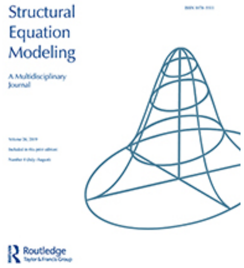

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

# A Nonlinear Dynamic Latent Class Structural Equation Model

Augustin Kelava & Holger Brandt

*Structural Equation Modeling: A Multidisciplinary Journal*, 26: 509–528, 2019

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1555692



# A Nonlinear Dynamic Latent Class Structural Equation Model

Augustin Kelava[1] and Holger Brandt[2]

[1]*University of Tübingen*
[2]*University of Kansas*

In this article, we propose a nonlinear dynamic latent class structural equation modeling (NDLC-SEM). It can be used to examine intra-individual processes of observed or latent variables. These processes are decomposed into parts which include individual- and time-specific components. Unobserved heterogeneity of the intra-individual processes are modeled via a latent Markov process that can be predicted by individual- and time-specific variables as random effects. We discuss examples of sub-models which are special cases of the more general NDLC-SEM framework. Furthermore, we provide empirical examples and illustrate how to estimate this model in a Bayesian framework. Finally, we discuss essential properties of the proposed framework, give recommendations for applications, and highlight some general problems in the estimation of parameters in comprehensive frameworks for intensive longitudinal data.

**Keywords**: time-series analysis, dynamic structural equation model, intensive longitudinal data, Bayesian methods

In the past years, so-called ambulatory assessment of intensive longitudinal data has become technically very easy. Electronic devices such as wearable devices and smartphones allow for a high number of repeated measures. These cover self-report data, peripheral physiological measures, and objective data, such as movements (Trull & Ebner-Priemer, 2014). Parts of these data are relatively stable or collected only once at the beginning of a study (e.g., dispositions, traits, covariates). They describe inter-individual differences and may be used to predict future developments. Other parts of the data are frequent measures within individuals (which describe intra-individual changes). Time-specific variables such as interventions or other events can be of importance. Their relevance can be investigated with regard to how they affect the inter- and intra-individual differences in the development process. However, comprehensive statistical frameworks are needed which are capable of describing intra-individual changes that depend on inter-individual differences and time-specific effects.

For a small number of repeated measures, latent growth curve models (e.g., Bollen & Curran, 2006) or —equivalently—multilevel models (e.g., Asparouhov & Muthèn, 2016; Rabe-Hesketh, Skrondal, & Pickles, 2004) can be used, which describe intercepts and changes as random latent variables. For a larger number of repeated measures, autoregressive effects can be included. In the past, many approaches were proposed for the analysis of these data, such as dynamic factor models (Molenaar, 1985, 2017; Zhang, Hamaker, & Nesselroade, 2008; Zhang & Nesselroade, 2007), (multivariate) time-series models (e.g., Box, Jenkins, Reinsel, & Ljung, 2015; Durbin & Koopman, 2001), latent Markov models (e.g., Altman, 2007), and many more.

Recently, Asparouhov, Hamaker, and Muthén (2017a) proposed a dynamic latent class analysis (DLCA) approach. There are two important properties of the DLCA approach: (1) Each individual is a member of a latent class at each time point with a specific probability. The latent class membership follows a Hidden Markov Model process, that is, future class-membership depends on previous class-membership. The

Correspondence should be addressed to Augustin Kelava. E-mail: augustin.kelava@uni-tuebingen.de

individual-specific transition probabilities are estimated as (between-level) random effects which are parameterized by a structural equation model or factor model. For example, inter-individual differences (such as differences in personality traits, e.g., vulnerabilities) predict the probabilities of changes between the latent classes (e.g., states of intentions to quit college). (2) The within-level model is a dynamic (time-series) model with autoregressive effects of the latent variables. For example, intra-individual changes of affective states could be part of the model (e.g., trajectories of moods). The DLCA approach is an important combination of mixture modeling and latent time-series modeling.

Another comprehensive framework, the dynamic structural equation model (DSEM), was proposed by the same authors (Asparouhov, Hamaker, & Muthén, 2017b). (1) The DSEM framework separates (1.a) subject-specific and (1.b) time-specific random effects (on the between-level). (2) There is a dynamic latent variable model, which describes, for example, the intra-individual (within-level) changes using an autoregressive process of latent variables. (3) Each random within-level parameter is explained by the subject-specific and time-specific random effects (see above). Therefore, the DSEM approach describes heterogeneity of the intra-individual trajectories by inter-individual differences and time-specific effects. Especially, the separation of time-specific and person-specific random effects can be seen as an important extension of single-level dynamic latent variable models.

However, there are also specific aspects of these large frameworks which need consideration: First, the DLCA framework does not separate time- and person-specific random effects. Second, the DSEM framework does not include a Hidden Markov Model for unobserved heterogeneity. Transitions between states can be of importance. For example, they could reflect unobserved heterogeneity in decision processes (e.g., to quit college). Third, neither the DLCA framework nor the DSEM framework include nonlinear effects at the between or within level. Nonlinear effects (such as simple interactions) are part of substantive theories, e.g., expectancy-value theories of motivation propose that high expectancies and high values (as a combination) lead to high motivation.

As a result, it is plausible to combine and to extend both frameworks in order to obtain a comprehensive approach which is capable of (a) intra-individual changes (as a DSEM), (b) inter-individual differences, which have an effect on the individual trajectories, (c) time-specific effects (as random effects), and (d) dynamic latent class memberships, which capture heterogeneity of the trajectories or which can reflect nominal latent variables (such as knowledge mastery). Furthermore, it has been advocated to include (e) flexible nonlinear effects (e.g., splines or interactions) in models (e.g., Marcoulides & Khojasteh, 2018), in order to account for (multiple) complex relationships between the latent variables (e.g., Brandt, Cambria, & Kelava, 2018). To the best of our knowledge, the proposed model is the first multilevel dynamic latent variable framework that unifies these five elements and allows for the specification of nonlinear effects on both the within- and between levels and that unifies these five aspects.

Besides this technical motivation, the NDLC-SEM approach is a natural extension of existing SEM approaches to develop possibilities to test existing theories on dynamic changes in motivation, personality, attitudes, interests, attachment, etc. in more detail (e.g., Fraley, 2002; Krapp, 2002; Pintrich, Marx, & Boyle, 1993; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991).

### Aims of the article

In this article, we first present the nonlinear dynamic latent class structural equation modeling (NDLC-SEM) approach. Second, we give examples of sub-models which are special cases of the more general NDLC-SEM framework. Third, we provide a detailed simulated and an empirical example and explain the estimation of the model using a Bayesian approach. Finally, we discuss the properties of the proposed framework, give recommendations for its application, and highlight general problems of estimating parameters in such comprehensive frameworks.

## THE NONLINEAR DYNAMIC LATENT CLASS STRUCTURAL EQUATION MODEL

The proposed NDLC-SEM contains several submodels. In its comprehensive version, the NDLC-SEM incorporates both inter-individual and time-specific random effects. The individual specific random effects can be used to specify a DLCA (e.g., Asparouhov et al., 2017a), which models latent transitions processes using Markov chain models. A variety of models can be specified as special cases, such as two-level dynamic SEM, dynamic LCA models, single-level dynamic CFA and SEM, state-space models, and time-series models.

We start by decomposing the observed scores. Let $Y_{it}$ be a $(J \times 1)$ dimensional vector of measurements for individual $i$ at time $t$. Note that each individual $i$ is observed at times $t = 1, 2, ..., T_i$ and that $T_i$ might be individual specific (individuals might differ in the number of measurement occasions.). The decomposition in the general NDLC-SEM is as follows:

$$Y_{it} = Y_{1it} + Y_{2i} + Y_{3t} \quad (1)$$

where $Y_{2i}$ and $Y_{3t}$ are $(J \times 1)$ dimensional individual-specific and time-specific parts, respectively. $Y_{1it}$ is the deviation of an individual $i$ at time $t$ from $Y_{2i}$ and $Y_{3t}$. Note that $Y_{1it}$ might depend on a latent state $s$ of an individual $i$ at time $t$, which leads to $[Y_{1it}|S_{it} = s]$ instead of $Y_{1it}$. For example, affective states could be measured repeatedly with different items. $Y_{2i}$ then refers to the individual item score level across time, $Y_{3t}$ refers to the average item score across subjects at each time point $t$, and $Y_{1it}$ is the person- and time-specific deviation of a subject's score relative to the person's and time-related item levels.

In the next subsection, we describe both the individual-specific and time-specific (between-level) models. After this, we describe the within-level model for $Y_{1it}$ and the Markov switching model for the latent states.

## The between-level models

### *Individual-specific component* $Y_{2i}$

The individual-specific component $Y_{2i}$ is explained by a measurement equation and a structural equation ("between-level" in multilevel models). It models the inter-individual differences in the developmental trajectories that are independent of time:

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

$$\eta_{2i} = \alpha_2 + B_2 \eta_{2i} + \Omega_2 h_2(\eta_{2i}) + \Gamma_2 X_{2i} + \zeta_{2i}. \tag{3}$$

$X_{2i}$ is a $(G_2 \times 1)$ dimensional vector of individual-specific, time-invariant (baseline) covariates (e.g., gender). $\eta_{2i}$ is a $(M_2 \times 1)$ dimensional vector of individual-specific, time-invariant latent variables (e.g., representing latent constructs at baseline measure such as baseline depression).

$h_2(\eta_{2i})$ is $R_2$ dimensional vector of functions of $\eta_{2i}$. For example, if interaction and quadratic effects are specified, $h(\eta_{2i}) = vech\,(\eta_{2i}\eta_{2i}')$ is a $(R_2 = M_2(M_2+1)/2 \times 1)$ dimensional vector of products of latent variables. $\epsilon_{2i}$ $(J \times 1)$ and $\zeta_{2i}$ $(M_2 \times 1)$ are multivariate residuals with zero expectations. At this point, we do not assume a specific distribution of the residuals, which should rely on theoretical considerations. In the context of Bayesian estimation, even complicated multivariate distributions can be easily specified. In the frequentist context, practical consideration of available implementations might lead to the assumption of multivariate normal distributions. The same holds for the following equations with residuals. $\nu_2$ $(J \times 1)$, $\Lambda_2$ $(J \times M_2)$, $K_2$ $(J \times G_2)$, $\alpha_2$ $(M_2 \times 1)$, $B_2$ $(M_2 \times M_2)$, $\Omega_2$ $M_2 \times R_2$, and $\Gamma_2$ $(M_2 \times G_2)$ are matrices of fixed effects (including intercepts and coefficients).

$h_2()$ can be used very flexibly such as for the specification of splines that approximate unknown relationships between the latent variables. Then, $h_2$ consists of unidimensional functions $h_{2m}$ $(m = 1, \ldots M_2)$ for each latent predictor $\eta_{2mi}$ that are defined as (see, e.g., Guo, Zhu, Chow, & Ibrahim, 2012):

$$h_{2m}(\eta_{2mi}) = \sum_{\nu_m=1}^{N_m} h^*_{2m\nu_m}(\eta_{2mi}), \tag{4}$$

where $h^*_{2m\nu_m}$ are basis functions with dimension $N_m$, for example, cubic splines with $N_m$ nodes (see details in Hastie, Tibshirani, & Friedman, 2009; Wood, 2017). Note that standard identification rules for splines apply (i.e., with regard to the intercept and linear effects in Equation (3) that need to be considered when including basis expansions, see, Guo et al., 2012).

Further, it is possible to add semiparametric interactions between two predictors using two-dimensional functions $h_{2mm'}, m' > m = 1 \ldots M_2 - 1$ which are operationalized as tensor product bases of the form

$$h_{2mm'}(\eta_{2mi}, \eta_{2mi'}) = \sum_{\nu_m=1}^{N_m} \sum_{\nu_{m'}=1}^{N_{m'}} h^*_{2m\nu_m}(\eta_{2mi}) h^*_{2m'\nu_{m'}}(\eta_{2m'ik}). \tag{5}$$

Splines are a non-parametric regression technique that can be seen as extensions of linear regression models. They have the advantage that they can flexibly adapt to nonlinearities and interactions of variables, even if the functional form of the relationship between the variables is unknown. Compared with the strict parametric functional representation of nonlinear effects (i.e., $\Omega_2 \eta_{2i} \eta_{2i}'$), splines come with less assumptions and can approximate even higher order nonlinear effects by piecewise polynomials. Recently, the extension of SEM to incorporate splines has received increasing attention (e.g., Feng, Wang, Wang, & Song, 2015; Feng, Wu, & Song, 2015, 2017b, Guo et al., 2012; Song, Li, Cai, & Ip, 2013) because they considerably extend the possibilities to model relationships between latent variables.

Note that the measurement model in Equation (2) is conceptualized as linear. By introducing additional restrictions in Equation (3), nonlinear measurement models can be obtained. Again, the same holds true for the following measurement models.

### *Time-specific component* $Y_{3t}$

Similarly, the time-specific component $Y_{3t}$ is explained by a measurement equation and a structural equation. It models the average time trend in the data which are independent of the individuals and do not contain information about inter-individual differences in the growth process:

$$Y_{3t} = \nu_3 + \Lambda_3 \eta_{3t} + K_3 X_{3t} + \epsilon_{3t} \tag{6}$$

$$\eta_{3t} = \alpha_3 + B_3 \eta_{3t} + \Omega_3 h_3(\eta_{3t}) + \Gamma_3 X_{3t} + \zeta_{3t}. \tag{7}$$

$X_{3t}$ is a $(G_3 \times 1)$ dimensional vector of time-specific, individual-invariant/independent covariates (e.g., situation effects such as hospital size where the treatment is conducted or season of the year when the study is conducted). $\eta_{3t}$ is a $(M_3 \times 1)$ dimensional vector of time-specific, individual-invariant latent variables. $h_3(\eta_{3t})$ is a $(R_3 \times 1)$ dimensional vector of (nonlinear) functions for the latent variables (i.e., interactions and quadratic effects). $\epsilon_{3t}$ $(J \times 1)$ and $\zeta_{3t}$ $(M_3 \times 1)$ are multivariate residuals with zero expectations. $\nu_3$ $(J \times 1)$, $\Lambda_3$ $(J \times M_3)$, $K_3$ $(J \times G_3)$, $\alpha_3$ $(M_3 \times 1)$, $B_3$ $(M_3 \times M_3)$, $\Omega_3$ $(M_3 \times R_3)$, and $\Gamma_3$ $(M_3 \times G_3)$ are matrices of fixed effects (including intercepts and coefficients).

## The within-level model

The within-level model is given by the following measurement model and structural model. It models the intra-individual differences occurring during the development process:

$$[Y_{1it}|S_{it} = s] = \nu_{1s} + \Sigma_{l=0}^{L}\Lambda_{1ls}\eta_{1i(t-l)} + \Sigma_{l=0}^{L}R_{ls}Y_{1i(t-l)} + \Sigma_{l=0}^{L}K_{1ls}X_{1i(t-l)} + \epsilon_{1it} \quad (8)$$

$$[\eta_{1it}|S_{it} = s] = \alpha_{1s} + \Sigma_{l=0}^{L}B_{1ls}\eta_{1i(t-l)} + \Sigma_{l=0}^{L}\Sigma_{l'=0}^{L'}\Omega_{1ll's}h_{1ll'}(\eta_{1i(t-l)}, \eta_{1i(t-l')}) + \Sigma_{l=0}^{L}Q_{ls}Y_{1i(t-l)} + \Sigma_{l=0}^{L}\Gamma_{1ls}X_{1i(t-l)} + \zeta_{1it} \quad (9)$$

$X_{1it}$ $(G_1 \times 1)$ dimensional is a vector of individual- and time-dependent covariates (e.g., dosage of psychopharmaka). $\eta_{1it}$ is a $(M_1 \times 1)$ dimensional vector of individual- and time-dependent latent variables (such as a latent factor representing the relative deviance of the depression score compared to average person and time-specific score). Note that $\eta_{1it}$ depends on a latent state $s$ of individual $i$ at time point $t$. $h_{1ll'}(\eta_{1i(t-l)}, \eta_{1i(t-l')})$ is a $(R_1 \times 1)$ dimensional vector of (nonlinear) functions of latent variables at different lags (i.e., interactions and quadratic effects). $\epsilon_{1it}$ $(J \times 1)$ and $\zeta_{1it}$ $(M_1 \times 1)$ are multivariate residuals with zero expectations. $\nu_{1s}$ $(J \times 1)$, $\Lambda_{1ls}$ $(J \times M_1)$, $R_{ls}$ $(J \times J)$, $K_{1ls}$ $(J \times G_1)$, $\alpha_{1s}$ $(M_1 \times 1)$, $B_{1ls}$ $(M_1 \times M_1)$, $\Omega_{1ll's}$ $(M_1 \times R_1)$, and $\Gamma_{1ls}$ $(M_1 \times G_1)$ are matrices with (fixed or random) coefficients. As can be seen, the measurement and structural model are time-series models which include latent classes.

The NDLC-SEM is capable of including observed categorical variables by a parameterization of threshold parameters. For a specific item $j$ with the variable $[Y_{1jit}|S_{it} = s]$ with a variable-specific number of categories $k = 1, ..., m_j$, we assume a normally distributed latent variable $[Y_{1jit}^*|S_{it} = s]$ and threshold parameters $\tau_{j1s}, ..., \tau_{j(m_j-1)s}$ such that:

$$[Y_{1jit} = k|S_{it} = s] \Leftrightarrow \tau_{j(k-1)s} \leq [Y_{1jit}^*|S_{it} = s] < \tau_{jks} \quad (10)$$

with $\tau_{j0s} = -\infty$ and $\tau_{j(m_j)s} = \infty$ for all latent states $s = 1, ..., K$. For reasons of simplicity, we will assume continuous observed variables such that $[Y_{1jit}^*|S_{it} = s] = [Y_{1jit}|S_{it} = s]$ in Equations (8) and (9). Note that in Equation (6) and (7) $[Y_{1it}|S_{it} = s]$ refers to a $J \times 1$ vector of items and that for an item $j$, $[Y_{1jit}|S_{it} = s]$ is a specific element of this vector $[Y_{1it}|S_{it} = s]$.

## The Markov switching model

The latent state variable $S_{it}$ is a person- and time-dependent variable which follows a Markov switching model with person- and time-specific transition probability:

$$P(S_{it} = d|S_{i(t-1)} = c) = \frac{\exp(\alpha_{itdc})}{\Sigma_{k=1}^{K}\exp(\alpha_{itkc})} \quad (11)$$

$\alpha_{itdc}$ are person- and time-specific random effects (see below) with $\alpha_{itKc} = 0$. For example, the latent state could represent the (time- and person-specific) adherence to the treatment regimen. Often, this variable is an unobserved (or unobservable) entity that nonetheless effects the development of the outcome variable (e.g., a decline in depression for persons that are adherent and a different growth pattern for those non-adherent).

## Random effects

At the within level, we allow for random effects (loadings, intercepts, and slopes). Any random within-level parameter $p_{it}$ (e.g., elements from $\nu_{1s}$, $\Lambda_{1ls}$ etc.) can be decomposed as

$$p_{it} = p_{2i} + p_{3t} \quad (12)$$

where $p_{2i}$ is a subject-specific random effect which is an element of vector $\eta_{2i}$ in the between-level model. $p_{3t}$ is an time-specific random effect which is an element of vector $\eta_{3t}$.

If we introduce the indices $i$ and $t$ for the parameters in the within-level models in Equations (8) and (9), we get the following (generalized) measurement and structural models:

$$[Y_{1it}|S_{it} = s] = \nu_{1s} + \Sigma_{l=0}^{L}\Lambda_{1ilts}\eta_{1i(t-l)} + \Sigma_{l=0}^{L}R_{ilts}Y_{1i(t-l)} + \Sigma_{l=0}^{L}K_{1ilts}X_{1i(t-l)} + \epsilon_{1it} \quad (13)$$

$$[\eta_{it}|S_{it}=s]=\alpha_{1s}+\Sigma_{l=0}^{L}B_{1ilts}\eta_{1i(t-l)} + \Sigma_{l=0}^{L}\Sigma_{l'=0}^{L'}\Omega_{1ill'ts}h_{1ll'}(\eta_{1i(t-l)},\eta_{1i(t-l')}) + \Sigma_{l=0}^{L}Q_{ilts}Y_{1i(t-l)}+\Sigma_{l=0}^{L}\Gamma_{1ilts}X_{1i(t-l)}+\zeta_{1it} \quad (14)$$

Every random parameter is decomposed as in Equation (12).

Furthermore, we allow residual variances $v_{it}$ on the within level to be random effects. This means that elements (i.e., parameters) of the matrices $Var(\epsilon_{1it})$ and $Var(\zeta_{1it})$ can be random: This allows researchers, for example, to model heteroscedasticity that is due to person- or time-specific sources (that are not represented by observed variables on the respective levels).

$$v_{it}=\exp(p_{2i}+p_{3t}) \quad (15)$$

Again, $p_{2i}$ is a subject-specific random effect, which is an element of vector $\eta_{2i}$ in the between-level model. $p_{3t}$ is a time-specific random effect, which is an element of vector $\eta_{3t}$.

Note that in the structural models from above interactions and quadratic effects have been formulated. Alternatively, it is plausible to use semiparametric spline effects, if the nonlinear structural relationship is unknown (e.g., Marcoulides & Khojasteh, 2018).

## Continuous time models

The proposed NDLC-SEM applies to empirical research where the time $t$ strictly follows a discrete variable, i.e., each person is observed at time $1,2,...T_i$. In applications with daily diaries or daily measurements with smart phones which follow a strict pattern, this is a plausible assumption. In cases, when observations follow an irregular pattern, assumptions of discrete time models are easily violated. The time variable is then no longer discrete, but real valued. This leads to biased estimates if discrete models are applied without accounting for this problem.

There are two ways to address this problem. First, it can be addressed by a simple procedure which is described in (see Appendix A in Asparouhov et al., 2017b) in detail (for the concept of phantom variables, see also Rindskopf, 1984). This procedure involves a transformation of the time variable by "rescaling, shifting, and rounding." As a result of the transformation, the real-valued time variable is approximated by a discrete time variable. Essentially, the time variable is multiplied by $1/\delta$ (where $\delta>0$ is a small number). The result is rounded up to the nearest integer. The data are filled with missing values for those integer time values that were not the nearest for an observed continuous time point, which is not a problematic assumption (according to Asparouhov et al., 2017b). After this procedure, the discrete dynamic model can be used again. The authors also provide an algorithm to obtain an optimal $\delta$ and explain how to transform the time variable when cross-classified models are used. The time shift transformation aligns the time scale between individuals, when time-specific effects are estimated.

Second, this problem can be addressed by representing the models as continuous time models which have a long tradition in several disciplines (for an introduction to the application in the SEM context, see Voelkle, Oud, Davidov, & Schmidt, 2012). Continuous time models are represented by stochastic differential equations and directly adapt for unequal and irregular intervals of measurements. In principle, an implementation of these models in this framework can be conducted by using nonlinear parameter constraints that relate parameters describing continuous change to discrete observations (e.g., a drift matrix or a continuous time error process matrix). For linear models, this process is straightforward and the matrix $B_{1ilts}$ that includes the autoregressive and cross-lagged effects of the latent variables can be used to estimate a drift matrix $B^{*}_{1ilts}$ (using constraints as proposed on pp. 183–184 in Voelkle et al., 2012). An extension to the proposed nonlinear modeling framework needs further technical work that is beyond the scope of this article (e.g., additional nonlinear constraints need to be derived for $\Omega_{1ll'ts}$).

# SUBMODELS OF THE NDLC-SEM

In this section, we will give two brief examples as special cases of the NDLC-SEM framework, which might be interesting from an applied perspective. First, we will describe the two-level nonlinear DLCA. Second, we will describe a semiparametric DSEM.

## Two-level nonlinear dynamic latent class analysis

In situations, when intra-individual processes/trajectories are examined, inter-individual differences (such as traits, dispositions, etc.) might be used to describe unobserved heterogeneity and transitions between latent states. In the following, a two-level nonlinear DLCA is proposed. We start with the within-level description.

### *Within level*

In the within level, the measurement and structural models are given as:

$$([Y_{1it}|S_{it}=s]=\nu_{1s}+\Sigma_{l=0}^{L}\Lambda_{1ls}\eta_{1i(t-l)}+\epsilon_{1it} \quad (16)$$

$$[\eta_{1it}|S_{it}=s]=\alpha_{1s}+\Sigma_{l=0}^{L}B_{1ls}\eta_{1i(t-l)} + \Sigma_{l=0}^{L}\Omega_{1ls}vech(\eta'_{1i(t-l)}\eta_{1i(t-l)}) + \Gamma_{1s}X_{1it}+\zeta_{1it} \quad (17)$$

In Equations (16) and (17), lags of observed variables/covariates (e.g., $Y_{1i(t-l)}$ and $X_{1i(t-l)}$) are omitted, but could be included (cp. Equations (8) and (9)).

### Transition probabilities

The unobserved heterogeneity in the trajectories is described in the Markov switching model from Equation (11). This means that person-specific individual characteristics, which are summarized by a random effect $\alpha_{idc}$, determine the probability of switching between states or stay within a specific state (i.e., $P(S_{it} = d|S_{i(t-1)} = c)$).

The latent states create class-specific relationships which are described in Equations (16) and (17). The random effect $\alpha_{idc}$ is part of the between-level model (i.e., $\eta_{2i}$).

### Between-level

In the nonlinear DLCA, the between-level model is reduced to a person-specific measurement and structural model (cp. Equations (2) and (3)):

$$Y_{2i} = \nu_2 + \Lambda_2\eta_{2i} + \epsilon_{2i} \quad (18)$$

$$\eta_{2i} = \alpha_2 + B_2\eta_{2i} + \Omega_2 vech(\eta_{2i}\eta_{2i}') + \Gamma_2 X_{2i} + \zeta_{2i} \quad (19)$$

As can be seen, in the measurement model the covariate $X_{2i}$ was omitted, but could be included. Furthermore, there could be (theoretical or exploratory) reasons to include semiparametric effects, for example in the structural model from Equation (19), instead of interactions. For example, in studies with a large number of scales/covariates, which were measured at the beginning of the study, an unknown semiparametric relationship could be examined by formulating a smooth function using a bases expansion (for cross-sectional SEM, see, e.g., Guo et al., 2012).

A simplified version of this model is depicted in Figure 1. This model is based on a dynamic extension of work in the area of achievement research (e.g., Wigfield et al., 1991). On the within level, $\eta_{11}$ could represent achievement at different time points and $\eta_{12}$ could represent effort at this time point. Both variables predict their future score. In addition, both effort at time $t$ predicts future achievement at time $t+1$; this effect is moderated by achievement at time $t$ (represented by $\omega_{112}$), that is, the more effective the effort is, the higher the achievement is at time $t$. The latent class $S$ could represent a state of cognitive mastery of the area (i.e., if one masters the area, effort does not have an effect at all on achievement).

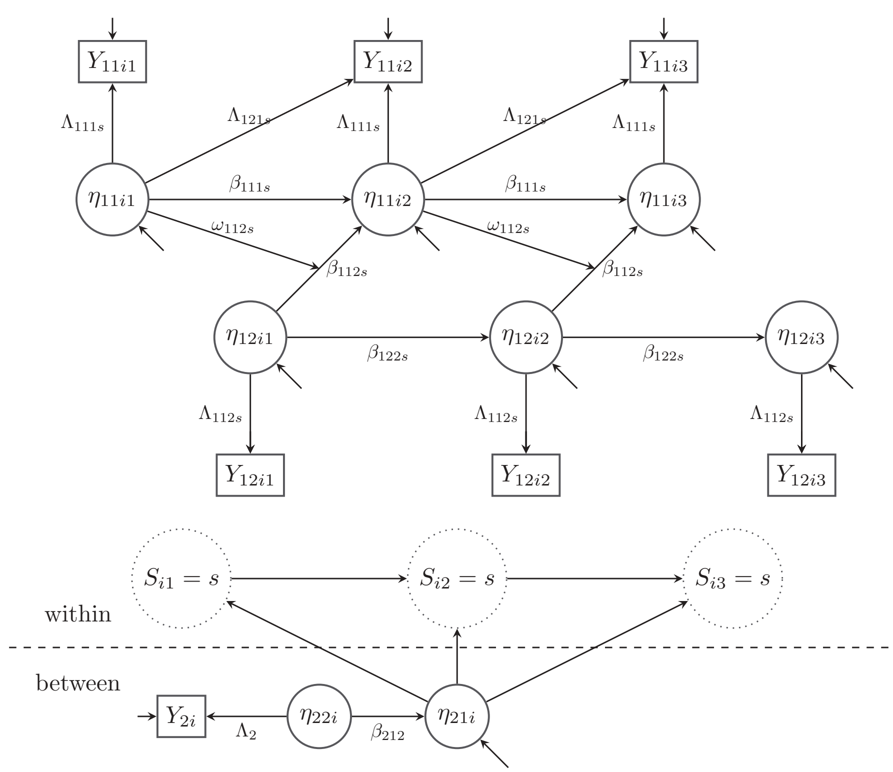


FIGURE 1 Path diagram for a two-level nonlinear model for the first three measurement occasions. On the within level, two factors are specified where the first factor has a lag 0 and lag 1 measurement model. The lag 1 structural model includes an interaction effect (see text for more details).

### Single-level semiparametric dynamic structural equation model

In the single-level semiparametric DSEM, a smooth relationship between the lagged latent variables is allowed. This makes the autoregressive structural model capable of more flexible lagged effects.

The measurement and structural model can be described as follows:

$$Y_{1it} = \nu_1 + \Sigma_{l=0}^{L} \Lambda_{1l} \eta_{1i(t-l)} + \epsilon_{1it} \quad (20)$$

$$\eta_{1it} = \alpha_1 + \Sigma_{l=0}^{L} \Omega_{1l} h_{1l}(\eta_{1i(t-l)}) + \Gamma_1 X_{1it} + \zeta_{1it} \quad (21)$$

In Equation (20), the lagged effects can be simplified, such that only a traditional factor analytic relationship between the measured variable and the latent variable is allowed. In Equation (21), $h_{1l}$ describes a smooth function related to lag $l$, which contains basis expansions of the lagged multivariate latent variable $\eta_{1i(t-l)}$ (e.g., a regression spline function). In this way, it is possible to obtain the nonlinear lagged effects of the latent variable $\eta_{1i(t-l)}$, instead of simple linear effects. A simplified version of this model for $L = 1$ is depicted in Figure 2.

### Different time-series models on the within level

Discrete time models for intensive repeated measures include autoregressive (AR) and moving average (MA) models, as well as their combinations such as ARIMA models. Within the proposed framework, each of these models can be specified. Which model is appropriate depends on the research questions, and on the process investigated. For example, AR and MA models assume that there is no mean drift over time (stationarity), whereas the ARIMA model can be used to model such drifts. AR and MA models describe different patterns of autocorrelation in the data.

For example, a simple ARIMA(1,1,0) (differenced first-order autoregressive model) for a single latent variable can be specified by $h_{ill'} = (\eta_{1i(t-l)} - \eta_{1i(t-l')})$ from Equation (9):

$$\eta_{1it} = \alpha_1 + \eta_{1i(t-1)} + \Omega_{112}(\eta_{1i(t-1)} - \eta_{1i(t-2)}) + \zeta_{1it} \quad (22)$$

with $B_{11s} = 1$. We illustrate this model in the empirical example section in more detail.

## EXAMPLE—COLLEGE DROP-OUT

In this section, we give a simulated example of a model, which is suitable to describe unobserved intra-individual processes (i.e., changes between states). These latent changes between states are driven by inter-individual differences (e.g., traits). We will assume $N = 500$ college students of mathematics and explain their intention to quit their studies (i.e., college drop-out). This intention (to quit vs. to stay) can be assumed to be a Markov process. The individual-specific transition probabilities between the states of this process can be explained by a random effect on the between level. The between-level model assumes inter-individual differences which have an effect on the unobserved intention. For example, some vulnerability factors (e.g., lack of conscientiousness and neuroticism) could increase the risk of an intention to quit.[1] Some factors could have a synergistic dysfunctional consequence if their joint occurrence has very adverse effects. This synergy is then a nonlinear effect (of the latent vulnerability variables) on the transition probability. Furthermore, daily affective states could be measured on the intra-individual level.

We assume that the population model on the *between level* is given as:

$$\eta_{21i} = 3.85 + .5 \cdot \eta_{22i} + .5 \cdot \eta_{23i} + .5 \cdot \eta_{22i} \cdot \eta_{23i} + \zeta_{21i} \quad (23)$$

$$Y_{2i} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \eta_{22i} \\ \eta_{23i} \end{bmatrix} + \begin{bmatrix} \epsilon_{21i} \\ \epsilon_{22i} \\ \epsilon_{23i} \\ \epsilon_{24i} \\ \epsilon_{25i} \\ \epsilon_{26i} \end{bmatrix} \quad (24)$$

where $\eta_{22}$ and $\eta_{23}$ are standard normally distributed with a correlation of .3. $\zeta_{21}$ is normally distributed with zero expectation and fixed variance of $\pi^2/3$. Note that the random effect $\eta_{21}$ has no measurement model. $\eta_{22}$ and $\eta_{23}$ (e.g., conscientiousness and neuroticism) have three indicators each. The residuals $\epsilon_{21}, \epsilon_{22}, \ldots, \epsilon_{26}$ are standard normally distributed and uncorrelated and have variances of .25.

$\eta_{21}$ defines the random effect, which describes the *transition probabilities* for the states $S_{it} = s$ with $(s = 1, 2)$:

$$P_{11i} = \frac{\exp(\eta_{21i})}{\exp(\eta_{21i}) + 1}, \quad P_{12i} = \frac{1}{\exp(\eta_{21i}) + 1}, \quad P_{21i} = 0,\ P_{22i} = 1 \quad (25)$$

For persons who have no intention to quit at time point $t$, we assume that the probability to maintain this intention at time point $t + 1$ is $P_{11}$ and to change the intention (i.e., to quit) is $P_{12}$. For simplicity, we assume that the transition probabilities to reverse the decision and to continue studies

[1] Note that there are elaborated theories on college drop-out that deal with several types of risk factors (e.g., Bean, 2005; Burrus et al., 2013; Tinto, 1993). These risk factors also include variables which are not just part of the personality of the students, but of their circumstances of life, or institutional characteristics and many more.

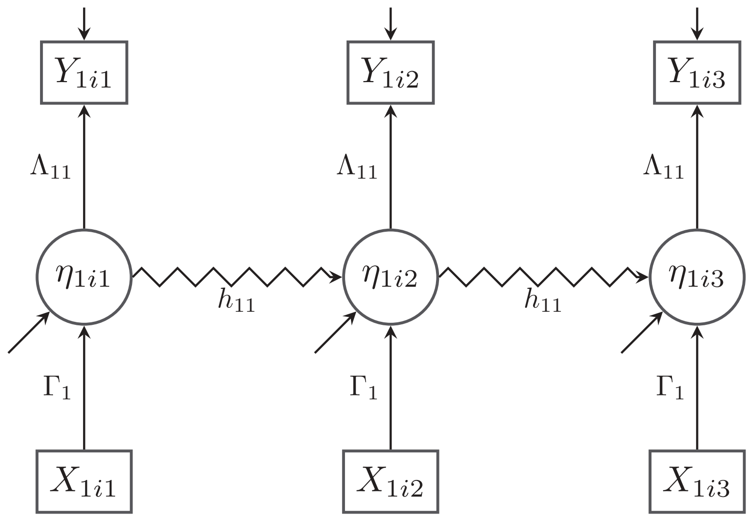


FIGURE 2 Path diagram for a single-level semiparametric model for the first three measurement occasions. Here, we simplified the model by using only a lag 1 structural model and a lag 0 measurement model which assumed that variables only loaded on the latent factors at the same time point.

at time point $t+1$ after having decided to quit at time point $t$ is equal to $P_{21}=0$; as a consequence, the probability to maintain the decision to quit is equal to $P_{22}=1$.

We assume that we have 50 time points (measurement occasions). The population model on the *within level* is given as:

$$[\eta_{1it}|S_{it}=s] = .5\cdot\eta_{1i(t-1)} + .2\cdot\eta_{1i(t-2)} + \zeta_{1it} \tag{26}$$

$$[Y_{1it}|S_{it}\ = s] = 1\cdot\eta_{1it} + \epsilon_{1it} \tag{27}$$

where $\zeta_{1t}$ is normally distributed with unit variance for both states. The expectation of $\zeta_{1t}$ is 0 in the first state and 2 in the second state. $Y_1$ represents three indicators. The three residuals in $\epsilon_{1t}$ are multivariate normally distributed and uncorrelated with unit variances. The process modeled is an auto-regressive AR(2) process.

The complete model is depicted in Figure 3.

## Estimation

### *Between level*

The following distributions for the variables were specified at the between level of the model (see Eqs. (23) and (24)):

$$Y_{2ji}\sim N(\mu_{Y_{2ji}},\sigma^2_{\epsilon_{2j}}),\quad \text{for } j=1,...,6 \tag{28}$$

with

$$\mu_{Y_{21i}}=\eta_{22i},\quad \mu_{Y_{22i}}=\lambda_{22}\eta_{22i},\quad \mu_{Y_{23i}}=\lambda_{23}\eta_{22i} \tag{29}$$

$$\mu_{Y_{24i}}=\eta_{23i},\quad \mu_{Y_{25i}}=\lambda_{25}\eta_{23i},\quad \mu_{Y_{26i}}=\lambda_{26}\eta_{23i} \tag{30}$$

and

$$(\eta_{22i},\eta_{23i})'\sim N(\mathbf{0},\Phi_2) \tag{31}$$

The person-specific random effect was specified as follows:

$$\eta_{21i}\sim N(\mu_{\eta_{21i}},\sigma^2_{\zeta_{21}}) \tag{32}$$

$$\mu_{\eta_{21i}}=\beta_{20}+\beta_{21}\eta_{22i}+\beta_{22}\eta_{23i}+\omega_{223}\eta_{22i}\eta_{23i} \tag{33}$$

The latent class membership ($t=2\ldots T$) was implemented using Equation (25). At time $t=1$, all persons were assumed to be in state $S=1$.

### *Within level*

At the within level, the structural model was time-/state-dependent:

$$\eta_{1it}\sim N(\mu_{\eta_{1it}},\sigma^2_{\zeta_{1s}}) \tag{34}$$

$$\mu_{\eta_{1it}}=\alpha_{1s}+\beta_{11}\eta_{1i(t-1)}+\beta_{12}\eta_{1i(t-2)} \tag{35}$$

The measurement model was given as:

$$Y_{1jit}\sim N(\mu_{Y_{1jit}},\sigma^2_{\epsilon_{1j}}),\quad \text{for } j=1,..,3 \tag{36}$$

$$\mu_{Y_{11it}}=\eta_{1it},\quad \mu_{Y_{12it}}=\lambda_{12}\eta_{1it},\quad \mu_{Y_{13it}}=\lambda_{13}\eta_{1it} \tag{37}$$

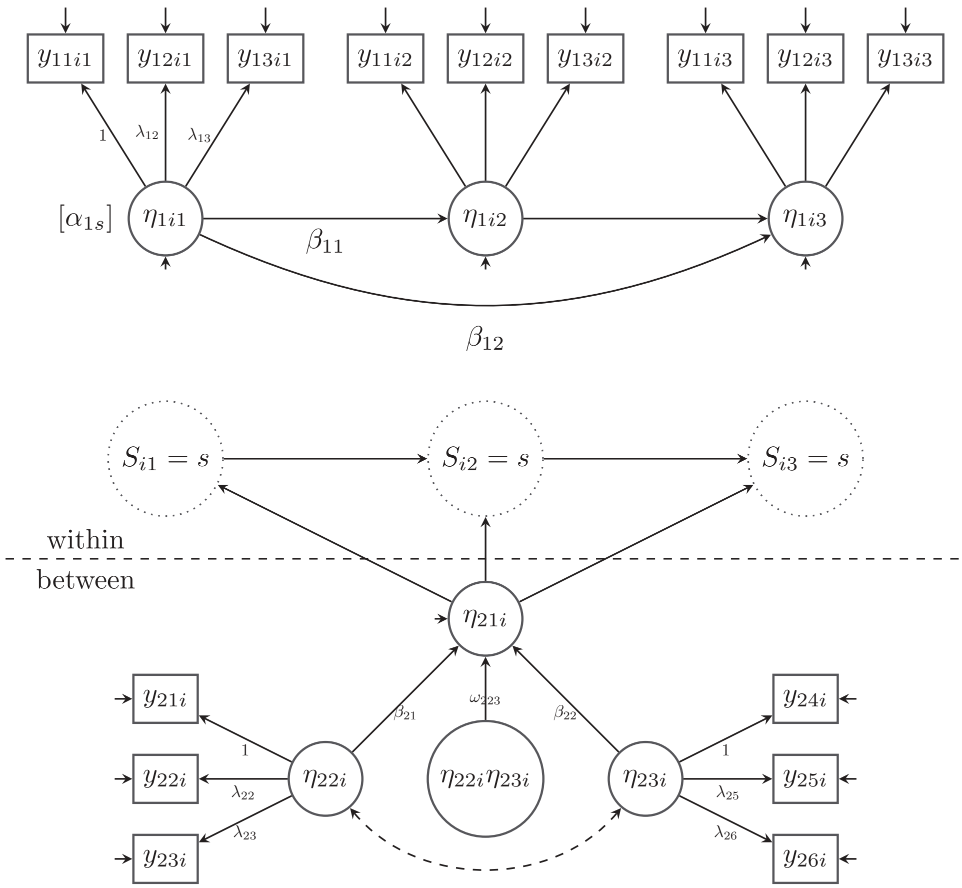


FIGURE 3 Path diagram for the simulated example (college student drop out) for the first three measurement occasions. On the within level, an AR(2) process with two states $S_{it} = 1, 2$ is assumed for one latent variable with three indicators. On the between level, a nonlinear model with two latent factors and one interaction is modeled.

### Prior distributions

Priors were chosen as weakly informative for the measurement models and for the level 1 auto-regressive coefficients

$$\lambda_{1j} \sim N(1, 1), \text{ for } j = 2, 3 \tag{38}$$

$$\lambda_{2j} \sim N(1, 1), \text{ for } j = 2, 3, 5, 6 \tag{39}$$

$$\beta_{1l} \sim N(0, 1), \text{ for } l = 1, 2 \tag{40}$$

or uninformative for the level 2 structural model and the level 1 intercepts

$$\beta_{20} \sim unif(2, 6) \tag{41}$$

$$\beta_{2p} \sim unif(0, 1), \text{ for } p = 1, 2 \tag{42}$$

$$\omega_{223} \sim unif(0, 1) \tag{43}$$

$$\alpha_{11} = 0 \tag{44}$$

$$\alpha_{12} \sim unif(0, 4) \tag{45}$$

where the constraint $\alpha_{11} = 0$ was necessary for model identification. Note that this constraint always holds in this model if data are rescaled by $Y^c_{1jit} = Y_{1jit} - \bar{Y}_{111}$ because $Y_{11i1} = \eta_{1i1}$ and all persons are in state $S_{i1} = 1$ at the first measurement occasion.

Standard priors were chosen for the precisions as

$$\sigma^{-2}_{\epsilon_{1j}} \sim Gamma(9, 4), \text{ for } j = 1 \ldots 3 \tag{46}$$

$$\sigma^{-2}_{\epsilon_{2j}} \sim Gamma(9, 4), \text{ for } j = 1 \ldots 6 \tag{47}$$

$$\sigma_{\zeta_{1s}}^{-2} \sim Gamma(9,4), \text{ for } s = 1,2 \quad (48)$$

$$\sigma_{\zeta_{21}}^{-2} \sim Gamma(9,4) \quad (49)$$

$$\Phi_2^{-1} \sim Wishart(\Phi_0^{-1}, 4) \quad (50)$$

where $\Phi_0$ was a $2 \times 2$ identity matrix.

### Implementation

The model was implemented in Jags (Plummer, 2003) using R2jags (Su & Yajima, 2015) in R (R Core Team, 2016). Three chains with 4000 iterations were run, from which the first 2000 iterations were discarded as burn-in. Convergence was checked by investigating within and between chain variability graphically, and by using the Rhat statistic with a rule of thumb that the model converged when $Rhat \leq 1.1$ for all parameters.

Code for the model can be obtained by the first author or the second author's website.

## Results

In Table 1, estimation results for the simulated example are presented. All coefficients indicated convergence with a maximal Rhat of 1.10 (for $\alpha_{12}$). Percentile confidence intervals included the population values for all coefficients. Between-level coefficients' standard deviations (SDs) were considerably larger than those from the within level as expected.

Figure 4 illustrates the estimated state membership over time (gray lines are those persons in the population that switched to state 2). Ninety-eight percent of the switchers and non-switchers were correctly identified at the last measurement occasion $T$ ($\chi^2 = 456.94, df = 1, p < .0001$). This ratio was rather constant over time with an average correct

TABLE 1
Parameter Estimates for the Example (Generated Data for College Student Drop Out)

| | *Mean* | SD | *2.5%* | *50%* | *97.5%* | *Rhat* |
|---|---|---|---|---|---|---|
| $\alpha_{11}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| $\alpha_{12}$ | 1.96 | 0.05 | 1.86 | 1.96 | 2.05 | 1.10 |
| $b_{20}$ | 4.06 | 0.12 | 3.82 | 4.05 | 4.31 | 1.02 |
| $b_{21}$ | 0.62 | 0.14 | 0.35 | 0.62 | 0.88 | 1.01 |
| $b_{22}$ | 0.72 | 0.13 | 0.47 | 0.72 | 0.96 | 1.02 |
| $\omega_{223}$ | 0.65 | 0.14 | 0.38 | 0.65 | 0.92 | 1.00 |
| $\lambda_{12}$ | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.01 |
| $\lambda_{13}$ | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.01 |
| $\lambda_{22}$ | 1.03 | 0.04 | 0.95 | 1.03 | 1.11 | 1.00 |
| $\lambda_{23}$ | 1.00 | 0.04 | 0.93 | 1.00 | 1.08 | 1.01 |
| $\lambda_{25}$ | 0.98 | 0.04 | 0.91 | 0.98 | 1.06 | 1.01 |
| $\lambda_{26}$ | 1.03 | 0.04 | 0.95 | 1.03 | 1.11 | 1.00 |
| $b_{11}$ | 0.51 | 0.01 | 0.49 | 0.51 | 0.53 | 1.07 |
| $b_{12}$ | 0.20 | 0.01 | 0.18 | 0.20 | 0.21 | 1.03 |
| $\sigma^2_{\zeta_{11}}$ | 1.00 | 0.02 | 0.97 | 1.00 | 1.04 | 1.03 |
| $\sigma^2_{\zeta_{12}}$ | 0.97 | 0.02 | 0.93 | 0.97 | 1.01 | 1.01 |
| $\sigma^2_{\zeta_{21}}$ | 2.85 | 0.57 | 1.84 | 2.81 | 4.06 | 1.03 |
| $\phi_{211}$ | 0.91 | 0.07 | 0.78 | 0.91 | 1.06 | 1.01 |
| $\phi_{212}$ | 0.32 | 0.05 | 0.22 | 0.32 | 0.43 | 1.00 |
| $\phi_{221}$ | 0.32 | 0.05 | 0.22 | 0.32 | 0.43 | 1.00 |
| $\phi_{222}$ | 1.06 | 0.09 | 0.90 | 1.06 | 1.25 | 1.00 |
| $\sigma^2_{\epsilon_{11}}$ | 1.01 | 0.01 | 0.98 | 1.01 | 1.03 | 1.00 |
| $\sigma^2_{\epsilon_{12}}$ | 1.00 | 0.01 | 0.98 | 1.00 | 1.02 | 1.00 |
| $\sigma^2_{\epsilon_{13}}$ | 1.01 | 0.01 | 0.99 | 1.01 | 1.03 | 1.00 |
| $\sigma^2_{\epsilon_{21}}$ | 0.27 | 0.02 | 0.22 | 0.26 | 0.32 | 1.00 |
| $\sigma^2_{\epsilon_{22}}$ | 0.30 | 0.03 | 0.25 | 0.30 | 0.36 | 1.01 |
| $\sigma^2_{\epsilon_{23}}$ | 0.27 | 0.03 | 0.23 | 0.27 | 0.32 | 1.00 |
| $\sigma^2_{\epsilon_{24}}$ | 0.30 | 0.03 | 0.25 | 0.30 | 0.36 | 1.01 |
| $\sigma^2_{\epsilon_{25}}$ | 0.29 | 0.03 | 0.24 | 0.29 | 0.34 | 1.01 |
| $\sigma^2_{\epsilon_{26}}$ | 0.30 | 0.03 | 0.24 | 0.30 | 0.36 | 1.01 |

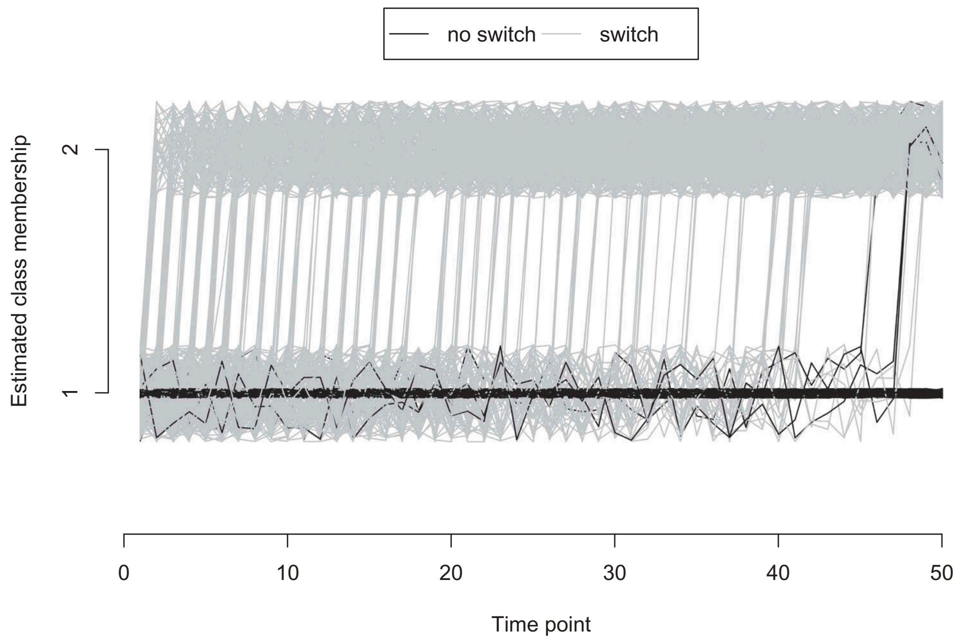


FIGURE 4 Estimated dynamic class membership for switchers (gray) and non-switchers (black).

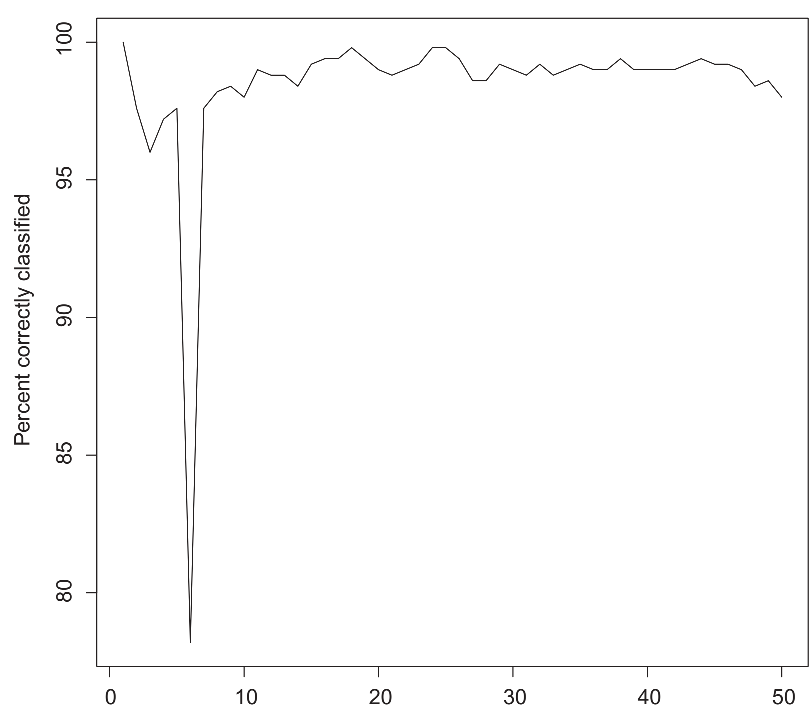


FIGURE 5 Estimated dynamic class membership for switchers (gray) and non-switchers (black).

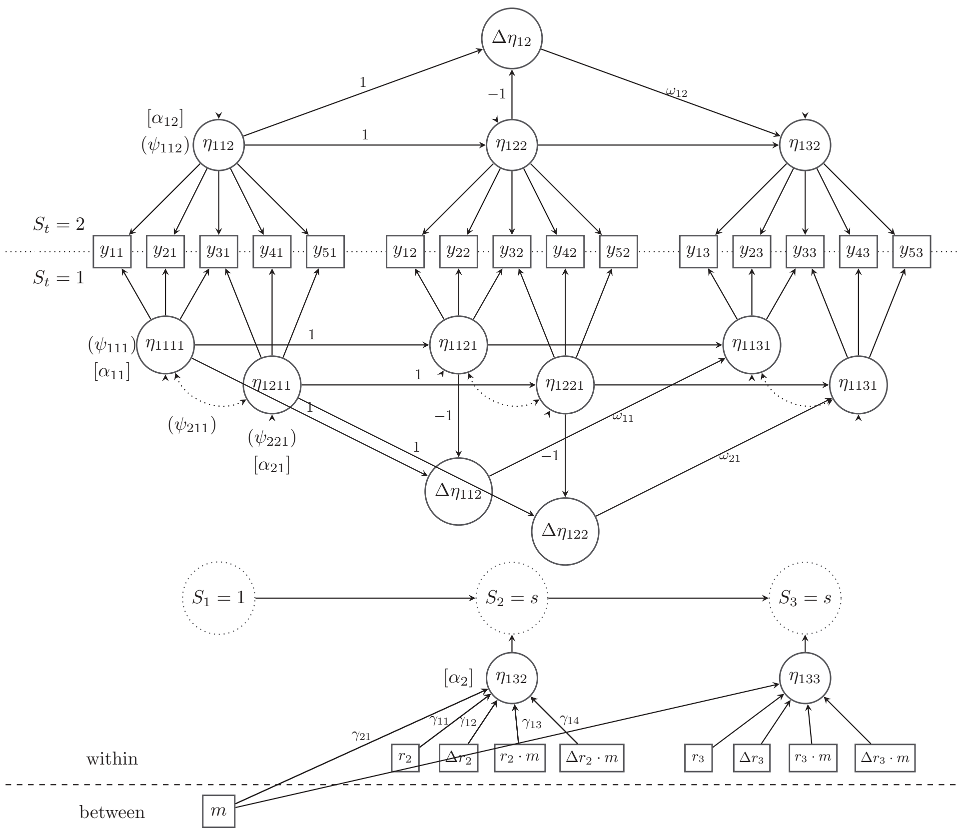


FIGURE 6 Path diagram and estimates for the empirical example for the first three measurement occasions. $\eta_{13t}$ is the within-level (person- and time-specific) latent variable predicting class membership via reading skills ($r_t$) and reading skill change ($\Delta r_t = r_t - r_{t-1}$), fine motor skills at baseline ($m$), and their interactions. On the within level, the two constructs for $S = 1$ are basic and abstract math skills, in $S = 2$ these math skill facets are represented by a single math skill level that students achieve at some time point during the study. The math development is modeled using an ARIMA(1,1,0) model.

classification of 98.4% (cf. Figure 5). The switching points for the states and their estimates correlated at .99.

## EXAMPLE—DEVELOPMENT OF MATH SKILLS

Next, we will present an empirical example to illustrate the capabilities to model complex hypotheses in this framework. The (real) data are taken from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 (ECLS-K; Tourangeau, Nord, Lê, Pollack, & Atkins-Burnett, 2009). In this study, students were measured at seven measurement occasions from kindergarten to grade 8. Here, we use a random subsample of 500 students that were measured at each of the seven measurement occasions using math and reading items. The focus lies in the development of math skills for which we chose five out of nine scales (ordinality/sequence, add/subtract, multiply/divide, place value, rate & measurement)[2] because the remaining scales were either to simple at the end of measurement (all students had 100% correct answers) or to complicated at the beginning (all students had 0% correct answers).

[2] Coded with C*t*R4MPB3 to C*t*R4MPB7 where *t* indicates the time point.

TABLE 2
Results for Structural ARIMA Model for the Empirical Example (ECLS-K). Parameterl Labels According to Figure 6

| | *Mean* | SD | *2.5%* | *50%* | *97.5%* | *Rhat* |
|---|---|---|---|---|---|---|
| $\alpha_{111}$ | 1.56 | 0.05 | 1.47 | 1.56 | 1.65 | 1.00 |
| $\alpha_{121}$ | −0.76 | 0.07 | −0.89 | −0.76 | −0.63 | 1.00 |
| $\alpha_{112}$ | 1.66 | 0.02 | 1.62 | 1.66 | 1.70 | 1.00 |
| $\omega_{11}$ | −0.24 | 0.02 | −0.29 | −0.24 | −0.19 | 1.00 |
| $\omega_{21}$ | 0.85 | 0.04 | 0.76 | 0.85 | 0.93 | 1.00 |
| $\omega_{12}$ | −0.73 | 0.02 | −0.78 | −0.74 | −0.68 | 1.00 |
| $\alpha_2$ | 2.37 | 0.52 | 1.34 | 2.37 | 3.38 | 1.01 |
| $\gamma_{11}$ | −2.31 | 0.65 | −3.61 | −2.30 | −1.04 | 1.01 |
| $\gamma_{12}$ | −0.33 | 0.71 | −1.72 | −0.34 | 1.09 | 1.00 |
| $\gamma_{21}$ | 0.30 | 0.10 | 0.12 | 0.30 | 0.50 | 1.01 |
| $\gamma_{13}$ | −0.49 | 0.12 | −0.73 | −0.49 | −0.26 | 1.01 |
| $\gamma_{14}$ | −0.57 | 0.12 | −0.82 | −0.57 | −0.33 | 1.00 |
| $\psi_{111}$ | 3.62 | 0.13 | 3.36 | 3.62 | 3.88 | 1.00 |
| $\psi_{211}$ | 5.09 | 0.18 | 4.75 | 5.09 | 5.45 | 1.00 |
| $\psi_{221}$ | 10.85 | 0.32 | 10.24 | 10.85 | 11.48 | 1.00 |
| $\psi_{112}$ | 0.17 | 0.01 | 0.15 | 0.17 | 0.20 | 1.01 |

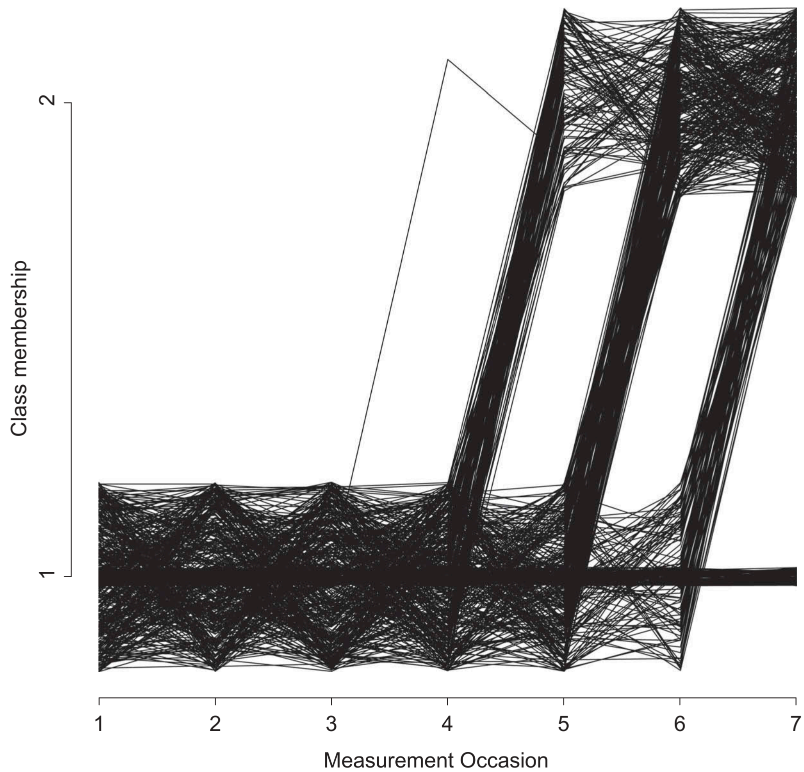


FIGURE 7 Estimated dynamic class membership for the students (the jittering is only imposed to illustrate individual lines).

Additional information were used as fine motor skills at the initial measurement occasion. Time-specific reading skills were used as observed covariates (based on the average of 10 scales measuring reading skills[3]). Math and reading skills were available as percentage of correct items and were logit-transformed before analysis.

In this example, we hypothesize that a nominal change in the math skill constructs occurs over time. Whereas at the beginning two constructs (concrete and abstract math) are necessary to describe the performance in the five different scales, only one construct for math skills will be sufficient at a later time point (all hypotheses are illustrated in the path diagram in Figure 6). This nominal change is operationalized as the latent state variable $S_{it}$ which represents a cognitive state that we want to name math mastery: Students in state $S_{it} = 1$ have not yet mastered the math test at time $t$, they have lower scores in the constructs, and their capabilities in concrete and abstract math questions are represented by two (correlated) constructs. Students in state $S_{it} = 2$ mastered the math test at time $t$ and there is no differentiation in their capabilities of concrete and abstract math skills necessary. For students in state 1, a switch to state 2 can be predicted by the person- and time-specific reading skills and the individual change from the last to each respective time point. The second part models the hypothesis that persons who have a strong increase from one time point to the next in reading skills will also be more likely to master math. Fine motor skills in kindergarten (such as holding a pen or using a scissor) were also assumed to predict a switch from state 1 to 2 because early fine motor skills are related to abstract thinking (Luo, Jose, Huntsinger, & Pigott, 2007). We assume that increasing reading skills can compensate low fine motor skills, which implies a nonlinear interaction effect between these variables. For the development of each factor of math skills $\eta_{1kit}$, we assume an ARIMA (1,1,0) model:

$$\begin{aligned}
[\eta_{1ki1}|S_{i1} = s] &= \alpha_{1ks} + \zeta_{1ki1} \\
[\eta_{1ki1}|S_{i2} = s] &= \alpha_{1ks} + \eta_{1ki1} + \zeta_{1ki2} \\
[\eta_{1ki1}|S_{i} = s] &= \alpha_{1ks} + \eta_{1ki(t-1)} + \omega_{ks}\underbrace{(\eta_{1ki(t-1)} - \eta_{1ki(t-2)})}_{\Delta\eta_{1ki(t-1)s}} \\
&\quad + \zeta_{1kit} \text{ for } t > 2
\end{aligned} \tag{51}$$

[3] $CtR4RPB1$ to $CtR4RPB10$.

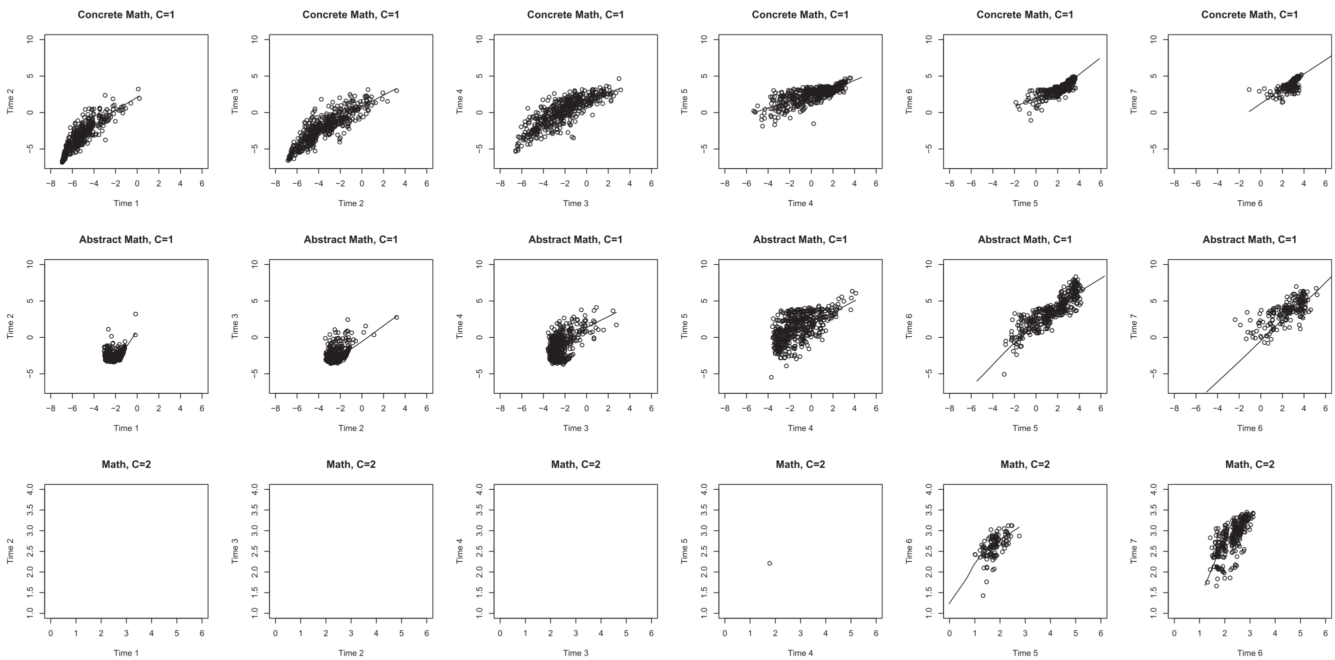


FIGURE 8 Scatter plots for each measurement occasion and factor. Persons start to switch to class 2 at time point 5 (3rd grade). For illustration purposes, factor scores were centered.

with $\Psi_{1kts} = \Psi_{1ks}$ for all $t$. Measurement models were assumed to be time invariant.

For estimation, four chains each with 16,000 iterations were run in Jags with 8,000 burn-in iterations. Details on prior selection can be found in the appendix (generally speaking, we used weakly informative priors on all coefficients in a similar way as in the simulated example presented above).

The results are presented in Table 2 and are illustrated in Figures 7 and 8. Parameter estimates converged with an Rhat below 1.10 for each parameter. All coefficients could be considered significant with 95% probability intervals not including zero with the exception of the effect of $\Delta r_t$ on $\eta_{2t}$. $\beta$ was negative for concrete math in state 1 and math skills in state 2, indicating a ceiling effect over time. This ceiling effect can be seen in Figure 8. Across time, the factor score distribution moved upward but bent to the ceiling. Abstract math skills under state 1 started very low and developed over time. At time point 5 (3rd grade) students started to master math and switch to state 2 and most students mastered math in grades 5 or 8 (time points 6 and 7) (cf. Figure 7).

## DISCUSSION

Ambulatory assessment has become a standard technique for the examination of human behavior, psychophysiological parameters, and subjective measures. It can be assumed that intra-individual changes of these variables are influenced by dispositions (e.g., inter-individual differences such as traits) and time-specific variables (e.g., events, interventions). There is a need to integrate individual-specific and time-specific effects in psychometric models. Furthermore, unobserved (latent) heterogeneity can be a substantial part of intra-individual trajectories, which can be an expression of specific states individuals are located in (e.g., an intention to quit colleague). Person- or time-specific effects can influence the probabilities of changes between these latent states.

Here, we proposed a general framework that can integrate (i) discrete latent processes, (ii) individual-specific, and (iii) time-specific effects. Furthermore, (iv) flexible structural relationships on both the between level and within level were included that are an important feature of more realistic models. In the past, there have been very elaborated models that combined different features (e.g., (i) and (ii) or (ii) and (iii)) but lacked a combination of all four aspects. Two examples are the DLCA (Asparouhov et al., 2017a) and DSEM (Asparouhov et al., 2017b) frameworks. In the DLCA framework, discrete changes of latent states are described as a Hidden Markov process. Its transition probabilities are (person-specific) random effects, which are driven by inter-individual differences. In the DSEM framework, individual-specific and time-specific effects are used to explain random effects, but unobserved discrete states are not part of the overall model. The proposed

NDLC-SEM framework is intended to combine these frameworks and to add (semi-parametric) nonlinear effects which account for flexible relationships of the latent variables at the between and within level. As another extension the intra-individual Hidden Markov process is a result of both individual-specific and time-specific effects.

In a first example, we examined simulated data of college students in mathematics and explained their intentions to quit their studies (i.e., college drop-out). We assumed that there was an unobserved intra-individual process (with latent states) which reflected their daily affect. The occurrence probabilities of the latent states, which accounted for the heterogeneity in the intra-individual trajectories, were driven by inter-individual differences (e.g., some vulnerability factors, such as lack of conscientiousness and emotional stability) in a nonlinear model. There are college drop-out theories which rely on complex (nonlinear) dependencies and several types of predictors (e.g., Bean, 2005; Burrus et al., 2013; Tinto, 1993) and which address several levels of data/predictors (incl. person- and time-specific covariates of processes, but also institutional and social factors).

In a second empirical example with data from the ECLS-K Study, we showed how the framework can be used to model continuous and nominal change of math skill levels. It is an ongoing challenge and discussion about how math mastery should be modeled. While most research assumes continuous latent factors, categorical latent factors are plausible, too (e.g., Doignon & Falmagne, 2012). Here, we showed how both can be integrated in a single model by using a time- and person-specific state membership that indicated math mastery. Students who had mastered math could be described with a single math skill factor. Students who had not yet mastered math needed two constructs (concrete and abstract math skills) to account for the underlying dimensions in the data. State membership could be predicted with both the time-varying covariates reading skills and a baseline covariate of fine motor skills (e.g., Luo et al., 2007). We found an interaction effect between these two variables that indicated that high scores in one of the constructs (reading skills or fine motor skills) could compensate low scores in the other one in the transition to math mastery.

In the NDLC-SEM framework, the integration and estimation of these different effects has been achieved by using Bayesian MCMC methods. Bayesian methods allow researchers to flexibly implement complex relationships between variables, hierarchical data structures, dynamic latent classes, or nonparametric relationships (e.g., regression splines). The implementation of the example has been achieved with the JAGS software (Plummer, 2003) using the R2jags package (Su & Yajima, 2015). We provide syntax for applied researchers which contains characteristic parts of the NDLC-SEM, such that the syntax can be expanded easily for more complex models. This implementation (e.g., the time-dependent class membership of individuals) goes beyond what is currently possible in commercial software packages. The implementation and estimation of models within the NDLC-SEM framework though builds on some assumptions which we would like to reflect critically.

## Estimation

Generally speaking, estimation of models in the NDLC-SEM framework is possible with both Bayesian and Frequentist approaches. However, both perspectives imply different challenges, which need to be considered.

### *Bayesian estimation*

As can be seen from the example section, the implementation in a Bayesian context requires the specification of well-considered prior distributions. In our implementation, the prior distribution of (a) coefficients on the between level (see Equations (41)–(43)) and (b) mean structures of the residuals of the different latent classes (see Equation (45)) were kept rather uninformative to avoid estimates that are too dependent on the chosen priors. More informative priors may speed up convergence, but need to be based on reliable information.

Results not reported here showed that an alternative parameterization, which allowed for a free estimation of the class-specific means on the within level needed highly informative priors (i.e., variances of 0.1 around the true means) in order to detect the two latent classes. Even though the parameter might not be identified in a traditional frequentist meaning, the choice of informative priors allows for "approximate identification" (for a similar modeling approach in differential item functioning, see Shi, Song, Liao, Terry, & Snyder, 2017). We avoided this by imposing a constraint that holds under the assumption that all persons are in state 1 at the first measurement occasion (for example, all students have the intention to stay at the first day). This assumption might not always hold and needs further evidence provided by researchers or needs to be adapted to the situation.

Another important aspect of Bayesian estimation is the need for a sparse estimation. The proposed NDLC-SEM offers many parameters, e.g., fixed effects (which might be related, for example, to large initial assessments with many covariates), random effects, and mixtures of distributions. In order to obtain sparse models, it is useful to consider regularized estimation. While for fixed effects, there is literature in Bayesian modeling, which deals with suitable prior distributions for penalized models (e.g., Feng et al., 2017b; Feng, Wu, & Song, 2017a; Guo et al., 2012), for mixture models, the field of sparse estimation of the number of mixture components is of increasing importance (e.g., Liu & Song, 2018; Papastamoulis & Iliopoulos, 2009; Richardson & Green, 1997). In addition, research on Bayesian penalized estimation so far focused predominantly on observed variable models (e.g., Bhadra, Datta, Polson, & Willard, 2017; Bhattacharya, Pati, Pillai, & Dunson, 2015; Carvalho, Polson, & Scott, 2010; Piironen & Vehtari, 2017) and future investigations are necessary to show if similar recommendations on prior usage hold.

### *Frequentist estimation*

A frequentist alternative to the Bayesian estimator that does not need priors are maximum likelihood estimators based on the expectation maximization algorithm (EM; Dempster, Laird, & Rubin, 1977). However, estimation of nonlinear models is already limited for single-level situations; their computational burden increases (exponentially) with the number of nonlinear terms (e.g., interactions) (e.g., Klein & Moosbrugger, 2000) and random effects (e.g., Muthén & Asparouhov, 2009). For clustered (i.e., multilevel) data structures as well as for large-scale or time-intensive data sets, this computational burden easily increases to an amount that is not feasible to estimate with today's available computer power. The reason for that is that likelihood-based estimators include integrals, which marginalize random effects (such as latent variables) (e.g., Rabe-Hesketh et al., 2004). These likelihood integrals need to be approximated by elaborated quadratures, for example, within a new adaptation of the EM algorithm that still needs to be developed. The inclusion of latent classes, as they are given in Markovian processes, increases the complexity of the estimation. The multilevel structure of the data leads to additional latent variables which increase the dimensionality of the integral to be optimized. The number of knots in the quadratures explodes. Relaxation of distributional assumptions of latent variables or residuals requires additional optimization routines which, for example, approximate nonnormality (e.g., Kelava & Brandt, 2014; Kelava, Nagengast, & Brandt, 2014). In order to obtain a very first implementation of a frequentist estimation procedure for the NDLC-SEM framework, it might be plausible to start with simple assumptions of multivariate normality.

Extensions of the EM algorithm for large-scale data have recently been discussed for simpler Gaussian mixtures (e.g., Huang, Peng, & Zhang, 2017; Ju & Liu, 2010; Medeiros, Araújo, Macedo, Chella, & Matos, 2014; Ordonez & Omiecinski, 2002; Sato & Ishii, 2000; Vlassis & Likas, 2002). For example, Verma, Dwivedi, and Sevakula (2015) state three possibilities that might also be applicable to the framework proposed here: First, in large-scale data sets, the data itself can be summarized within a defined grid (e.g., by calculating weights for identical data points) and the resulting smaller data set is then analyzed. Second, a parallelization of the problem could be conducted, where randomly drawn (smaller) data sets from the complete data set are analyzed in parallel (by so-called workers); this allows researcher to better exploit the power of parallel processors. And third, the inclusion of an additional approximation step that takes previous (errors of) estimates into account. As can be seen, further research is needed to apply these ideas to (high-dimensional) models formulated in the proposed framework (or equally for the DLCA/DSEM frameworks). For reasons of availability of feasible estimation procedures, we decided to conduct our analyses within a Bayesian approach. The development of frequentist estimation techniques for multilevel time-series models is an important task which needs to addressed; however, it is beyond the scope of this article.

## Recommendations

Current research (based on extensive simulation studies) shows that estimation quality in multilevel SEM frameworks depends on many aspects, such as model structure, types of variables, number of subjects and measurement occasions, and estimators (for overviews, see Depaoli & Clifton, 2015; Holtmann, Koch, Lochner, & Eid, 2016). Thus, giving general recommendations is difficult.

Depaoli and Clifton (2015) showed that the choice of prior distributions for between-level random effects is important. Uninformative priors for between-level random effects are sub-optimal with respect to parameter recovery in small samples. In these situations, the authors recommended that applied researchers should consider "the magnitude of between-group variation present in their data to inform the choice of priors." Although they focused on multilevel SEM in their publication, parameters on the between level should be sensitive to small samples in the NDLC-SEM, DLCA, and DSEM frameworks, too, and similar recommendations hold.

The same is true for mixture models (e.g., Depaoli, Yang, & Felt, 2017). In their work, the authors showed that parameter estimates related to the latent classes are substantially influenced by the choice of the prior distributions. In line with their work, we recommend a sensitivity analysis of the prior distributions to understand the stability of the parameter estimates.

Schultzberg and Muthén (2018) showed in the context of the DSEM framework that larger samples (e.g., $N = 200$) and few measurement occasions (e.g., $T = 10$) performed substantially better than samples with fewer subjects (e.g., $N = 20$) and many measurement occasions (e.g., $T = 100$). Large $N$ seem to be more beneficial than large $T$. Therefore, we also recommend to have a substantially larger number of subjects than measurement occasions. The reader may refer to extensive simulation studies as given by Schultzberg and Muthén (2018), when DSEM-type models have to be estimated. In order to reliably use these methods for situations with few persons in intensive longitudinal sets, as they exist, for example, in psychophysiological experiments, further research on optimal sets of priors needs to be done. Generally speaking, although there is literature on frequentist and Bayesian multilevel SEM, further research is needed to understand the requirements for reliable parameter estimation in more complex frameworks, such as the one proposed or DLCA and DSEM.

Finally, we would like to give a cautionary note on the specification and estimation of the proposed NDLC-SEM framework. Theoretically, the proposed framework could be estimated in its generality (pending on typical identification assumptions of latent variable mixture modeling and dynamic SEM). However, its main contribution is that it allows researchers to choose from many important conceptual elements, which are only partially given in other frameworks. For example, the model equations offer many features like separation of intra-individual changes from inter-individual differences and time-specific effects, unobserved heterogeneity as part of the Hidden Markov Models, a large number of (autoregressive) lagged effects and flexible nonlinear effects (and covariates) which can be included on both the within and between level. We would like to emphasize that with the complexity of the framework and the technical capabilities, the responsibility for sparse models grows (in the sense of Occam's razor). We give two recommendations related to sparsity. First, we advise against a naive estimation of a model that is not sparse enough and encourage to specify those features of the framework which are important from a theoretical perspective, but cannot be specified by other dynamic latent variable frameworks (e.g., nonlinear effects). A model that includes too many parameters will probably overfit the data. In each empirical situation, we recommend that the decision concerning the necessity to include different parts of the model depends on both theoretical considerations (such as inclusion of nonlinear effects in substantive theories) and model fit comparisons of competing models. Thus, the proposed framework serves as a starting point of model specification.

Second, if high-dimensional data, for example, from ambulatory assessment with many repeated measures are analyzed, models are necessarily complex. Comprehensive modeling techniques offer an inclusion of a high number of predictors/covariates (e.g., on the between level), especially when there is a huge initial assessment at the beginning of a study, in which many scales were administered (e.g., in the COGITO study, Schmiedek, Lövdén, & Lindenberger, 2010). Or, strong assumptions about measurement invariance over many repeated measures are too strong and need to be relaxed (Bauer, 2018; Liang, Yang, & Huang, 2018). In these cases, it might be meaningful to reduce the number of actually relevant parameters (i.e., effective number of degrees of freedom) by using penalized estimation techniques (e.g., Brandt et al., 2018; Jacobucci, Grimm, & McArdle, 2016). At this time point, there still is a need of technical developments in this field that addresses sparsity in complex models as in this framework that has not been investigated yet.

## Future directions and conclusion

In sum, Bayesian estimation (as applied here) allows a great flexibility of specifying models. However, there is a need for frequentist estimation, which builds on less assumptions than Bayesian estimation. To the best of our knowledge, there are no comprehensive frequentist estimators which are capable of estimating all effects simultaneously that can be described by the proposed NDLC-SEM framework.

## FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft [KE 1664/1-2];

## REFERENCES

Altman, R. M. (2007). Mixed hidden Markov models. *Journal of the American Statistical Association*, *102*(477), 201–210. doi:10.1198/016214506000001086

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017a). Dynamic latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 257–269. doi:10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017b). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–30. doi:10.1080/10705511.2017.1406803

Asparouhov, T., & Muthèn, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163–192). Charlotte, NC: Information Age.

Bauer, D. J. (2018). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*, 507–526. doi:10.1037/met0000077

Bean, J. P. (2005). Nine themes of college student retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (pp. 215–244). New York, NY: Rowman & Littlefield.

Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, *12*, 1105–1131. doi:10.1214/16-BA1028

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet-Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, *110*, 1479–1490. doi:10.1080/01621459.2014.960967

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2015). *Time series analysis: Forecasting and control, 5th edition*. Hoboken, NJ: Wiley.

Brandt, H., Cambria, J., & Kelava, A. (2018). An adaptive Bayesian lasso approach with spike-and-slab priors to identify multiple linear and nonlinear effects in structural equation models. *Structural Equation Modeling*, *25*, 946–960. doi:10.1080/10705511.2018.1474114

Burrus, J., Elliott, D., Brenneman, M., Markle, R., Carney, L., Moore, G., … Roberts, R. D. (2013). Putting and keeping students on track: Toward a comprehensive model of college persistence and goal attainment. *ETS Research Report Series*, *2013*(1), i–61. doi:10.1002/j.2333-8504.2013.tb02321.x

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*, 465–480. doi:10.1093/biomet/asq017

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, Ser. B*, *39*, 1–38. Retrieved from http://www.jstor.org/stable/2984875

Depaoli, S., & Clifton, J. P. (2015). A bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 327–351. doi:10.1080/10705511.2014.937849

Depaoli, S., Yang, Y., & Felt, J. (2017). Using bayesian statistics to model uncertainty in mixture models: A sensitivity analysis of priors. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 198–215. doi:10.1080/10705511.2016.1250640

Doignon, J.-P., & Falmagne, J.-C. (2012). *Knowledge spaces*. Berlin: Springer Science & Business Media.

Durbin, J., & Koopman, S. (2001). *Time series analysis by state space methods*. Oxford, UK: Oxford University Press.

Feng, X.-N., Wang, G.-C., Wang, Y.-F., & Song, X.-Y. (2015). Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Statistics in Medicine*, *34*, 1527–1547. doi:10.1002/sim.6410

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2015). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods & Research, (Online)*, 1–28. doi:10.1177/0049124115610349

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017a). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods & Research*, *46*(4), 926–953. doi:10.1177/0049124115610349

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017b). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(3), 341–358. doi:10.1080/10705511.2016.1257353

Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review*, *6*, 123–151. doi:10.1207/S15327957PSPR0602_03

Guo, R., Zhu, H., Chow, S.-M., & Ibrahim, J. G. (2012). Bayesian lasso for semiparametric structural equation models. *Biometrics*, *68*, 567–577. doi:10.1111/j.1541-0420.2012.01751.x

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York, NY: Springer Science & Business Media.

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research*, *51*(5), (PMID: 27594086), 661–680. doi:10.1080/00273171.2016.1208074

Huang, T., Peng, H., & Zhang, K. (2017). Model selection for Gaussian mixture models. *Statistica Sinica*, *27*(1), 147–169.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793

Ju, Z., & Liu, H. (2010, July). Applying fuzzy EM algorithm with a fast convergence to GMMs. In *International conference on fuzzy systems* (pp. 1–6). Barcelona, Spain. doi: 10.1109/FUZZY.2010.5584456

Kelava, A., & Brandt, H. (2014). A general nonlinear multilevel structural equation mixture model. *Frontiers in Quantitative Psychology and Measurement*, *5*(748). doi:10.3389/fpsyg.2014.00748

Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for non-normally distributed latent predictor variables. *Structural Equation Modeling*, *21*, 468–481. doi:10.1080/10705511.2014.915379

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi:10.1007/BF02296338

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*, 383–409. doi:10.1016/S0959-4752(01)00011-1

Liang, X., Yang, Y., & Huang, J. (2018). Evaluation of structural relationships in autoregressive cross-lagged models under longitudinal approximate invariance: A Bayesian analysis. *Structural Equation Modeling*, *25*, 558–572. doi:10.1080/10705511.2017.1410706

Liu, H., & Song, X. Y. (2018). Bayesian analysis of mixture structural equation models with an unknown number of components. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 41–55. doi:10.1080/10705511.2017.1372688

Luo, Z., Jose, P. E., Huntsinger, C. S., & Pigott, T. D. (2007). Fine motor skills and mathematics achievement in East Asian American and European American kindergartners and first graders. *British Journal of Developmental Psychology*, *25*(4), 595–614. doi:10.1348/026151007X185329

Marcoulides, K. M., & Khojasteh, J. (2018). Analyzing longitudinal data using natural cubic smoothing splines. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–7. doi:10.1080/10705511.2018.1449113

Medeiros, M., Araújo, G., Macedo, H., Chella, M., & Matos, L. (2014). Multi-kernel approach to parallelization EM algorithm for GMM training. In *Brazilian conference on intelligent systems (bracis), ieee* (pp. 158–165). Sao Paulo, Brazil. doi: 10.1109/BRACIS.2014.38

Molenaar, P. C. M. (1985, Jun 01). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202. doi:10.1007/BF02294246

Molenaar, P. C. M. (2017). Equivalent dynamic models. *Multivariate Behavioral Research*, *52*(2), (PMID: 28207288), 242–258. doi:10.1080/00273171.2016.1277681

Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Ordonez, C., & Omiecinski, E. (2002). FREM: Fast and robust EM clustering for large datasets. In *Proceedings of the eleventh international conference on information and knowledge management, acm* (pp. 590–599). McLean, VA. doi: 10.1145/584792.584889

Papastamoulis, P., & Iliopoulos, G. (2009). Reversible jump MCMC in mixtures of normal distributions with the same component means. *Computational Statistics & Data Analysis*, *53*(4), 900–911. doi:10.1016/j.csda.2008.10.022

Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*, 5018–5051. doi:10.1214/17-EJS1337SI

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*, 167–199. doi:10.3102/00346543063002167

Plummer, M. (2003). *Jags: A program for analysis of Bayesian graphical models using GIBBS sampling*. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria.

R Core Team. (2016). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. doi:10.1007/BF02295939

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Ser. B*, *59*, 731–792. doi:10.1111/1467-9868.00095

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49* (1), 37–47. doi:10.1007/BF02294204

Sato, M., & Ishii, S. (2000). On-line EM algorithm for the normalized Gaussian network. *Neural Computation*, *12*, 407–432. doi:10.1162/089976600300015853

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the cogito study. *Frontiers in Aging Neuroscience*, *2* (27). doi:10.3389/fnagi.2010.00027

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 495–515. doi:10.1080/10705511.2017.1392862

Shi, D., Song, H., Liao, X., Terry, R., & Snyder, L. A. (2017). Bayesian SEM for specification search problems in testing factorial invariance. *Multivariate Behavioral Research*, *52*(4), 430–444. doi:10.1080/00273171.2017.1306432

Song, X.-Y., Li, Z.-H., Cai, J.-H., & Ip, E. H.-S. (2013). A Bayesian approach for generalized semiparametric structural equation models. *Psychometrika*, *78*, 624–647. doi:10.1007/s11336-013-9323-7

Su, Y.-S., & Yajima, M. (2015). R2jags: Using r to run 'jags' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=R2jags(R package version 0.5-7)

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (*2nd* ed.). Chicago, IL: University of Chicago Press.

Tourangeau, K., Nord, C., Lê, T., Pollack, J. M., & Atkins-Burnett, S. (2009). Early childhood longitudinal study, kindergarten class of 1998-99 (ECLS-K), combined user's manual for the ECLS-K eighth-grade and K-8 data files and electronic codebooks (NCES 2009-004) [Computer software manual]. Washington, DC. Retrieved from https://nces.ed.gov/ecls/data/ECLSK_K8_Manual_part1.pdf

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*(6), (PMID: 25530686), 466–470. doi:10.1177/0963721414550706

Verma, N. K., Dwivedi, S., & Sevakula, R. K. (2015, Dec). Expectation maximization algorithm made fast for large scale data. In *2015 IEEE workshop on computational intelligence: Theories, applications and future directions (WCI)* (pp. 1–7). Kaupur, India: IEEE. doi:10.1109/WCI.2015.7495515

Vlassis, N., & Likas, A. (2002). A greedy EM algorithm for Gaussian mixture learning. *Neural Processing Letters*, *15*, 77–87. doi:10.1023/A:1013844811137

Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176. doi:10.1037/a0027543

Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*, 552–565. doi:10.1037/0012-1649.27.4.552

Wood, S. N. (2017). *Generalized additive models: An introduction with r*. Boca Raton, FL. Chapman and Hall/CRC.

Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(3), 377–402. doi:10.1080/10705510802154281

Zhang, Z., & Nesselroade, J. R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research*, *42*(4), 729–756. doi:10.1080/00273170701715998

## APPENDIX

### Model setup and priors for the empirical example

***Distribution of variables.*** All observed ($math_{1jit}$) and latent variables $\eta$ were assumed to follow a normal distribution with the respective mean structure and variance:

$$[math_{1jit}|S_{it} = s] \sim N(\mu_{1jits}, \sigma^2_{\epsilon_{1j}}) \quad (52)$$

$$\sim MVN(\mu_{\eta,it1}, \Phi_{\zeta_{11}}) \quad (53)$$

$$\sim N(\mu_{\eta,it2}, \sigma^2_{\zeta_{12}}) \quad (54)$$

where $N(\mu, \sigma^2)$ was the normal distribution with mean $\mu$ and variance $\sigma^2$. The ARIMA(1,1,0) latent variable model is described in Equation (51) and the factor loading pattern is illustrated in Figure 6.

***Markov Switching Model.*** The probabilities for state membership were modeled using a time- and person-specific latent variable $\alpha_{itcd}$ for $t > 1$ (all persons were assumed to be in state $S_{i1} = 1$ at $t = 1$).

$$P(S_{it} = 1|S_{i(t-1)} = 1) = \frac{\exp(\alpha_{it11})}{\Sigma_{k=1}^{2} \exp(\alpha_{itk1})} \quad (55)$$

$$P(S_{it} = 2|S_{i(t-1)} = 1) = 1 - P(S_{it} = 1|S_{i(t-1)} = 1) \quad (56)$$

$$P(S_{it} = 1|S_{i(t-1)} = 2) = 0.01 \quad (57)$$

$$P(S_{it} = 2|S_{i(t-1)} = 2) = 1 - P(S_{it} = 1|S_{i(t-1)} = 2) \quad (58)$$

where we chose a very small probability for those students that mastered math to switch back to a non-mastery state of $\pi = 0.01$. The latent variable $\alpha_{it11}$ was specified as (see Figure 6)

$$\begin{aligned}\alpha_{it11} = \alpha_{11} + \beta_{11} \cdot read_{i,t-1} + \omega_{13} \\ \cdot (read_{it} - read_{i(t-1)}) + \beta_{21} \cdot motor_i + \omega_{21} \\ \cdot motor_i \cdot read_{i(t-1)} + \omega_{22} \cdot motor_i \\ \cdot (read_{it} - read_{i(t-1)})\end{aligned} \quad (59)$$

***Prior distributions***. Priors were chosen as weakly informative priors throughout the model. For the measurement model on the within level, factor loading and intercept priors were specified as

$$\lambda_{1j} \sim N(1, 1), \text{ for } j = 1 \ldots 5 \quad (60)$$

$$\tau_{1j} \sim N(0, 2), \text{ for } j = 1 \ldots 4. \quad (61)$$

For the structural models coefficients on the within and between levels, again weakly informative priors were chosen:

$$\beta_{11} \sim N(0, 1) \quad (62)$$

$$\beta_{21} \sim N(0, 2) \quad (63)$$

$$\omega_{13} \sim N(0, 1) \quad (64)$$

$$\omega_{2p} \sim N(0, 2) \text{ for } p = 1, 2 \quad (65)$$

$$\alpha_{11} \sim N(0, 2) \quad (66)$$

where the constraint $\alpha_{11} = 0$ was necessary for model identification. Note that this constraint always holds in this model if data are rescaled by $Y^c_{1jit} = Y_{1jit} - \bar{Y}_{111}$ because $Y_{11i1} = \eta_{1i1}$ and all persons are in state $S_{i1} = 1$ at the first measurement occasion.

Standard priors were chosen for the precisions as

$$\sigma^{-2}_{\epsilon_{1j}} \sim Gamma(9, 4), \text{ for } j = 1 \ldots 5 \quad (67)$$

$$\Phi^{-1}_{\zeta_{11}} \sim Wishart(\Phi^{-1}_0, 4) \quad (68)$$

$$\sigma^{-2}_{\zeta_{12}} \sim Gamma(9, 4) \quad (69)$$

where $\Phi_0$ was a $2 \times 2$ identity matrix.