(SC_{dt}=2 \mid SC_{d(t-1)}=2, DC_d=1)}\right)$	-2.3	-2.0	-2.2	-2.0	-1.8	-2.0	-2.2
27	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d(t-1)}=2, DC_d=1)}{P(SC_{dt}=2 \mid SC_{d(t-1)}=2, DC_d=1)}\right)$	-2.1	-1.9	-2.6	-2.0	-1.9	-2.0	-2.2
28	$\ln\left(\frac{P(SC_{dt}=1 \mid SC_{d(t-1)}=2, DC_d=2)}{P(SC_{dt}=2 \mid SC_{d(t-1)}=2, DC_d=2)}\right)$	-1.4	-1.1	-1.8	-0.9	-1.3	-1.3	-1.4
29	$\ln\left(\frac{P(SC_{dt}=3 \mid SC_{d(t-1)}=2, DC_d=2)}{P(SC_{dt}=2 \mid SC_{d(t-1)}=2, DC_d=2)}\right)$	-0.7	-0.8	-1.35	-0.7	-0.1	-0.4	-0.6
30	$\ln\left(\frac{P(SC_{dt}=1 \mid SC_{d(t-1)}=3, DC_d=1)}{P(SC_{dt}=3 \mid SC_{d(t-1)}=3, DC_d=1)}\right)$	-2.3	-2.8	-2.8	-15*	-1.3	-2.1	-2.7
31	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d(t-1)}=3, DC_d=1)}{P(SC_{dt}=3 \mid SC_{d(t-1)}=3, DC_d=1)}\right)$	-0.2	-0.4	-0.12	-0.3	0.9	0.4	0.0
32	$\ln\left(\frac{P(SC_{dt}=1 \mid SC_{d(t-1)}=3, DC_d=2)}{P(SC_{dt}=3 \mid SC_{d(t-1)}=3, DC_d=2)}\right)$	-2.7	-2.9	-3.5	-15*	-2.1	-2.6	-2.9
33	$\ln\left(\frac{P(SC_{dt}=2 \mid SC_{d(t-1)}=3, DC_d=2)}{P(SC_{dt}=3 \mid SC_{d(t-1)}=3, DC_d=2)}\right)$	-1.3	-1.1	-0.9	-0.9	-0.6	-0.8	-1.1

Note. Models in the application study: DT=Discrete-time model; CT=Continuous-time model; CT cov = Continuous-time model with time-varying covariates. Models in the simulation study: CT=data generating continuous-time model; DT= estimated discrete time model. DT values stated were calculated from CT parameters and used in the assessment of estimation performance. Unless otherwise stated, DT model values equal the corresponding CT values (day-class size and measurement part) or the value of the neighboring DT model (day-class transition, initial transition). Values with an asterisk (*) are fixed parameters. Values marked grey were the four additional transition parameters in DT models not included in the assessment of estimation performance.

Table 2

Goodness-of-Fit Statistics for models in the application.

Model	LL	No. par.	BIC
DT	-10,166	33	20,500
CT	-10,177	33	20,522
CT cov	-9,717	57	19,724

Note. DT=Discrete-time model; CT=Continuous-time model; CT cov = Continuous-time model with time-varying covariates; LL = logLikelihood; par. = parameters; BIC = Bayesian Information Criterion.

Table 3

Day-class specific transition probabilities in different models, depending on time interval (CT model) and reported events (CT covariate model).

				Day Class 1			Day Class 2		
				State			*State*		
Model	*d*	*Event*	*State-1*	1	2	3	1	2	3
DT	-	-	1: ☹	.53	.44	.03	.33	.43	.24
			2: 😐	.09	.82	.10	.14	.57	.29
			3: ☺	.05	.43	.52	.05	.21	.74
CT	.6 (1 hour)	-	1: ☹	.67	.32	.01	.49	.36	.16
			2: 😐	.06	.87	.07	.11	.69	.20
			3: ☺	.03	.30	.66	.03	.15	.82
	1.8 (3 hours)	-	1: ☹	.34	.59	.07	.20	.44	.36
			2: 😐	.12	.76	.13	.14	.47	.39
			3: ☺	.09	.57	.34	.08	.29	.63
CT cov	1 (100 min)	None	1: ☹	.53	.45	.01	.36	.49	.15
			2: 😐	.08	.88	.04	.09	.73	.18
			3: ☺	.05	.54	.40	.03	.29	.67
		Positive	1: ☹	.26	.61	.12	.03	.37	.60
			2: 😐	.02	.74	.24	.03	.52	.45
			3: ☺	.01	.31	.68	.01	.12	.87
		Negative	1: ☹	.77	.23	.00	.72	.24	.03
			2: 😐	.37	.62	.01	.36	.55	.09
			3: ☺	.42	.51	.07	.30	.33	.36

Note. DT = Discrete-time model; CT = Continuous-time model; CT cov = Continuous-time model with covariates; *d* = length of time interval in 100 minutes; State-1 = occupied state at preceding time point; State 1: ☹ = unpleasant mood state; State 2: 😐 = pleasant mood state; State 3: ☺ = very pleasant mood state. Values for the infrequent concurrent report of a positive and a negative event do not appear in the table.

Table 4

Loglinear effects of events on the transition intensities in the CT covariate model (standard errors in parentheses). The respective "no transition" category served as reference. Negative values imply a decrease in transition intensity (increase in stability), positive values describe an increase.

		Day Class 1			*Day Class 2*		
		State			*State*		
	State-1	1	2	3	1	2	3
Positive Event	1: ☹	0	0.72 (0.09)	1.35 (0.57)	0	0.98 (0.34)	2.42 (0.42)
	2: 😐	-0.93 (0.37)	0	1.61 (0.22)	0.22 (0.44)	0	0.87 (0.22)
	3: ☺	-30.96 (1.34)	-0.66 (0.29)	0	-0.88 (0.79)	-0.85 (0.30)	0
Negative Event	1: ☹	0	-0.63 (0.12)	-2.91 (2.95)	0	-0.84 (0.26)	-2.06 (0.44)
	2: 😐	1.63 (0.18)	0	-0.77 (0.44)	1.19 (0.32)	0	-0.20 (0.32)
	3: ☺	2.66 (1.07)	0.77 (0.39)	0	2.64 (0.99)	0.51 (0.35)	0

Figure captions

Figure 1. Transition probability from the unpleasant to the pleasant mood state for Day Class 1 as a function of the time interval as estimated in the continuous-time model. Marked are the time intervals used in Table 3 (d=0.6 and d=1.8).

Figure 2. Proportion of 500 simulated replications affected by estimation problems such as non-convergence. The left panels contain results for continuous-time models, the right panels contain results for discrete time models. Upper panels represent conditions with D=3 days, lower panels represent conditions with D=10 days. Lines denote different numbers of measurement occasions.

Figure 3. Estimation quality (root median squared error) of the state-class transition parameters for simulated conditions. Conditions with fewer than 100 valid replications are omitted.

Figure 4. Estimation quality (root median squared error) of the day-class transition parameters for simulated conditions. Conditions with fewer than 100 valid replications are omitted.

Figure 5. Standard error estimation (median width of the confidence interval) for state-class transition parameters and simulated conditions. Conditions with fewer than 100 valid replications are omitted.

Figure 6. Standard error estimation (median width of the confidence interval) for day-class transition parameters and simulated conditions. Conditions with fewer than 100 valid replications are omitted.

Figure 1

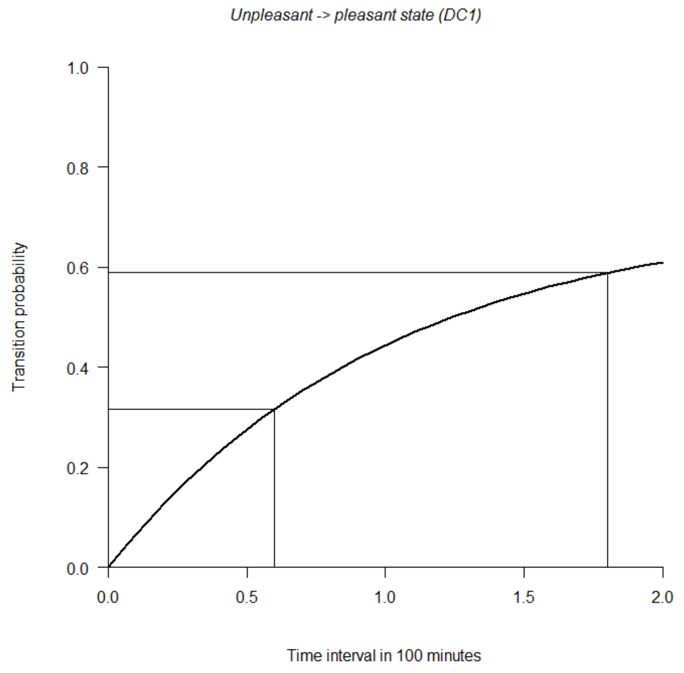


Figure 2

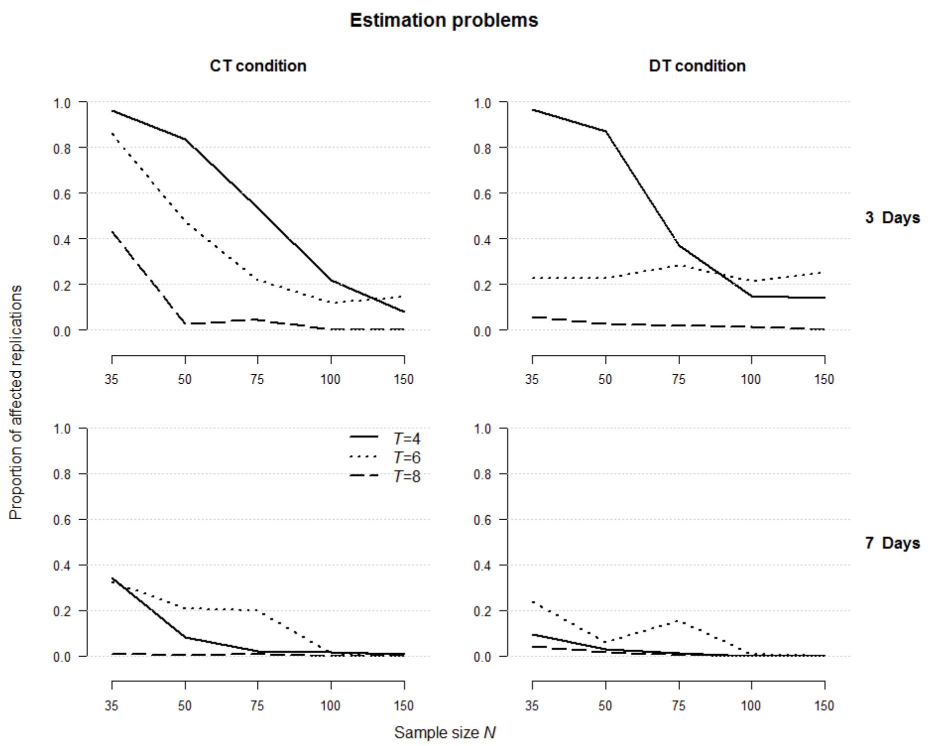


Figure 3

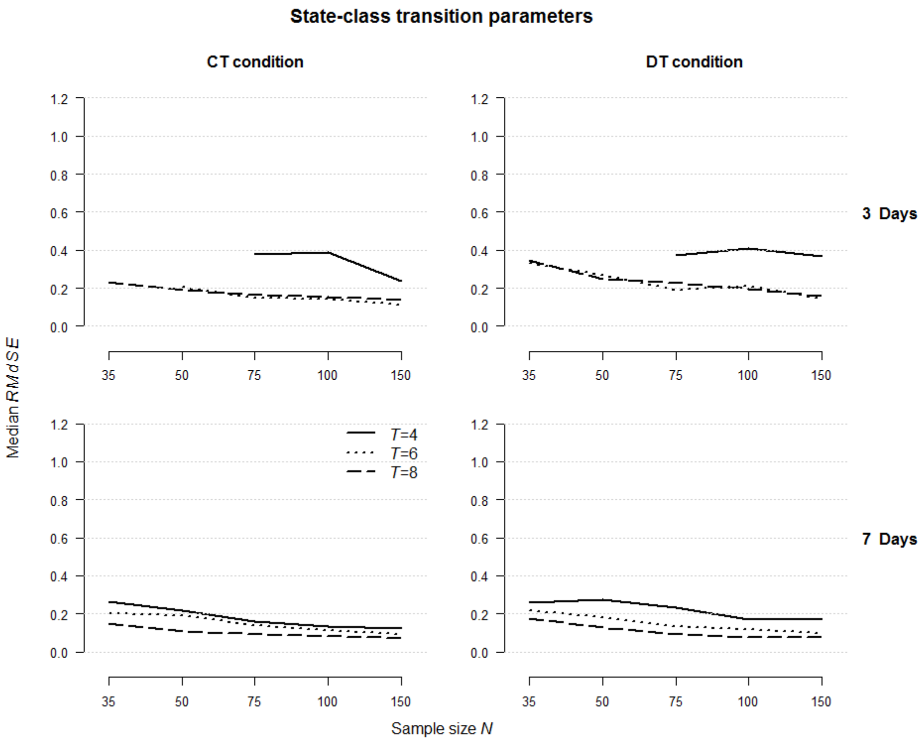


Figure 4

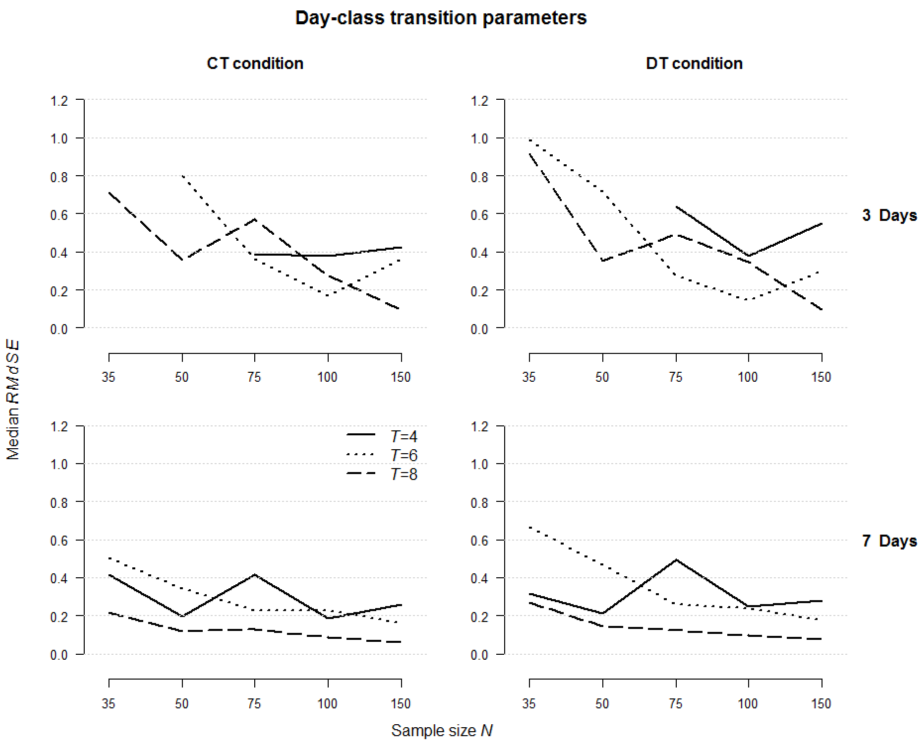


Figure 5

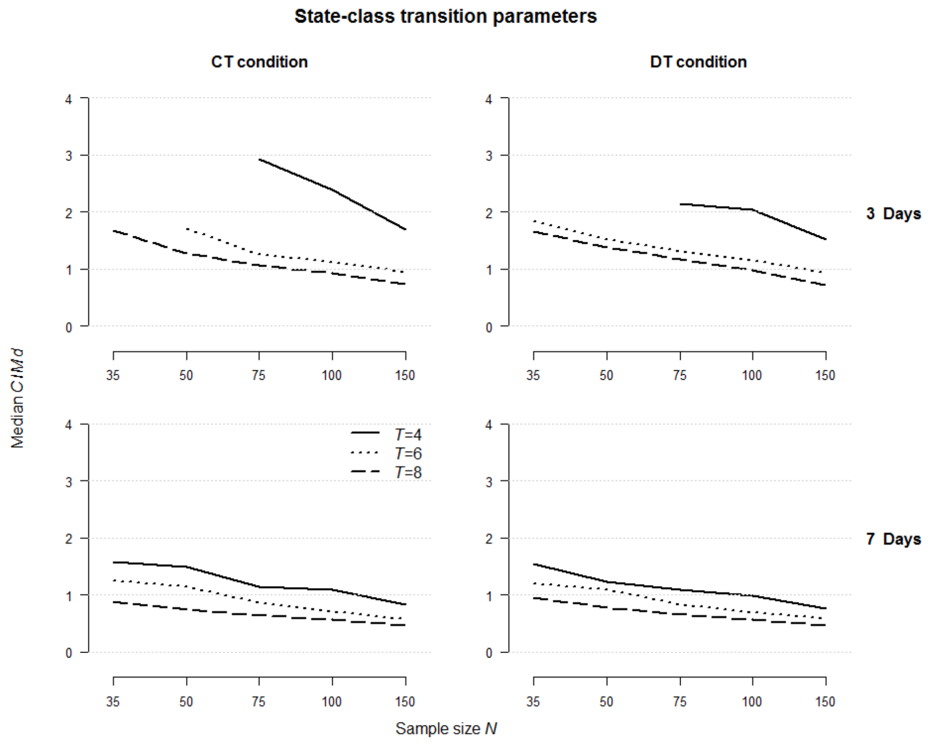
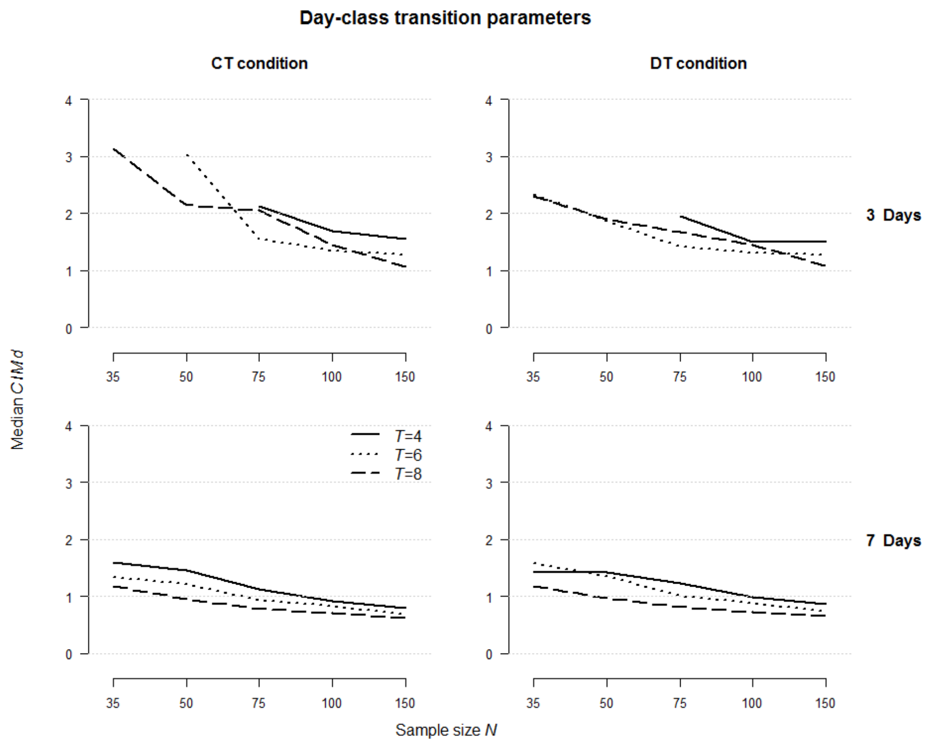

State-class transition parameters
CT condition
DT condition
3 Days
7 Days
Median CIMd
T=4
T=6
T=8
Sample size N
0
1
2
3
4
35
50
75
100
150

Figure 6


Day-class transition parameters
CT condition
DT condition
3 Days
7 Days
Median CIMd
T=4
T=6
T=8
Sample size N
0
1
2
3
4
35
50
75
100
150