Structural Equation Modeling: A Multidisciplinary Journal

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

Dynamic Latent Class Analysis

Tihomir Asparouhov, Ellen L. Hamaker & Bengt Muthén

To cite this article: Tihomir Asparouhov, Ellen L. Hamaker & Bengt Muthén (2017) Dynamic Latent Class Analysis, Structural Equation Modeling: A Multidisciplinary Journal, 24:2, 257-269, DOI: 10.1080/10705511.2016.1253479

To link to this article: http://dx.doi.org/10.1080/10705511.2016.1253479

Published online: 21 Dec 2016.

Structural Equation Modeling: A Multidisciplinary Journal, 24: 257–269, 2017

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2016.1253479

Dynamic Latent Class Analysis

Tihomir Asparouhov,[1] Ellen L. Hamaker,[2] and Bengt Muthén[1]
[1]*Muthén and Muthén, Los Angeles, CA*
[2]*Utrecht University*

This article describes the general time-intensive longitudinal latent class modeling framework implemented in M*plus*. For each individual a latent class variable is measured at each time point and the latent class changes across time follow a Markov process (i.e., a hidden or latent Markov model), with subject-specific transition probabilities that are estimated as random effects. Such a model for single-subject data has been referred to as the regime-switching state-space model. The latent class variable can be measured by continuous or categorical indicators, under the local independence condition, or more generally by a class-specific structural equation model or a dynamic structural equation model. We discuss the Bayesian estimation based on Markov chain Monto Carlo, which allows modeling with arbitrary long time series data and many random effects. The modeling framework is illustrated with several simulation studies.

INTRODUCTION

Latent class analysis is primarily used in cross-sectional studies where subjects are observed at one occasion only. In longitudinal studies where observations are obtained at several occasions and the number of occasions is small we can use classic latent transition models to conduct latent class analysis at each time point and study the changes in latent class membership across time. This has been a popular approach for panel data (i.e., a small number, say <6, of repeated measurements obtained from a relatively large sample of individuals or cases). In the last several years, however, intensive longitudinal data with many repeated measurements (say >20) from a large number of individuals or cases, have become much more common. These data are often collected using smart phones or other electronic devices, such that a latent construct can be measured weekly, daily, or even hourly for extended periods of time. This type of data is referred to as ambulatory assessments (AA), *daily diary data, ecological momentary assessment* (EMA) *data*, or *experience sampling methods* (ESM) data (cf. Trull & Ebner-Priemer, 2014). The accumulation of these types of data naturally leads to an increasing demand for statistical methods that allow us to model the dynamics over time as well as individual differences therein using intensive longitudinal data.

The goal of the novel modeling framework we describe here is to allow for the study of (a) a latent (or hidden) Markov model that accounts for the switching between different states (also referred to as latent classes or regimes), with (b) individual differences in transition probabilities modeled as random effects, and (c) the option of dynamic relationships within each state through time series analysis and multilevel extensions of this. Although there have been combinations of some of these three elements before (e.g., the regime-switching state-space model proposed by Kim and Nelson (1999) combines a hidden Markov model with dynamic relationships for single-subject data, and Altman (2007) combined a hidden Markov model with random transition probabilities to allow for individual differences in the switching process across individuals), to date there has not been a framework that combines all three of these elements simultaneously. The new framework presented here is based on combining two other frameworks: dynamic structural equation modeling (DSEM, see Asparouhov, Hamaker and Muthén (2016), which is one of the major innovations of M*plus* Version 8, and an extension of the existing general multilevel mixture framework developed by Asparouhov and Muthén (2008). Both are briefly discussed next.

Correspondence should be addressed to Tihomir Asparouhov, Muthén and Muthén, 3463 Stoner Avenue, Los Angeles, CA 90066. E-mail: tihomir@statmodel.com

The DSEM framework that is implemented in M*plus* Version 8 uses time series models for observed and latent variables to account for the dependencies between observations over time. Such models date back to Kalman (1960) and are applied extensively in engineering and econometrics. In most such applications, however, multivariate time series data of a single case (i.e., N = 1) is analyzed. In contrast, the intensive longitudinal data that are currently gathered in the social sciences typically come from a relatively large sample of individuals, which gives rise to a need for statistical techniques that allow us to analyze the time series data from multiple independent individuals simultaneously; such an approach is based on borrowing information from other cases, while keeping the model flexible enough to allow for subject-specific model parameters. The DSEM framework implemented in M*plus* accommodates this more complex modeling need.

The other development that is fundamental for this presentation consists of an extension of the existing multilevel mixture framework in Mplus. This framework is generalized in three ways. The first generalization is formed by Bayesian estimation of multilevel mixture models, which makes it possible to have much larger numbers of random effects. This on its own is an important advantage because a number of models that previously were unavailable or were mostly theoretical due to extremely large computational times are now possible using Bayes estimation. We illustrate this with several simulation examples.

The second generalization that we describe here is the combination of the hidden Markov model (HMM) with the multilevel mixture model. In the Asparouhov and Muthén (2008) multilevel mixture framework it is possible to model latent class intensive longitudinal data by estimating a two-level mixture model where the cluster consists of all the observations for one individual across time, but this model does not allow us to study the correlation in the latent class variable during consecutive periods (i.e., the autocorrelation). In that multilevel framework latent class variables are correlated due to being nested within the same individual but not due to being in consecutive periods. The HMM with subject-specific transition probabilities fills this gap. One of the goals of intensive longitudinal analysis is to model these two distinct sources of correlation: the within-individual correlations due to the subject-specific effect (two-level modeling) and the autocorrelation, that is, the correlation due to proximity of observations (time series modeling). These two types of correlations are easy to parse out from the data in sufficiently long longitudinal data.

The third generalization is the combination of DSEM with the multilevel HMM, which implies that observed or latent continuous variables can be autocorrelated both through the latent class autocorrelation (i.e., the HMM), and directly (i.e., through the dependencies over time in the form of, for instance, autoregressive relationships using DSEM). The latter appears to be essential if the observations are quite frequent; in the extreme, if the time between observations converges to 0, we should expect not only that the latent class variable will remain the same as in the previous period, but also that the observed or latent variables that are used as class indicators remain almost unchanged from one measurement moment to the next (which means the autocorrelation within each class will also be high). In practical applications we cannot determine a priori if the observations are taken frequently enough to warrant within-class autocorrelation and therefore one should always consider the possibility for that.

The combination of DSEM with the multilevel mixture model also implies that DSEM can be generalized to nonhomogeneous populations, the same way finite mixture structural equation modeling (SEM) models generalize cross-sectional SEM models to non homogeneous populations. Thus we can refer to this framework also as mixture-DSEM. Most of the issues that arise in finite-mixture models also arise in mixture-DSEM, such as how to decide on the number of classes, how to avoid label switching in Markov Chain Monk Carlo (MCMC) estimation, what happens when the distribution of the variables is nonnormal within a class, how to choose starting values for the estimation, and so on. One can draw a parallel to cross-sectional mixture modeling, and use that as a guiding principle for what to expect in the mixture-DSEM framework. The framework we present can also be thought of as the merger of time series, structural equation, multilevel, and mixture modeling concepts in a generalized modeling framework.

Consider the following hypothetical example. A group of patients answer daily a brief survey to evaluate their current state. Based on current observations, past history, most recent history, and similar behavior from other patients we classify the patient into one of three states: State 1: healthy, State 2: increased risk of relapse, State 3: relapse. The model needed for this kind of classification is included in this framework. If such an automatic diagnosis program is implemented, it can potentially reduce cost of care and improve outcome by identifying the critical needs of the patient. Although this example is purely hypothetical it certainly highlights the vast potential of this methodology.

The remainder of the article is organized as follows. First, we present the DSEM framework, which can be used to do single-subject time series analysis, as well as multilevel extensions of this. The latter has also been referred to as dynamic multilevel modeling. Second, we extend the DSEM framework to a mixture model, thus combining the (single or multilevel) time series models with a within level latent class modeling. We then consider some implications the Bayesian estimation has on the general multilevel mixture modeling, unrelated to time series modeling. Simulation studies are presented on three multilevel mixture models that are now possible because of the Bayesian estimation: multilevel latent class analysis with measurement noninvariance, the unrestricted two-level mixture model, and multilevel latent transition analysis (MLTA) with cluster specific transition probabilities. We then introduce the HMM for single-level models and illustrate the model with a simulation study. Finally we combine the previously discussed modeling techniques to formulate the general dynamic latent class analysis (DLCA) model and illustrate the model with three simulation studies: a simple two-class DLCA model, the multilevel markov switching autoregressive model (MMSAR), and DLCA model with regime switching for the latent factor. We conclude with a discussion for future research.

DYNAMIC STRUCTURAL EQUATION MODEL

Here we present an overview of the DSEM framework implemented in M*plus*. This is a general modeling framework intended to encompass diverse DSEM models that have already appeared in the literature, including many time series models. DSEM can be used to estimate structural models with intensive longitudinal data. A more complete discussion is available in Asparouhov et al. (2016). Consider the DSEM model of lag L. Let Y_{it} be an observed vector of measurements for individual i at time t. We begin with the usual within-between decomposition

$$Y_{it} = Y_{1,it} + Y_{2,i} \tag{1}$$

where $Y_{2,i}$ is the individual-specific random effect and $Y_{1,it}$ is the individual i deviation at time t. The two components are assumed to be normally distributed random vectors and are used to form two separate sets of structural equations—one on each level. The Level 2 structural equation model takes the usual form

$$Y_{2,i} = \nu_2 + \Lambda_2 \eta_{2,i} + \varepsilon_{2,i} \tag{2}$$

$$\eta_{2,i} = \alpha_2 + B_2 \eta_{2,i} + \Gamma_2 x_{2,i} + \xi_{2,i} \tag{3}$$

where $x_{2,i}$ is a vector of individual-specific time-invariant covariates and $\eta_{2,i}$ is a vector of individual-specific time-invariant latent variables. The variables $\varepsilon_{2,i}$ and $\xi_{2,i}$ are zero mean residuals as usual and the remaining vectors and matrices in these equations are nonrandom model parameters.

The within-level structural equation model consists of two equations that are used to model the contemporaneous and lagged relationships; that is,

$$Y_{1,it} = \sum_{l=0}^{L} \Lambda_{1,i,l} \eta_{1,i,t-l} + \varepsilon_{1,it} \tag{4}$$

$$\eta_{1,i,t} = \alpha_{1,i} + \sum_{l=0}^{L} B_{1,i,l} \eta_{1,i,t-l} + \Gamma_{1,i} x_{1,it} + \xi_{1,it}. \tag{5}$$

Here $x_{1,it}$ is a vector of observed covariates for individual i at time t and $\eta_{1,i,t}$ is a vector of latent variables for individual i at time t. The difference between the standard structural model and the model in Equations 4 and 5 is that all the elements in the observed and the latent vectors on the left side have the same time index, where as the latent variables on the right side are associated with both the same occasion, but also times $t-1, \ldots, t-L$, showing that these preceding latent variables can now be used as predictors for the observed and latent variables at time t.

In the proceding equations we allow random loadings (in the Λs), random structural coefficients (in the Bs), random slopes for the exogenous variables (in Γ), and random factor intercepts (in α) on the within level. Thus, every within-level parameter can be random or nonrandom; that is, invariant across individuals. All of the random effects parameters are modeled at the between level as latent variables, meaning they are part of the vector $\eta_{2,i}$ and are modeled in Equation (3). For identification purposes restrictions need to be imposed on the proceding model along the lines of standard structural equation models. The model is the time-series generalization of the time-intensive model described in Section 8.3 of Asparouhov and Muthén (2016).

Categorical variables can easily be accommodated in this model through the probit link function. For each categorical variable Y_{ijt} in the model, $j = 1, \ldots, p$, taking the values from 1 to m_j, we assume that there is a latent variable Y^*_{ijt} and threshold parameters $\tau_{1j}, \ldots, \tau_{m_j-1j}$ such that

$$Y_{ijt} = m \Leftrightarrow \tau_{m-1j} \leq Y^*_{ijt} < \tau_{mj} \tag{6}$$

where we assume $\tau_{0j} = -\infty$ and $\tau_{m_jj} = \infty$. This definition essentially converts a categorical variable Y_{ijt} into an unobserved continuous variable Y^*_{ijt}. The model is then defined using Y^*_{ijt} instead of Y_{ijt} in Equation (1).

The DSEM model just described is a two-level model, where the individual is the clustering variable, and can be used to estimate structural models for data sets with multiple individuals over an unlimited time period. This model is the two-level extension of the dynamic factor model described in Molenaar (1985), Zhang and Nesselroade (2007) and Zhang, Hamaker, and Nesselroade (2008).

Note that for $t = 1, \ldots, L$, the latent variables $\eta_{1,i,t-l}$ used as predictors in Equations 4 and 5 have a zero or negative time index variable. This is a well-known problem in the time series literature and various solutions have been proposed. To this end, we treat these as auxiliary parameters that have a prior distribution. Two different methods are implemented in Mplus. The first one requires the prior distribution to be specified before the estimation of the model. The second option starts assuming that these initial variables are zero for the first MCMC iteration and in the next 100 MCMC iterations the prior for these auxiliary variables are updated to be the normal distribution with mean and variance equal to the sample mean and the sample variance of the corresponding imputed latent variables with a positive time index. After the first 100 iterations the priors are no longer updated to retain proper MCMC estimation. Note here that in general the influence of the initial values, or more accurately stated, the influence of the priors for these initial conditions is minimal when the time series length is sufficiently long. For practical purposes a length of 50 is sufficiently large to eliminate almost entirely the effect of the initial value priors (but note that this is not necessarily the case for the effect of the other priors used for the model when the number of individuals in the sample is small).

The estimation of the DSEM is a combination of the Bayes estimation method described in Zhang and Nesselroade (2007) and Zhang et al. (2008) for single-level DSEM and the two-level estimation described in Asparouhov and Muthén (2010); when $[Y_{2,i}|*]$ is generated in the MCMC, the latent variables are conditioned on, thus the two procedures can easily be combined. That is, once the between-level parts have been generated $[Y_{2,i}|*]$, then $Y_{1,it} = Y_{it} - Y_{2,i}$ can be computed, and for $Y_{1,it}$ the model is a single-level DSEM model such that the Zhang and Nesselroade (2007) procedure can be applied to it. The latter includes the generation of the latent variables $\eta_{1,i,t}$, which are then multiplied by the corresponding loadings and subtracted from Y_{it}. At that point the Asparouhov and Muthén (2010) two-level algorithm applies because the within-cluster data are no longer correlated.

DSEM MIXTURE MODEL

Let S_{it} be a categorical latent variable for individual i at time t; that is, S_{it} is a within-level latent class. In the time series literature such a latent class variable is more often referred to as a latent state variable; therefore, we use S as the variable name rather than the traditional C. Suppose that S_{it} takes values $1, 2, \ldots, K$ where K is the number of classes in (or states of) the model. The DSEM mixture model consists of Equations 1 through 5, however Equations 4 and 5 now depend on S_{it} as follows:

$$[Y_{1,it}|S_{it} = s] = \nu_{1,s} + \sum_{l=0}^{L} \Lambda_{1,l,s}\eta_{i,t-l} + \varepsilon_{it} \tag{7}$$

$$[\eta_{i,t}|S_{it} = s] = \alpha_{1,s} + \sum_{l=0}^{L} B_{1,l,s}\eta_{i,t-l} + \Gamma_{1,s}x_{it} + \xi_{it}. \tag{8}$$

The residual covariance matrices of ε_{it} and of ξ_{it} are also state specific, meaning they depend on s. If the model includes categorical variables, then Equation 6 also becomes state specific:

$$[Y_{ijt} = m|S_{it} = s] \Leftrightarrow \tau_{m-1js} \leq Y^*_{ijt} < \tau_{mjs}. \tag{9}$$

The distribution of S_{it} is given by

$$P(S_{it} = s) = \frac{Exp(\alpha_{is})}{\sum_{s=1}^{K} Exp(\alpha_{is})} \tag{10}$$

where α_{is} are normally distributed random effects. For identification purposes $\alpha_{iK} = 0$. These random effects α_{is} are included as part of the between-level latent variable vector $\eta_{2,i}$. This implies also that individual-level predictors $x_{2,i}$ can be used to predict $P(S_{it} = s)$ through the structural Equation 3.

To estimate this model we can utilize the single-level mixture model estimation described in Section 8 of Asparouhov and Muthén (2010). Conditional on α_{is} and the rest of the between-level random effects, the updating of S_{it} is the same as in single-level mixture models. To update α_{is} we use a Metropolis step with a multivariate normal symmetric jumping distribution. Given the current estimates α_{is}, the proposed $\hat{\alpha}_{is}$ values are selected from the following distribution:

$$\hat{\alpha}_{is} \sim N(\alpha_{is}, \Sigma)$$

where Σ is proportional to the identity matrix. The parameter in Σ is updated within a burn-in period to maintain proper accept–reject ratios, and after the burn-in period the parameter is no longer updated to ensure proper MCMC estimation. The burn-in period is not used for parameter inference, but only to stabilize the estimate for the jumping distribution. In M*plus* by default 1,000 burn-in iterations are used.

The new draw $\hat{\alpha}_{is}$ is accepted with probability

$$Acceptance\,ratio = min\left(1, \frac{Prior(\hat{\alpha}_{is})Likelihood(S_{it}|\hat{\alpha}_{is})}{Prior(\alpha_{is})Likelihood(S_{it}|\alpha_{is})}\right)$$

where $Prior(\alpha_{is})$ is the density function implied by the between-level model and is conditional on all other between-level variables including other between-level random effects, and the $Likelihood(S_{it}|\alpha_{is})$ is simply the likelihood for the nominal variables S_{it} given the implied distribution based on Equation 10:

$$Likelihood(S_{it}|\alpha_{is}) = \prod_t \frac{Exp(\alpha_{is_{it}})}{\sum_{s=1}^{K} Exp(\alpha_{is})}.$$

Note here that the Metropolis step is performed for each cluster separately and that the jumping distribution is identical in all clusters. This Metropolis step can be further fine-tuned to improve the mixing. Possible avenues for improvement are to have cluster-specific jumping distributions, and jumping distributions with Σs that are proportional to the sample covariance matrix of α_{is} rather than the current diagonal matrix.

This above model is suitable for modeling intensive longitudinal data but not completely. Within each class DSEM allows us to model autocorrelation directly on the observed variables via the time series model for the latent continuous variables. The model in Equation 10 implies that the latent class distribution changes across individuals, but it does not allow us to model the autocorrelation of this latent class variable. Put differently, the model in Equation 10 implies that the values of the latent class variable at two consecutive time points for the same individual are conditionally independent (conditional on the random effects α_{is}). For intensive longitudinal modeling applications, this would be an unrealistic assumption. We address this issue later by including the HMM, but first we focus on exploring the proceding modeling framework as a two-level mixture modeling framework.

MULTILEVEL MIXTURE EXAMPLES

In this section we consider several examples that are now feasible due to the fact that we are using the Bayesian estimation and can estimate models with an unlimited number of random effects. Although such models were theoretically feasible even with the maxumum likelihood ML estimation, using Monte Carlo integration for example, the heavy computational demand has made these impractical. Consider, for example, a model that has no other between-level random effect except α_{is}. The ML estimation uses $K-1$ dimensions of numerical integration, making this model computationally demanding for models with more than three classes. To reduce the dimensions of numerical integration we can assume that α_{is} are proportional, but this is not a realistic assumption. With the Bayesian estimation we avoid this problem and can estimate completely unrestricted variance covariance for α_{is} with no substantial increase in the computational time.

We temporarily depart from the framework of intensive longitudinal data and focus on the standard two-level setup instead. Although the observations are nested within clusters, they are naturally unordered within each cluster (i.e., there is no time ordering of the within-cluster observations), and they are conditionally independent given the cluster-level random effects. The indexes i and t are thus replaced by j and i, where i refers to the individual and j refers to the cluster.

Example 1: Latent Class Analysis with Clustered Data and Measurement Noninvariance

The most common approach for estimating a latent class analysis model with clustered data is to use robust ML estimation. The point estimates essentially ignore the clustering, and the standard errors are adjusted upward to account for the dependence of the observations within the clusters using the sandwich estimator or the jackknife estimator (see Patterson, Dayton, & Graubard, 2002). There are several problems with this approach. First, the approach does not allow cluster-specific class distribution; that is, information from other members of the same cluster cannot contribute to the estimation of class membership for an individual. Class membership is estimated using the point estimates only, which ignore the clustering. Second, the approach is based on full measurement invariance; that is, it is assumed that item thresholds are identical across clusters. For continuous latent variables accommodating measurement noninvariance in, for instance, cross-cultural studies is essential (e.g., Davidov, Dulmer, Schluter, & Schmidt, 2012; De Jong, Steenkamp, & Fox 2007). A similar issue arises also in latent class analysis. For example, if a market segmentation study uses a sample from multiple countries, it will be unrealistic to assume latent class analysis invariance across the countries. If measurement invariance does not hold, assuming it will likely yield spurious classes. Furthermore, when the number of groups is more than a few, the differences across groups should be modeled as random effects rather than as fixed effects to preserve the parsimonious nature of the model.

In this section we consider a latent class analysis measurement noninvariance model that resolves these problems. A similar model is also considered in De Jong and Steenkamp (2009), who also used Bayesian estimation. The model is included here as it is encompassed by the general framework that we propose.

We illustrate the latent class analysis measurement noninvariance model with a simple example and a small simulation study. Consider a model with $K = 3$ latent classes measured by eight binary indicators. Let U_{pij} represent the score on indicators p for individual i in cluster j. The conditional probability for scoring 1 on such a binary indicator can be expressed as

$$P(U_{pij} = 1|C_{ij} = k) = \Phi(\tau_{pk} + \varepsilon_{pj})$$

where τ_{pk} is a nonrandom parameter (the usual threshold parameter) and ε_{pj} is a measurement noninvariance zero mean random effect that allows certain indicators to have larger or smaller probability in cluster j than the population values, beyond what the latent class distribution explains. For example, certain measurement instruments might not be universally accurate, or might not be universally good at separating the classes. Note that this probability is conditional on the individual being in the kth class. The probability that individual i from cluster j is a member of the kth class is expressed as

$$P(C_{ij} = k) = \frac{Exp(\alpha_k + \alpha_{jk})}{\sum_{s=1}^{K} Exp(\alpha_s + \alpha_{js})}.$$

The parameters α_k are nonrandom effects that fit the population-level class distribution and α_{jk} are zero mean random effects that allow cluster-specific class distribution. As usual for identification purposes $\alpha_K = \alpha_{jK} = 0$.

Using this model we generate 100 clusters of size 50 for a total sample size of 5,000 using the following parameter values $\tau_{p1} = 1$, $\tau_{p2} = -1$, $\tau_{p3} = -1$ for $p \leq 4$, $\tau_{p3} = 1$ for $p>4$, $Var(\varepsilon_{pj}) = 0.2$, $Var(\alpha_{jk}) = 0.3$, $\alpha_1 = 0.8$, and $\alpha_2 = 0.4$. The ML estimation for this model will use 10-dimensional numerical integration (i.e., 8 noninvariance random effects ε_{pj} and 2 latent class distribution random effects α_{jk}) and will be very computationally demanding. With the newly developed option of Bayesian estimation it takes only 40 seconds for each replication. We generate and analyze 100 samples. Table 1 contains the results from this simulation for a sample of the parameters. From these results we see that the parameter estimates are unbiased and the coverage is near the nominal levels.

Example 2: Unrestricted Two-Level Mixture Model

Another example of a model that is now easy to estimate, because of the use of Bayesian estimation, is the unrestricted two-level mixture model. Let Y_{ij} be a vector of observed continuous variables for individual i in cluster j. The model we are interested in is given by the following equations:

TABLE 1
Simulation Results for a Noninvariant Latent Class Analysis

Parameter	*True Value*	*Abs. Bias*	*Coverage*
τ_{11}	1	.00	.92
τ_{12}	–1	.01	.97
τ_{13}	–1	.01	.96
$Var(\varepsilon_{1j})$	0.2	.00	.98
α_1	0.8	.01	.95
$Var(\alpha_{j1})$	0.3	.01	.93

$$Y_{ij} = Y_{b,j} + Y_{w,ij}$$

$$Y_{b,j} \sim N(0, \Sigma_b)$$

$$[Y_{w,ij} | C_{ij} = k] \sim N(\mu_k, \Sigma_{wk})$$

$$P(C_{ij} = k) = \frac{Exp(\alpha_k + \alpha_{jk})}{\sum_{s=1}^{K} Exp(\alpha_s + \alpha_{js})}$$

where C_{ij} is a latent class variable, $Y_{b,j}$ is the cluster-level random effect, and μ_k, Σ_{wk}, and Σ_b are unconstrained mean and variance covariance parameters.

One simple reason to be interested in this model is that the model is the saturated two-level mixture model. Thus any two-level mixture model is nested within this model and can be compared to this model to detect misfit. The model is also of interest in the case of observed latent classes; that is, the two-level multiple group model when the grouping variable is a within-level grouping variable. Such a model would require numerical integration if estimated with ML in M*plus* (although theoretically it is not needed), making the estimation prohibitive in multivariate settings.

This model is one of the seven multiple group multilevel models discussed in Asparouhov and Muthén (2012) intended to explore the various possible relationships between cluster and group effects, which can be used to determine, for example, if cluster effects are equal or unequal in the different groups. Several other models from Asparouhov and Muthén (2012) that were difficult to estimate in multivariate settings will also be easily accessible within this multilevel mixture framework based on the Bayesian estimation. With the Bayesian estimation this unrestricted model can also easily include categorical variables and thus two-level LCA with conditional dependence can be estimated similar to the single-level model described in Asparouhov and Muthén (2011).

We illustrate the unrestricted two-level mixture model with a small simulated example. Consider a two-class mixture model where the latent class variable is measured by (3) continuous indicators. The entries of the within-level variance covariance matrix for class k are denoted by σ_{ijk}, and the mean of the ith variable in class k by μ_{ik}. We generate data using the following parameters: $\mu_{i1} = -1$ and $\mu_{i2} = 1$; the within-level variances in Class 1 are 1 and the three covariance values are 0.2, 0.3, 0.3; the within-level variances in Class 2 are 0.6 and the covariance vaues are 0.3, 0.4, 0.2; and the between-level variances are 1 and the covariances are 0.3, 0.4, 0.1. The class distribution parameters are $\alpha_1 = 0.8$ and $Var(\alpha_{j1}) = 0.5$. The sample consists of 200 clusters of size 50. We generate and analyze 100 data sets. Table 2 contains the results of the simulation for a selection of the parameters. Each replication takes about 14 seconds to complete. It is

TABLE 2
Simulation Results for an Unrestricted Two-Level Mixture Model

Parameter	*True Value*	*Abs. Bias*	*Coverage*
σ_{111}	1	.00	.94
σ_{121}	0.2	.00	.93
μ_{11}	–1	.00	.96
σ_{112}	0.6	.00	.96
σ_{122}	0.3	.00	.95
μ_{12}	1	.00	.97
α_1	0.8	.01	.95
$Var(\alpha_{j1})$	0.5	.01	.90

clear from the results that the parameter estimates are unbiased and the coverage is near the nominal level of 95%.

Example 3: Multilevel Latent Transition Analysis with Cluster-Specific Transition Probabilities

Our primary interest in multilevel latent transition analysis stems from the idea that to be able to model latent class autocorrelation across time, we have to develop as a building block the relationship between two latent class variables in multilevel settings. Once such a building block is established, we can use it in the intensive longitudinal setting as the model relating two consecutive latent class variables, thus accounting for the sequential dependency there might be between the states (i.e., classes) a person is in at consecutive time points. In this section, however, we still postpone the discussion of intensive longitudinal data and we shall consider the multilevel level taltent transition analysis model in its own right, focusing on more traditional panel data (consisting of a small number of repeated measures).

In Asparouhov and Muthén (2008) a two-level latent transitional model is discussed and estimated with the ML estimation method. The latent transition analysis is based on the following example: Students are nested within schools and are classified in (2) classes at two separate occasions. We are interested in how the transition probability $P(C_2|C_1)$ varies across schools, where C_1 and C_2 represent the latent class variables at these separate occasions. Figure 1 gives a graphical representation of this model, but the precise model used in Asparouhov and Muthén (2008) is given by the following equations:

$$P(C_{1,ij} = c) = \frac{Exp(\alpha_{1jc})}{\sum_{k=1}^{K} Exp(\alpha_{1jk})} \quad (11)$$

$$P(C_{2,ij} = d|C_{1,ij} = c) = \frac{Exp(\alpha_{2jd} + \gamma_{cd})}{\sum_{k=1}^{K} Exp(\alpha_{2jk} + \gamma_{ck})} \quad (12)$$

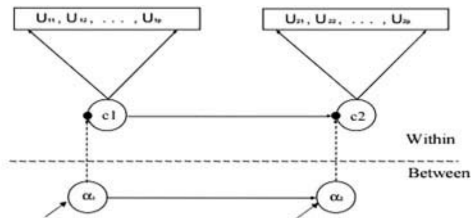

FIGURE 1 Two-level latent transition model.

where for identification purposes $\alpha_{1jK} = \alpha_{2jK} = \gamma_{cK} = \gamma_{Kd} = 0$, as usual.

Consider the simple case of $K = 2$ classes. The parameter γ_{11} represents the regression effect from C_1 to C_2 and gives a way for C_1 to affect the distribution of C_2. Because that is a fixed coefficient, however, the model has just two random effects α_{1j1} and α_{2j1}, which implies that the ML estimator would use two-dimensional numerical integration to estimate this model. In contrast, the joint distribution of the two binary latent variables has 3 *df* and to be able to fit that distribution for every cluster there should really be three random effects.

Here we propose a new two-level multilevel latent transition analysis model that resolves this problem. The Bayesian framework can easily accommodate any number of random effects and thus we can easily estimate the full (3) random effects model that is needed for the two-class situation. The new model is given by the following equations:

$$P(C_{1,ij} = c) = \frac{Exp(\alpha_{jc})}{\sum_{k=1}^{K} Exp(\alpha_{jk})} \quad (13)$$

$$P(C_{2,ij} = d|C_{1,ij} = c) = \frac{Exp(\alpha_{jcd})}{\sum_{k=1}^{K} Exp(\alpha_{jck})} \quad (14)$$

where again for identification purposes $\alpha_{jK} = \alpha_{jcK} = 0$. In the 2-class example the three random effects are α_{j1}, α_{j11}, and α_{j21} and they can be reformulated as follows

$$\alpha_{j1} = log\{P(C_1 = 1|j)/P(C_1 = 2|j)\} \quad (15)$$

$$\alpha_{j11} = log\{P(C_2 = 1|j, C_1 = 1)/P(C_2 = 2|j, C_1 = 1)\} \quad (16)$$

$$\alpha_{j21} = log\{P(C_2 = 1|j, C_1 = 2)/P(C_2 = 2|j, C_1 = 2)\}. \quad (17)$$

This new multilevel latent transition analysis model is sufficiently flexible to be able to fit any cluster-specific

TABLE 3
Simulation Results for a Multilevel Latent Transition Analysis

Parameter	*True Value*	*Abs. Bias*	*Coverage*
$E(Y_1 \mid C_1 = 1)$	1	.01	.98
$E(Y_1 \mid C_1 = 2)$	–1	.01	.99
$Var_w(Y_1)$	1	.00	.96
$Var_b(Y_1)$	1	.04	.94
$E(\alpha_{1j})$	.5	.00	.95
$Var(\alpha_{1j})$	.05	.00	.92
$E(\alpha_{11j})$	-.5	.00	.94
$Var(\alpha_{11j})$	.05	.02	.89

transition model. The model defined in Equations 11 and 12 is equivalent to the model defined in Equations 13 and 14, if and only if $Var(\alpha_{j21} - \alpha_{j11}) = 0$; that is, if the difference between these two random effects is a constant (equal to γ_{11}) independent of the cluster *j*. Note also that, as with any other between-level effects, the random effects of the transition probabilities can be regressed on between-level predictors.

The estimation algorithm for the multilevel ltaent transition analysis model is again the MCMC algorithm. For the update of the random effects α_{jcd} and α_{jc} we use a Metropolis step similar to the one latent class variable case. Note here that if cluster sizes are smal, the joint latent class distribution tables will have empty cells that will lead to logits of infinity; that is, the random effects will have arbitrary large values that in turn will result in biases and overestimation for $Var(\alpha_{jcd})$. We recommend cluster sizes with at least 50 observations to avoid this problem.

We illustrate the multilevel latent transition analysis model with the following simulation study. Each of the two latent class variables are measured by (4) continuous indicators with means in Class 1 set to 1, means in Class 2 set to –1, and within-level and between-level variances for the indicators all set to 1. The means of the random effects from Equations (15 through 17) are 0.5, –0.5, 1, and the variances are set to 0.05. We generate and analyze 100 data sets using 100 clusters of size 50. Table 3 shows the results of the simulation study for a subset of the parameters. The parameter estimates show almost no bias and the coverage is near the nominal level of 95%.

THE HIDDEN MARKOV MODEL

In this section we discuss the single-level HMM. We use a single-level model to simplify the discussion. From a practical point of view, the single-level HMM-3 can be used to analyze time series data from a single person; see, for example, Hamaker and Grasman (2012) and Hamaker, Grasman and Kamphuis (2016). Analyzing data from a single person, or analyzing data separately for each individual in a replicated time series design, has the advantage that different models can be used for different individuals, that is, the best fitting model can be different for different individuals and can be used to identify person-specific dynamics that might be characteristic of, for instance, a psychological disorder. Such an idiographic approach has been advocated for decades in psychology (cf. Molenaar, 2004). The advantage of analyzing a sample of individuals from the entire population, rather than a single person, is that information is accumulated and borrowed to obtain more stable estimates. In addition, when we analyze a sample of the population, we can make inference for the entire population, wehereas when we analyze a single person we can make inference only about the future behavior of that one person. In this section we focus on the time series data for a single person and in particular the single-level HMM.

The HMM has two parts: a measurement part and a Markov switching part. The measurement part is like any other mixture model. It is defined by $P(Y_t \mid C_t)$, where Y_t is a vector of observed variables and C_t is the latent class or state variable at time *t*, which takes on values $1, \ldots, K$. The Markov switching (or regime switching) part is given by the transition matrix $P(C_t \mid C_{t-1})$, which allows us to correlate the latent class variable over time with itself. In single-level models we use the transition matrix directly as model parameters so that we can use the Dirichlet conjugate priors for these parameters and avoid the Metropolis step in MCMC. The size of the transition matrix $Q = P(C_t \mid C_{t-1})$ is K by K, but because the columns add up to 1, the number of independent parameters in the Markov part of the model is $K(K-1)$.

In the HMM model p-$P(C_t)$ is not a model parameter. These marginal probabilities are implicitly modeled and represent the stationary distribution of C_t, that is, the distribution of C_t as *t* increases to infinity. It can be obtained implicitly from the stationary assumption that $P(C_t)$ is independent of *t*, that is, from the equation $Qp = p$. Because the first $K-1$ equations in this linear system added up give the last equation, the rank of the matrix Q is $K-1$ and thus the linear system $Qp = p$ alone cannot be used to solve for *p*. The most common way to solve for *p* is to replace the last equation in that linear system with the equation $p_1 + \ldots + p_K = 1$.

The HMM we consider here is essentially an autoregressive model of order 1, meaning that the state variable C_t affects the state variable in the next period C_{t+1} but it does not have a direct effect on C_{t+2}; that is, C_t only affects C_{t+2} indirectly through the value of C_{t+1}.

To estimate the HMM we modify the latent class updating step of the MCMC mixture estimation algorithm given in Asparouhov and Muthén (2010). In the HMM estimation the latent class variables are updated sequentially. We first update C_1 given the conditional distribution $P[C_1 \mid *]$, conditional on everything else, including the latent class variable at all other times. We then update C_2 from the conditional distribution $P[C_2|*]$, and so on. It turns out that the conditional distribution of C_t depends only on C_{t-1}, C_{t+1} and the observed class indicators Y_t at time t. Using the transition matrix $Q = P(C_t|C_{t-1})$ we first compute

$$P(C_t = k|C_{t-1}, C_{t+1}) = \frac{P(C_{t+1}|C_t = k)P(C_t = k|C_{t-1})}{\sum_{k=1}^{K} P(C_{t+1}|C_t = k)P(C_t = k|C_{t-1})}$$

and then we use that to compute the posterior distribution for C_t

$$P(C_t = k|C_{t-1}, C_{t+1}, Y_t) = \frac{P(Y_t|C_t = k)P(C_t = k|C_{t-1}, C_{t+1})}{\sum_{k=1}^{K} P(Y_t|C_t = k)P(C_t = k|C_{t-1}, C_{t+1})}.$$

For completeness we have to specify how we treat the initial condition C_0. Just like in DSEM, we treat that as an auxiliary parameter that can have its own Dirichlet prior distribution. The prior can be prespecified or it can be automatically determined by the algorithm based on the distribution of C_t obtained during a burn-in period.

The updating of the transition matrix Q in the MCMC estimation is straightforward. Let n be the matrix of current frequencies; that is, n_{ji} is the number of time periods t for which $C_{t-1} = i$ and $C_t = j$. Consider the updating of the ith column of the transition matrix $q_i = P(C_t|C_{t-1} = i)$. If the prior of q_i is the Dirichlet distribution $D(r_i)$ then the posterior distribution $[q_i|C_t]$ is the Dirichlet distribution $D(r_i + n_i)$ where n_i is the ith column of n.

We illustrate the HMM with the following simulation study. Consider a two-class HMM where each class is measured by three binary variables, $P(U_{pt} = 0|C_t = 1) = \Phi(-1)$ and $P(U_{pt} = 0|C_t = 2) = \Phi(1)$; that is, the threshold parameters in Class 1 are $\tau_{p1} = -1$ and in Class 2 are $\tau_{p2} = 1$. The transition matrix is specified as follows $q_{11} = P(C_t = 1|C_{t-1} = 1) = 0.9$ and $q_{12} = P(C_t = 1|C_{t-1} = 2) = 0.25$. These are all the parameters in this model: three thresholds in each class and two parameters in the transition matrix, making for a total of eight parameters. Using the method described earlier, above one can compute the marginal distribution of C_t, $P(C_t = 1) = 5/7$. We generate and analyze 100 samples of size 1,000. To avoid dependence on the initial value, we start with the first class being 1, but we discard the first 10 observations of the sample. It takes 1 second on average to estimate this model for each sample. The results of the simulation are presented in Table 4 for several of the model parameters. The parameter estimates are unbiased and the coverage is satisfactory.

TABLE 4
Simulation Results for a Hidden Markov Model

Parameter	*True Value*	*Abs. Bias*	*Coverage*
τ_{11}	–1	.00	.95
τ_{12}	1	.00	.91
q_{11}	0.9	.00	.94
q_{12}	0.25	.01	.97

Various other interesting models can be estimated within the single-level time series mixture framework in M*plus*. Single-level DSEM models can be combined with a single-level HMM to obtain a more rich set of examples. Several such examples are presented in Hamaker et al. (2016).

For the remainder part of this article we return to the framework of twolevel mixture models where the cluster is an individual and the observations within the cluster are the time series data for that individual, because we are now finally in a position to present a general latent class analysis models for intensive longitudinal data.

DYNAMIC LATENT CLASS ANALYSIS

In this section we give a complete definition of the general dynamic latent class analysis DLCA model. However, this definition is nothing more than a combination of the ideas presented in the previous sections; that is, the DSEM mixture model, the single-level HMM, and the multilevel latent transition model. Figure 2 summarizes how the different modeling ideas are combined to arrive at the DLCA model.

We begin with the decomposition of the observed variable Y_{it} of individual i at time t into a within and a between component; that is,

$$Y_{it} = Y_{1,it} + Y_{2,i}. \quad (18)$$

Let S_{it} be a latent class or state variable for individual i at time t. The model for $Y_{1,it}$ is a class-specific DSEM model

$$[Y_{1,it}|S_{it} = s] = \nu_{1,s} + \sum_{l=0}^{L} \Lambda_{1,l,s}\eta_{i,t-l} + \varepsilon_{it} \quad (19)$$

$$[\eta_{i,t}|S_{it} = s] = \alpha_{1,s} + \sum_{l=0}^{L} B_{1,l,s}\eta_{i,t-l} + \Gamma_{1,s}x_{it} + \xi_{it}. \quad (20)$$

If Y_{ijt}, that is, the jth variable of the observed vector Y_{it}, is a categorical variable

$$Y_{ijt} = m|S_{it} = s \Leftrightarrow \tau_{m-1js} \leq Y^*_{ijt} < \tau_{mjs} \quad (21)$$

The latent class variable S_{it} follows a Markov switching model with subject-specific transition probability

$$P(S_{it} = d|S_{i,t-1} = c) = \frac{Exp(\alpha_{idc})}{\sum_{k=1}^{K} Exp(\alpha_{ikc})} \tag{22}$$

where α_{idc} are subject-specific random effects. For identification purposes $\alpha_{iKc} = 0$.

Finally, at the between level we have

$$Y_{2,i} = \nu_2 + \Lambda_2\eta_{2,i} + \varepsilon_{2,i} \tag{23}$$

$$\eta_{2,i} = \alpha_2 + B_2\eta_{2,i} + \Gamma_2 x_{2,i} + \xi_{2,i} \tag{24}$$

where $\eta_{2,i}$ contains all subject-specific random effects, including the random transition probabilities effects α_{idc}, as well as all random intercepts, loadings, and slopes.

We estimate the model with MCMC where the estimation is nothing more than combing the updating steps described earlier.

DLCA EXAMPLES

In this section we illustrate the DLCA model with three simulation studies: a simple two-class DLCA model, the MMSAR, and DLCA model with regime switching for the latent factor.

Example 1: Two-Class DLCA

In this section we illustrate the DLCA with a simple two-class simulation example. The sample consists of 200 individuals, with 100 observation times each, where the latent class variable at each time point is measured by 4 binary indicators. We allow the model to be subject specific; that is, we allow the threshold parameters for the class indicators to vary slightly across individuals

TABLE 5
Simulation Results for the Two-Class Dyniamc Latent Class Analysis

Parameter	*True Value*	*Abs. Bias*	*Coverage*
τ_{11}	1	.00	.97
τ_{12}	-1	.00	.99
θ_1	0.1	.00	.92
α_1	1	.00	.96
σ_1	0.05	.01	.89

$$P(U_{pit} = 0|S_{it} = k) = \Phi(\tau_{pk} + \varepsilon_{ip})$$

where $\varepsilon_{ip} \sim N(0, \theta_p)$ is a subject-specific random effect for each indicator. In this simulation we use $\theta_p = 0.1$, $\tau_{p1} = 1$, and $\tau_{p2} = -1$. The latent class Markov switching model is given by

$$p_{i1} = P(S_{it} = 1|S_{it-1} = 1) = \frac{Exp(\alpha_{i1})}{1 + Exp(\alpha_{i1})}$$

$$p_{i2} = P(S_{it} = 1|S_{it-1} = 2) = \frac{Exp(\alpha_{i2})}{1 + Exp(\alpha_{i2})}$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j)$$

The state transition matrix is

$$\begin{pmatrix} p_{i1} & p_{i2} \\ 1 - p_{i1} & 1 - p_{i2} \end{pmatrix}.$$

In this simulation $\alpha_1 = 1$, $\alpha_2 = -0.5$, and $\sigma_j = 0.05$. This model has 17 parameters: Each indicator variable has one threshold in each class and a between-level variance for the subject-specific threshold deviation, making for a total of 12 parameters, plus the 5 parameters for mean and variance covariance for the transition matrix random effects α_{i1} and α_{i2}.

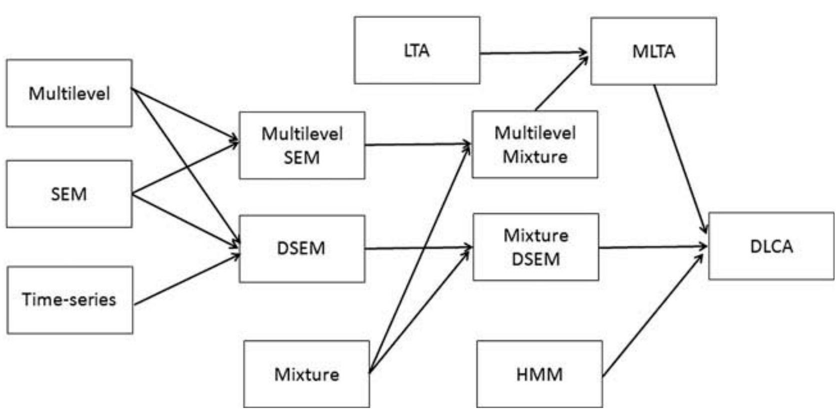

FIGURE 2 Arriving at the dynamic latent class analysis model.
Note. SEM=structural equation modeling; DSEM=dynamic structural equation modeling; LTA=latent trait analysis; MLTA=multilevel latent trait analysis; HMM=hidden markov model; DLCA=dynamic latent class analysis.

TABLE 6
Simulation Results for a Multilevel Markov Switching Autoregressive Model

Parameter	*True Value*	*Abs. Bias*	*Coverage*
μ_1	1	.00	.91
β_1	0.4	.01	.94
θ_1	0.9	.01	.93
α_1	1	.02	.92
σ_1	0.05	.01	.96
σ	1	.03	.96

We simulate and analyze 100 samples. It takes about 20 minutes to estimate this model for each data set. The results of the simulation for a subset of the parameters are given in Table 5. The estimates are unbiased and the coverage is near the nominal level of 95%.

Example 2: Multilevel Markov Switching Autoregressive Model

The markov switching autoregressive model (MSAR) was used in Hamaker et al. (2016) to analyze bipolar disorder using data from individual patients. Here we consider the multilevel version of this model that can be used to analyze not just a single patient, but an entire sample. The difference between this model and the model described in the previous section is that the autoregressive effect can be estimated not just for the latent state variable, but also directly for the latent state indicator. An MMSAR model with two regimes can be summarized with the following equations

$$Y_{it} = Y_{2,i} + Y_{1,it}$$

$$Y_{1,it} = \mu_{S_{it}} + \beta_{S_{it}} Y_{1,it-1} + \varepsilon_{it,S_{it}}$$

$$P(S_{it} = 1 | S_{it-1} = j) = \frac{Exp(\alpha_{ij})}{1 + Exp(\alpha_{ij})}$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j), Y_{2,i} \sim N(0, \sigma).$$

We conduct a simulation study using a sample with 100 individuals and 100 observations for each individual. There are 11 parameters in this model. For each class we have μ_k, β_k, and $\theta_k = Var(\varepsilon_{itk})$ for a total of six parameters. The remaining five parameters are α_k and σ_k for each of the two classes as well as the between-level variance parameter σ. The following parameter values were used for the data generation: state-specific means $\mu_1 = 1$ and $\mu_2 = -1$, autoregressive parameters $\beta_1 = 0.4$ and $\beta_2 = 0.2$, residual variances $\theta_1 = 0.9$ and $\theta_2 = 0.7$, mean logits $\alpha_1 = 1$ and $\alpha_2 = -0.5$, between variance $\sigma = 1$, and variance of logits random effects $\sigma_j = 0.05$. We generate and analyze 100 samples. It takes about 15 minutes on average to estimate one replication. The results for some of the parameters are presented in Table 6. The point estimates show almost no bias and the coverage is near the nominal level.

It is interesting to note in this model that the latent class variable appears to have just one indicator. However, due to the time series nature of the model, the latent class variable stays the same in time segments (i.e., for several consecutive time points), and thus one can assume that the latent class variable is measured not just by the observation at that point, but also, albeit to a smaller extent, by measurements at the neighboring time points.

The model in this section illustrates a time series model for one continuous variable. As time progresses the variable switches between two regimes (also referred to as states or classes); one of these regimes is characterized by a high average, and the other is characterized by a low average over time. Sometimes such models are referred to as regime switching models (cf. Kim & Nelson, 1999). Consider the distribution of Y_{it} for a fixed i. This distribution is bimodal due to the two classes. The observed sequence $Y_{i1}, Y_{i2}, Y_{i3}, \ldots$ is a nonindependent sample from that distribution, a sample where consecutive observations are correlated but nevertheless the sample will reliably reproduce the bimodal distribution of Y_{it} as t increases. Using this point of view, a bimodal distribution for a variable Y_{it} can be considered indisputable evidence for regime switching behavior. We should note here also that this line of argument goes beyond bimodal distributions. Many mixtures of normals do not result in bimodal distributions but simply in nonnormal or heavy tail distributions. Thus just like with cross-sectional Mixture models (cf. Bauer & Curran, 2003), nonnormality in the distribution can be viewed and interpreted as evidence for a regime switching model and vice versa. Regime switching models can be nothing more than nonnormality in the distribution. Proper substantive interpretation is imperative for regime switching models, and pure statistical evidence should not be used without substantive interpretation.

Example 3: Regime Switching for the Latent Factor

The regime switching model described in the previous section for an observed variable can also be estimated for a latent factor. In psychological studies often the main variable of interest is a latent variable measured through a factor analysis model. Consider again the hypothetical example from the introduction section where we monitor/measure a latent variable, and classify patients into one of 3 regimes: healthy, increased risk of relapse, and relapse. The regime switching can occur directly on the latent variable. In this section we present such a simulation example. Consider the following model where Y_{pit} is an observed variable $p = 1, \ldots, 4$ for individual i at time t measuring a

TABLE 7
Simulation Results for the Regime Switching for the Latent Factor Model

Parameter	*True Value*	*Abs. Bias*	*Coverage*
ν_1	0	.01	.96
$\theta_{w,1}$	1	.00	.95
$\theta_{b,1}$	1	.04	.93
λ_2	1.2	.00	.95
μ_2	2.5	.01	.94
β_2	0.2	.00	.92
θ_2	0.8	.02	.88
α_1	1	.00	.90
σ_1	0.05	.02	.94

latent variable η_{it}, which follows a two-class regime switching model

$$Y_{pit} = \nu_{pi} + \lambda_p \eta_{it} + \varepsilon_{pit}$$

$$\eta_{it} = \mu_{S_{it}} + \beta_{S_{it}} \eta_{it-1} + \xi_{it,S_{it}}$$

$$P(S_{it} = 1 | S_{it-1} = j) = \frac{Exp(\alpha_{ij})}{1 + Exp(\alpha_{ij})}$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j)$$

$$\nu_{pi} \sim N(\nu_p, \theta_{b,p})$$

$$\xi_{itk} \sim N(0, \psi_k)$$

$$\varepsilon_{pit} \sim N(0, \theta_{w,p}).$$

This model has six random effects, four random intercepts ν_{pi} for the four observed variables and two random effects α_{ij} used in the latent transition matrix. The model we consider here has 24 parameters; for each observed variable we have the within-level residual variance $\theta_{w,p} = Var(\varepsilon_{pit})$, the between-level residual variance $\theta_{b,p}$, the nonrandom intercept ν_p, and the loading parameter λ_p, for a total of 15 parameters, as the first variable loading is fixed to 1. The remaining nine parameters are as follows: α_j and σ_j account for four parameters, $\psi_k = Var(\xi_{itk})$, β_k, μ_k for the two classes give another six parameters, but for identification purposes we fix $\mu_1 = 0$. We generate data using 100 individuals observed at 100 time points, using the following parameter values: $\theta_{w,p} = 1$, $\theta_{b,p} = 1$, $\nu_p = 0$, $\lambda_2 = 1.2$, $\lambda_3 = 0.8$, $\lambda_4 = 0.8$, $\alpha_1 = 1$, $\alpha_2 = -0.5$, $\sigma_j = 0.05$, $\psi_1 = 1$, $\psi_2 = 0.8$, $\mu_2 = 2.5$, $\beta_1 = 0.4$, $\beta_2 = 0.2$. The entropy of the model is 0.8. We generate and analyze 100 replications. Each replication takes about 15 minutes to complete. The results of the simulation for a selection of the parameters is presented in Table 7. The parameter estimates show little or no bias and the coverage of the parameters is satisfactory.

DISCUSSION

DLCA modeling presented in this article is a broad and flexible framework that allows for a wide variety of single and multilevel models for intensive longitudinal data. It is based on combining three innovations in M*plus* Version 8: (a) DSEM, which allows for multilevel modeling based on time series analysis; (b) Bayesian estimation of multilevel mixture models, which makes it possible to have large numbers of random effects; and (c) the multilevel HMM, which allows for person-specific switches between different classes (also referred to as states or regimes in time series). By combining these features into one framework, we can now specify a random effects model that allows for individual differences in the transition probabilities of the HMM as well as in the parameters of the different time series models that describe the within-person dynamics within each of the regimes. In addition to these new modeling features for intensive longitudinal data, we have also presented a number of other new latent class modeling options for cross-sectional and panel data; these latter options have resulted from the innovations that were needed to develop DLCA.

DSEM modeling in M*plus* is somewhat more general than what is described earlier; see Asparouhov et al. (2016). For example, lag variable modeling can be done not just for the latent continuous variables but also for the observed variables. Other modeling features incorporated in the DSEM framework are log-normal modeling distribution for within-level variances, cross-sectional modeling where time-specific effects are modeled as well, and unequal and subject specific-times of observations. All of these features apply also to the mixture modeling framework and can improve the feasibility of the models in practical settings. Missing data can easily be handled in the MCMC estimation framework. This is important also because when observations are taken at different times for different individuals, missing data are used to align the times of observations between individuals. Factor scores for all random effects and latent variables are easily obtained because these are naturally generated within the MCMC estimation. Inputs and outputs for all of the simulation studies presented here are available online at statmodel.com.

Further work is needed in the area of model comparison. One possibility is to compute deviance information criterion (DIC) for these models; however, due to the nonindependence on the within-level latent variables the marginal likelihood is difficult to compute. It is possible to integrate out all within-level latent variables, but this leads to a large number of model parameters, which would require a large number of MCMC iterations to produce accurate results. It is not unusual that when trying to use DIC for such models that the values of competing models are too close and the precision of DIC too low to be able to meaningfully use DIC for comparison. The most straightforward way to compare models is via the credibility intervals for parameter estimates.

Many of the time series models we discussed earlier are considered stationary models; that is, models that stabilize over time. In many practical settings this is not realistic. For example we might be interested in models where the transition probabilities change over time while still remaining subject specific. The easiest way to break through the stationarity assumption is to introduce predictors in the model that change over time. With the change in the predictors the model can accommodate nonstationary models. The covariates and predictors can be as simple as the time variable itself; in fact, standard growth modeling essentially uses only the time variable as the main predictor for modeling change over time. Other nonstationary covariates can also be utilized. In fact for many covariates, the stationarity assumption is just as unrealistic to assume as it is for the dependent variables. An alternative approach is to utilize time-varying effects models or cross-classified modeling where time-specific random effects are utilized, as in the DSEM framework.

The models we discussed in this article use a latent class variable that changes over time; however, time-invariant latent class variables are also of interest. The combination of two latent class variables, a time-invariant latent class variable and a time-varying latent class variable, are also of interest. In Asparouhov and Muthén (2008) it is shown that the time-invariant latent class variables are a special case of the time-varying latent class variables. As the variance of the logits random effects increases to infinity the latent class variable becomes time invariant. Thus this framework can be used to estimate such models as well. For example, estimating a model with a binary time-invariant latent class variable and a binary time-varying latent class variable is equivalent to estimating a four-class model with two of the logits random effects having large variances.

REFERENCES

Altman, R. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, *102*, 201–210.

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G. R. Hancock, & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*, (pp. 27–51). Charlotte, NC: Information Age.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Technical Report, Version 3. Retrieved from http://statmodel.com/download/Bayes3.pdf

Asparouhov, T., & Muthén, B. (2011). Using Bayesian priors for more flexible latent class analysis. In *Proceeedings of the Joint Statistical Meeting* (pp. 4979–4993). Miami Beach, FL.

Asparouhov, T., & Muthén, B. (2012). *Multiple group multilevel analysis*. M*plus* Web Notes: No. 16. Los Angeles, CA: Muthén: Muthén.

Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In *Advances in multilevel Modeling for educational research: Addressing practical issues found in real-world applications*. J. Harring, L. Stapleton, S. Beretvas (Eds.), (pp. 155–182). Charlotte, NC: Information Age.

Asparouhov, T., Hamaker E., & Muthén, B. (2016). Dynamic structural equation models. Technical Report. In preparation.

Bauer, D., & Curran, P. (2003), Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363.

Davidov, E., Dulmer, H., Schluter, E., & Schmidt, P. (2012), Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*, 558–575.

De Jong, M. G., Steenkamp, J.-B. E. M., & Fox, J.-P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, *34*, 260–278.

De Jong, M. G., Steenkamp, J.-B. E. M. (2009) Finite mixture multilevel multidimensional ordinal IRT Models for large scale cross-cultural research, *psychometrika*, *75*, 3–32.

Hamaker, E. L., & Grasman, R. P. (2012) Regime switching state-space models applied to psychological processes: Handling missing data and making inferences, *Psychometrika*, *77*, 400–422.

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2016). Modeling BAS dysregulation in bipolar disorder illustrating the potential of time series analysis. *Assessment*, *23*, 436–446.

Kim, C. & Nelson, C. (1999). State-space models with regime switching. cambridge, M-A: MIT Press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*, 35–45.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement*: *Interdisciplinary Research and Perspectives*, *2*, 201–218.

Patterson, B., Dayton, M., and Graubard, B. (2002). Latent class analysis of complex sample survey data: Application to dietary data, *Journal of the American Statistical Association*, *97*, 721–729.

Trull, T., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science, *Current Directions in Psychological Science*, *23*, 466–470.

Zhang, Z., and Nesselroade, J. (2007). Bayesian estimation of categorical dynamic factor models, *Multivariate Behavioral Research*, *42*, 729–756.

Zhang, Z., Hamaker, E., and Nesselroade, J. (2008). Comparisons of four methods for estimating a dynamic factor model, *Structural Equation Modeling*, *15*, 377–402.