Statistics and Its Interface Volume 6 (2013) 27–44

A Bayesian analysis of generalized latent curve mixture models

Jun-Hao Pan*, Xin-Yuan Song†,‡ and Edward H. Ip

Latent curve models for longitudinal data have received increasing attention in medical, educational, psychological, and behavioral sciences. In these applied areas of research, heterogeneous longitudinal data are common. This paper proposes the use of generalized latent curve models for analyzing heterogenous longitudinal data. The basic model features a mixture of trajectories. It also employs a multinomial logit model for assessing the influence of fixed covariates and explanatory latent variables on the class membership probability within the mixture model. This broad class of models also handles non-normal data from the exponential family distributions. A Bayesian approach is implemented for data analysis. We report a simulation study that proves the satisfactory performance of the proposed approach. Furthermore, we analyzed a real data set extracted from the National Longitudinal Survey of Youth to illustrate the practical value of the proposed model and methodology.

Keywords and phrases: Latent curve mixture models, Heterogeneous longitudinal data, Markov chain Monte Carlo method, Modified deviance information criterion.

1. INTRODUCTION

Longitudinal data, which comprise repeated measurements of the same individuals on different periods, arise frequently in a wide range of fields, such as psychology, education, medicine, and public health. Latent curve modeling (LCM) [3, 23, 24, among others], which applies structural equation modeling (SEM) concepts and techniques to traditional growth curve analyses, has received increasing attention as a useful longitudinal technique in the analysis of change patterns. LCM can incorporate information about group and individuals. More importantly, LCM can be used to analyze and explain changes. For example, LCM relates growth factors to an individual's contextual data by extracting latent growth factors from individual trajectories. Based on the pioneer work of Meredith and Tisak [23, 24], several scholars demonstrated that LCM is a useful tool for studying developmental trends from both inter- and intra-individual perspectives [see, for example, 3, 8, 22, 26, 27, 29, 37, 41, among others].

Heterogeneity is an important issue in the analysis of longitudinal data. Statistical inference would be seriously distorted when a heterogeneous population is analyzed as homogeneous. Observed heterogeneity can be captured by fixed-effects models, including multi-group models (e.g., age groups). However, other types of heterogeneities such as heterogeneous trajectory classes cannot be handled by fixed-effects models because an individual's membership in a trajectory class is unknown in advance. In statistics, a discrete or continuous mixing distribution is often used to model heterogeneity. When a discrete mixing distribution is deployed, the model is often termed a finite mixture model. In longitudinal studies, the usefulness of mixture models has been increasingly recognized for identifying meaningful classes of individuals according to their developmental trends. Recently, Muthén et al. [see, for example, 26, 28, among others] extended the classical LCM to finite-mixture LCMs and proposed a growth mixture modeling (GMM) under which a population of interest is regarded as a mixture of classes, each defined by its distinct developmental trajectory. In their models, probabilities of class membership can be directly related to covariates through a multinomial logit model. Hence, the model for membership belonging to a specific trajectory class is enhanced with the information derived from the covariates. For instance, if male students are more likely to belong to a trajectory class of faster learners in mathematics, then the predictive power of the mixture model can be improved by including gender as a covariate.

There are two significant methodological challenges for applying GMMs to real data analysis, especially in the social and behavioral sciences. First, explanatory variables are often measured indirectly. Thus, the concept of a latent explanatory variable is useful when multiple indicators are used to describe a specific construct. Second, in practice, the makeup of multiple indicators for a specific construct is almost always multimodal. For example, binary and continuous (including normal and non-normal distributions) data could be present. Sometimes, ad hoc methods are used to

*Jun-Hao Pan's research is fully supported by grants from the National Natural Science Foundation of China (NSFC 11026192 and NSFC 91124004). This research is also supported by the Fundamental Research Funds for the Central Universities and the high-performance grid computing platform of Sun Yat-sen University.

†Prof. Xin-Yuan Song's research is fully supported by grants (GRF 446609 and GRF 404711) from the Research Grant Council of the Hong Kong Special Administration.

‡Corresponding author.

create single-mode data. For example, cutoffs are set for continuous data. However, direct models for mixed-mode data would greatly increase the flexibility of the analytic tool. To the best of our knowledge, limited work has been conducted to fully address these practical issues within the context of GMMs.

To illustrate the aforementioned statistical challenges, we used an example from the National Longitudinal Surveys of Youth (NLSY79) in the U.S. This example was also used to illustrate our proposed solution. NLSY79 is a nationally representative sample of young men and women 14 to 22 years of age at the time of the first survey in 1979. These individuals were interviewed annually until 1994 and are currently interviewed on a biennial basis. Their scores for the Peabody Individual Achievement Tests (PIATs) [10] in mathematics and other subjects were recorded at each period. The data analysis was conducted to investigate the longitudinal behaviors of the mathematics achievement of young students from 1990 to 1994 and to identify whether different classes of developmental trends exist. The effects of some important determinants on academic achievement were also determined. Unlike gender, which is directly observed, other explanatory variables of growth trajectory such as behavior problems and home environment are indirectly measured using the five indicators of the Behavior Problem Index (BPI) [43] and the three indicators of the Home Observation for Measurement of the Environment (HOME) Inventory [5]. The indicators of these latent constructs do not necessarily belong to a single data type. For example, the latent explanatory variable "home environment" is described by binary variables such as "live with mother" and by continuous variables such as "cognitive stimuli score". Although this example concerns a social survey, the two challenges–explanatory latent variables and mixed-mode indicators–are common in other sciences, including health and business/economics.

In this paper, we propose a generalized latent curve mixture model (GLCMM) to analyze heterogeneous longitudinal data. The proposed approach directly addresses the aforementioned statistical challenges. By building upon existing work in GMM, our proposed model aims to fill the following significant gaps in the literature by (1) incorporating explanatory latent variables for predicting latent growth factors and probability of class membership, and by (2) including a broad class of data types for indicators of the latent variables through the exponential family distribution (EFD) framework. We adopted a Bayesian approach, together with Markov chain Monte Carlo (MCMC) techniques such as the Gibbs sampler [15], the Metropolis-Hastings (MH) algorithm [17, 25], and the permutation sampling [14], to analyze GLCMM. A Bayesian model selection criterion, the modified deviance information criterion (DIC), was utilized for model selection. The sampling-based Bayesian approach is proposed for several reasons. First, the different types of latent variables (categorical and continuous) and complicated data structure make the statistical inference of GLCMM through the maximum likelihood (ML)-based method difficult. By contrast, the sampling-based Bayesian approach is powerful for analyzing complex models and data. Second, the ML-based method is likely to encounter multiple modes in a mixture likelihood, which could lead to uncertainty in inference if all the modes are reached [12]. Our proposed Bayesian approach employs permutation sampling to fix the problem of label switching in a mixture likelihood and to avoid getting trapped in a suboptimal mode. Finally, the Bayesian approach allows the use of genuine prior information and is less dependent on asymptotic theory, thereby producing more reliable results even with small samples [11, 19, 32, 37, among others].

The remainder of this paper is organized as follows. Section 2 describes GLCMM. Section 3 presents a Bayesian analysis of the proposed model. An MCMC method coupled with the permutation sampling is proposed to obtain the Bayesian estimates of the unknown parameters. The modified DIC is used for model selection. Section 4 shows a simulation study for the evaluation of the empirical performance of Bayesian estimation and model selection. Section 5 illustrates the methodology with a real data set from NLSY79. Section 6 concludes the paper with a discussion.

2. GENERALIZED LATENT CURVE MIXTURE MODEL (GLCMM)

2.1 Basic latent curve model

LCMs are popular longitudinal techniques for analyzing individual differences in changing patterns, which usually involve random intercepts and slopes, with each pair forming a different trajectory over time. The basic LCM can be viewed as the following common factor analysis model or trajectory equation by Bollen and Curran [3]:

$$\mathbf{y}_i = \mathbf{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_{yi}, \quad i = 1, \ldots, n \tag{1}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{i,p_1})^T$ is a $p_1 \times 1$ vector of continuous repeated measures at p_1 periods for the i-th individual, $\mathbf{\Lambda}_y$ is a $p_1 \times q_1$ matrix of sequentially known values of the growth curve records, $\boldsymbol{\eta}_i$ is a $q_1 \times 1$ latent growth factor, and $\boldsymbol{\epsilon}_{yi}$ is a $p_1 \times 1$ vector of error measurements, which is assumed to be independent of $\boldsymbol{\eta}_i$ and distributed as $N(\mathbf{0}, \mathbf{\Psi}_y)$ with a diagonal covariance matrix $\mathbf{\Psi}_y$. The pattern of $\mathbf{\Lambda}_y$ can be interpreted as a representation of a particular aspect of change in y_{ij} across p_1 periods. When $q_1 = 2$, Equation (1) for a linear LCM is expressed in the following matrix form:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,p_1} \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{p_1} \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} + \begin{pmatrix} \epsilon_{yi1} \\ \epsilon_{yi2} \\ \vdots \\ \epsilon_{yi,p_1} \end{pmatrix}.$$

The first column of $\mathbf{\Lambda}_y$ is used to define an intercept factor, which represents the initial status of change in y_{ij}. The second column of $\mathbf{\Lambda}_y$ represents known times of measurement. The values of t_j $(j = 1, \ldots, p_1)$ reflect the spacing between measurement periods, and the latent growth factor $\boldsymbol{\eta}_i$ contains the random intercept (initial status) η_{i1} and random slope (rate of change) η_{i2}. The two-factor LCM in this case is specified so that the intercept factor serves as the starting point for any growth across time and the slope factor captures the rate of change in the trajectory over time. The values of t_j $(j = 1, \ldots, p_1)$ can be specified using either fixed value restrictions (e.g. $0, 1, \ldots, p_1 - 1$) that represent a straight-line growth or unspecified value restrictions ($t_1 = 0$ and $t_2 = 1$ are fixed for model identification, and the remaining $t_3, \ldots, t_{p_1}$ are freely estimated) that allow estimation of an optimal pattern of change over measurement periods [24]. However, the method of unspecified value restrictions does not produce a straightforward interpretation of the resulting model parameters; see [3] for more details. In this article, we only consider the fixed value restrictions method.

To examine variabilities across individuals, the latent growth factor $\boldsymbol{\eta}_i$ is further modeled by

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\delta}_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

where $\boldsymbol{\mu}$ is a $q_1 \times 1$ vector of the population mean of latent individual growth factors, $\boldsymbol{\delta}_i$ is a $q_1 \times 1$ random vector of residuals that reflect differences between the mean growth factors and the individual growth factors, and $\boldsymbol{\delta}_i$ is assumed to be distributed as $N(\mathbf{0}, \mathbf{\Psi}_\delta)$ with a diagonal covariance matrix $\mathbf{\Psi}_\delta$. The diagonal elements $\psi_{\delta j}$ measure the degree of diversity in individual latent growth factors from their means.

For instance, in the study of NLSY79 described in Section 1 for investigating the longitudinal behavior of mathematics achievement of young students, we aim to specify an LCM to represent changes in math scores. To obtain a rough idea about the trajectory model, we plotted the average PIAT math scores across all individuals at different periods. Figure 1 shows that a linear trajectory model is adequate. We denote y_{ij} $(j = 1, 2, 3)$ as the PIAT scores [10] in mathematics in 1990, 1992, and 1994, during which the measurements were conducted. A linear LCM for $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})^T$ is

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} + \begin{pmatrix} \epsilon_{yi1} \\ \epsilon_{yi2} \\ \epsilon_{yi3} \end{pmatrix}, \tag{3}$$

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix}, \tag{4}$$

where the known times of measurement are set to 0, 1, and 2 to reflect the equal spacing between measurement periods from 1990 to 1994, and the subject developmental trend of

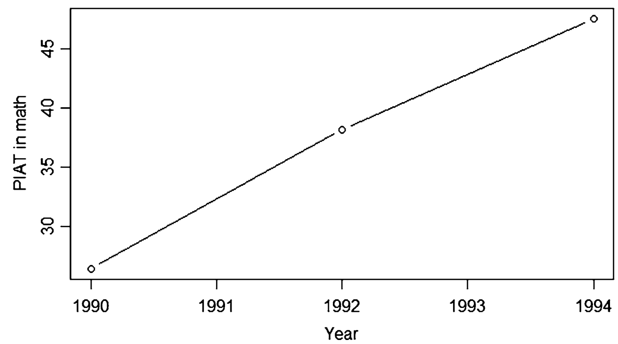

Figure 1. Trajectory of average PIAT math scores across all individuals at three measurement periods.

mathematics score is described by a linear trajectory with the random intercept η_{i1} and random slope η_{i2}.

In some circumstances, a linear LCM may not be flexible enough to model practical situations. The use of a polynomial would help capture a nonlinear trajectory. For instance, a quadratic LCM is defined by

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,p_1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0^2 \\ 1 & 1 & 1^2 \\ \vdots & \vdots & \vdots \\ 1 & p_1 - 1 & (p_1 - 1)^2 \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{pmatrix} + \begin{pmatrix} \epsilon_{yi1} \\ \epsilon_{yi2} \\ \vdots \\ \epsilon_{yi,p_1} \end{pmatrix},$$

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \\ \delta_{i3} \end{pmatrix},$$

where η_{i3} is the quadratic component for revealing the curvature presented in individual trajectories. In practice, a model selection procedure may be conducted to determine whether the inclusion of additional nonlinear terms leads to a significant improvement in model fit.

2.2 Generalized latent curve model

In substantive research, the longitudinal patterns of individuals may be influenced by other observed and latent explanatory variables. Moreover, the latent growth factors in $\boldsymbol{\eta}_i$ may be inter-correlated. To assess these effects and inter-relationships, Equation (2) in the basic LCM can be extended to

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{w}_i + \mathbf{\Pi}\boldsymbol{\eta}_i + \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i, \tag{5}$$

where $\boldsymbol{\mu}$ is a $q_1 \times 1$ vector of intercepts, $\mathbf{w}_i$ is an $r_1 \times 1$ vector of fixed covariates that may follow either continuous or discrete distributions, and $\boldsymbol{\xi}_i$ is a $q_2 \times 1$ vector of explanatory latent variables. $\mathbf{F}(\boldsymbol{\xi}_i) = (f_1(\boldsymbol{\xi}_i), \ldots, f_h(\boldsymbol{\xi}_i))^T$ is an $h \times 1$ vector-valued function that contains nonzero differentiable functions $f_1, \ldots, f_h$, the forms of which are given and may need to be tested. Furthermore, $\mathbf{A}$ and $\mathbf{\Gamma}$ are $q_1 \times r_1$ and $q_1 \times h$ matrices of unknown parameters reflecting the effects of $\mathbf{w}_i$ and $\mathbf{F}(\boldsymbol{\xi}_i)$ on $\boldsymbol{\eta}_i$, and $\boldsymbol{\Pi}$ is a $q_1 \times q_1$ matrix of unknown coefficients specifying the inter-relationships among the latent growth factors in $\boldsymbol{\eta}_i$. We assume that the distribution of $\boldsymbol{\delta}_i$ is $N(\mathbf{0}, \boldsymbol{\Psi}_\delta)$ with a diagonal covariance matrix $\boldsymbol{\Psi}_\delta$ and that $\boldsymbol{\delta}_i$ is independent of $\boldsymbol{\xi}_i$. Let $\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$, in which $\boldsymbol{\Pi}_0$ is assumed to be non-singular and $|\boldsymbol{\Pi}_0|$ is assumed to be independent of the elements of $\boldsymbol{\Pi}$. Li and Wang [21] reported that the violation of this assumption would cause a nonstandard and complex full conditional distribution of $\boldsymbol{\Pi}$, thereby requiring a tedious step for drawing observations from $p(\boldsymbol{\Pi}|\cdot)$ in the MCMC algorithm. In this article, we assume that $|\boldsymbol{\Pi}_0|$ is a constant independent of $\boldsymbol{\Pi}$ for the sake of simple computation. This assumption can be relaxed with a modification of the posterior sampling [21].

In the NLSY79 study, how "behavior problems" influence the developmental trend of PIAT mathematics score is of particular interest. However, "behavior problems, ξ_i" is a latent variable because it is measured by the multiple items of BPI [43]. A model incorporating the effects of ξ_i on the latent growth factors is

$$(6) \qquad \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \xi_i + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix},$$

which is a special case of Equation (5) with $\mathbf{A} = \mathbf{0}$, $\boldsymbol{\Pi} = \mathbf{0}$, and $\mathbf{F}(\boldsymbol{\xi}_i) = \xi_i$. The regression coefficients γ_1 and γ_2 represent the effect of behavior problems on the initial status and rate of change in mathematics score, respectively.

The latent vector $\boldsymbol{\xi}_i$ in Equation (5) can be measured as follows:

$$(7) \qquad \mathbf{x}_i = \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\epsilon}_{xi}, \quad i = 1, 2, \ldots, n,$$

where $\mathbf{x}_i$ is a $p_2 \times 1$ vector of observed indicators, $\boldsymbol{\Lambda}_x$ is a $p_2 \times q_2$ factor loading matrix, and $\boldsymbol{\epsilon}_{xi}$ is a $p_2 \times 1$ random vector of error measurements. Here, $\boldsymbol{\epsilon}_{xi}$ is assumed to be independent of $\boldsymbol{\xi}_i$, $\boldsymbol{\eta}_i$, $\boldsymbol{\epsilon}_{yi}$, and $\boldsymbol{\delta}_i$, and distributed as $N(\mathbf{0}, \boldsymbol{\Psi}_x)$ with diagonal $\boldsymbol{\Psi}_x$, and $\boldsymbol{\xi}_i$ is assumed to be distributed as $N(\mathbf{0}, \boldsymbol{\Phi})$. Note that Models (1) and (7) are confirmatory factor analysis (CFA) models because the patterns of their factor loading matrices can be pre-specified according to background data.

In the study of NLSY79 data, "behavior problems, ξ_i" is measured by the indicators, x_{i1}, x_{i2}, x_{i3}, x_{i4}, and x_{i5}, which are the BPI [43] subscales for antisocial, anxious, peer conflict, headstrong, and hyperactive behaviors, respectively. Equation (7) is then defined with $p_2 = 5$ and $q_2 = 1$:

$$(8) \qquad \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ x_{i5} \end{pmatrix} = \begin{pmatrix} 1.0 \\ \lambda_{x21} \\ \lambda_{x31} \\ \lambda_{x41} \\ \lambda_{x51} \end{pmatrix} \xi_i + \begin{pmatrix} \epsilon_{xi1} \\ \epsilon_{xi2} \\ \epsilon_{xi3} \\ \epsilon_{xi4} \\ \epsilon_{xi5} \end{pmatrix},$$

where 1.0 is fixed to identify the model and to introduce a scale to the latent variable ξ_i. Figure 2 depicts the path diagram of LCM for analyzing the NLSY79 data.

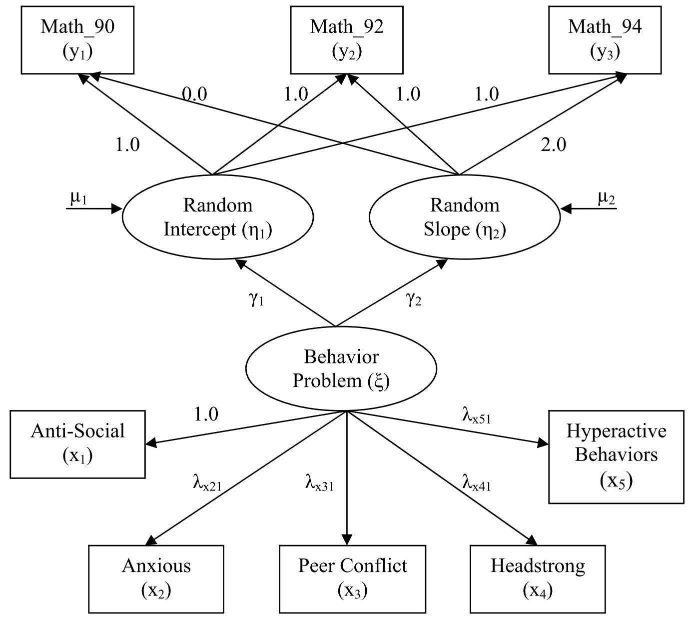

Figure 2. Path diagram of the generalized linear LCM for analyzing the NLSY79 data.

Although we only cope with the linear effect of ξ_i on $\boldsymbol{\eta}_i$ in Figure 2, Equation (5) is flexible enough to assess the effects of the observed covariates in $\mathbf{w}_i$ on $\boldsymbol{\eta}_i$, the interrelationships among the growth factors in $\boldsymbol{\eta}_i$, and the nonlinear effect of $\boldsymbol{\xi}_i$ on $\boldsymbol{\eta}_i$. For example,

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \pi_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{pmatrix} \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \xi_{i1}\xi_{i2} \end{pmatrix} + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix},$$

where π_{21} reveals the interrelationship between η_{i1} and η_{i2}, and γ_{13} and γ_{23} examines the interaction effects of ξ_{i1} and ξ_{i2} on η_{i1} and η_{i2}, respectively. The incorporation of nonlinear $\mathbf{F}(\boldsymbol{\xi}_i)$ in Equation (5) is important to provide a general model that reflects the true relationships among latent variables. The distribution of $\mathbf{y}_i$ is non-normal because of the nonlinearity of $\mathbf{F}(\boldsymbol{\xi}_i)$. Hence, the conventional computer packages developed based on the normal assumption of $\mathbf{y}_i$ cannot be directly applied to the current analysis.

As mentioned above, in Equation (5), $|\boldsymbol{\Pi}_0| = |\mathbf{I} - \boldsymbol{\Pi}|$ is independent of the elements of $\boldsymbol{\Pi}$. In some circumstances, this assumption may be violated. For example,

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} 0 & \pi_{12} \\ \pi_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix},$$

where $|\boldsymbol{\Pi}_0| = |\mathbf{I} - \boldsymbol{\Pi}| = 1 - \pi_{12}\pi_{21}$. The nonconstant $|\boldsymbol{\Pi}_0|$ causes a nonstandard posterior distribution of $p(\boldsymbol{\Pi}|\cdot)$. The MCMC method developed by Li and Wang [21] can be applied to sampling from $p(\boldsymbol{\Pi}|\cdot)$.

The generalized LCMs defined by Equations (1), (5), and (7) are not identified without imposing appropriate identification constraints. For instance, an equivalent form of (7) is $\mathbf{x}_i = \mathbf{\Lambda}_x^* \boldsymbol{\xi}_i^* + \boldsymbol{\epsilon}_{xi}$, where $\mathbf{\Lambda}_x^* = \mathbf{\Lambda}_x \mathbf{R}$, $\boldsymbol{\xi}_i^* = \mathbf{R}^{-1}\boldsymbol{\xi}_i$, and $\mathbf{R}$ is an arbitrary nonsingular matrix. One common method of solving this problem is to fix appropriate elements in $\mathbf{\Lambda}_x$ at fixed known values so that the only possible choice of $\mathbf{R}$ is the identity matrix [2, 18, 33]. In Equation (8), for example, we fix λ_{x11} at 1.0 for the purpose of identification. Similarly, appropriate elements in $\mathbf{\Lambda}_y$, $\mathbf{\Pi}$, and/or $\mathbf{\Gamma}$ may also be fixed at known values if necessary.

CFA models (1) and (7) can be synthesized into a unique model framework. Assuming that $\mathbf{u}_i = (\mathbf{y}_i^T, \mathbf{x}_i^T)^T$, $\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{pmatrix}$, $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$, and $\boldsymbol{\epsilon}_{ui} = (\boldsymbol{\epsilon}_{yi}^T, \boldsymbol{\epsilon}_{xi}^T)^T$, then (1) and (7) can be combined as

$$(9) \qquad \mathbf{u}_i = \mathbf{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_{ui}, \quad i = 1, 2, \ldots, n,$$

where $\mathbf{u}_i$ $(p \times 1)$ consists of p_1 repeated measurements in (1) and p_2 multiple indicators in (7), $\boldsymbol{\omega}_i$ $(q \times 1)$ consists of q_1 latent growth factors in $\boldsymbol{\eta}_i$ and q_2 explanatory latent variables in $\boldsymbol{\xi}_i$, where $p = p_1 + p_2$, and $q = q_1 + q_2$. With this synthesis, the generalized LCM can be simply defined by Equations (5) and (9).

2.3 Generalized latent curve mixture model

To accommodate possible heterogeneity in longitudinal data, we further introduce a latent allocation variable to the model described in Section 2.2. The model consists of two parts. The first part (the upper panel of Figure 3) defines a mixture model with latent trajectory classes, in which the probability of an individual belonging to a trajectory class is predicted by explanatory latent variables in $\boldsymbol{\zeta}$ and observed covariates in $\mathbf{b}$ through a multinomial logit model. The second part (the lower panel of Figure 3) is a within-class LCM that examines how explanatory latent variables in $\boldsymbol{\xi}$ and observed covariates in $\mathbf{w}$ affect latent growth factors in $\boldsymbol{\eta}$ within a trajectory class. Each part is presented in the following sections.

2.3.1 Mixture modeling for latent trajectory classes

Suppose that K different trajectory classes are present in a random sample of size n, where K is the number of trajectory classes and is usually determined based on a model selection procedure. A mixture model for $\mathbf{u}_i$ is defined as follows:

$$(10) \qquad p(\mathbf{u}_i) = \sum_{k=1}^{K} \pi_{ik} p_k(\mathbf{u}_i|\boldsymbol{\theta}_k), \quad i = 1, 2, \ldots, n,$$

where π_{ik} is the probability of the ith individual in the kth class such that $\sum_{k=1}^{K} \pi_{ik} = 1.0$, and $p_k(\cdot|\cdot)$ is the density function of $\mathbf{u}_i$ with parameter vector $\boldsymbol{\theta}_k$. The probability π_{ik} can be further modeled by the multinomial logit model:

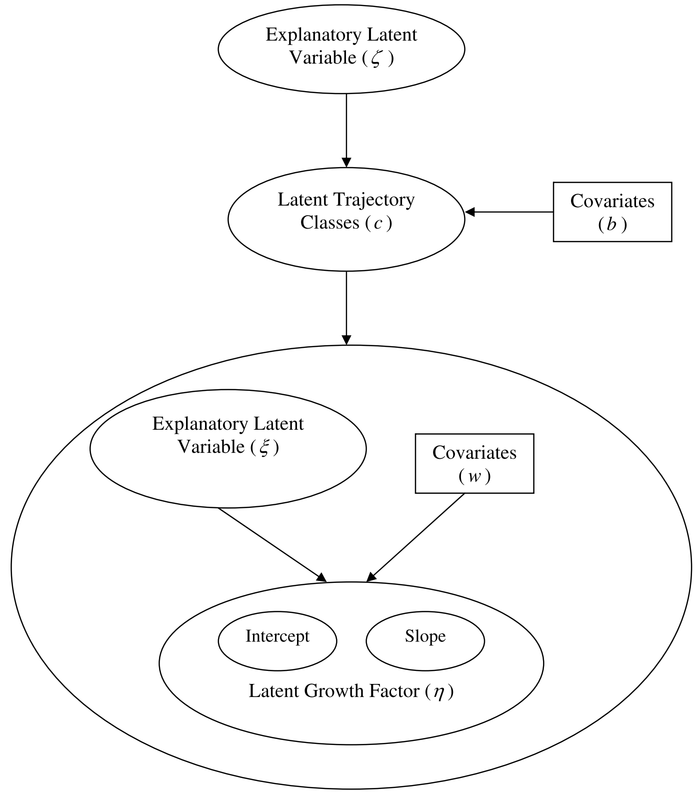

Figure 3. Structure of GLCMM.

$$(11) \qquad \pi_{ik} = p(c_i = k|\mathbf{b}_i, \boldsymbol{\zeta}_i) = \frac{\exp\left(\varphi_{k0} + \boldsymbol{\varphi}_k^{(1)^T}\mathbf{b}_i + \boldsymbol{\varphi}_k^{(2)^T}\boldsymbol{\zeta}_i\right)}{\sum_{l=1}^{K}\exp\left(\varphi_{l0} + \boldsymbol{\varphi}_l^{(1)^T}\mathbf{b}_i + \boldsymbol{\varphi}_l^{(2)^T}\boldsymbol{\zeta}_i\right)},$$

where c_i is a latent allocation variable, $\mathbf{b}_i$ is an $r_2 \times 1$ vector of fixed covariates, $\boldsymbol{\zeta}_i$ is a $q_3 \times 1$ vector of explanatory latent variables, and φ_{k0}, $\boldsymbol{\varphi}_k^{(1)} = (\varphi_{k1}, \varphi_{k2}, \ldots, \varphi_{k,r_2})^T$, and $\boldsymbol{\varphi}_k^{(2)} = (\varphi_{k,r_2+1}, \varphi_{k,r_2+2}, \ldots, \varphi_{k,r_2+q_3})^T$ are class-specific unknown parameters. To identify (11), $\varphi_{K0}, \varphi_{K1}, \ldots, \varphi_{K,r_2+q_3}$ are fixed at 0.0. The explanatory latent vector $\boldsymbol{\zeta}_i$ in (11) is measured by multiple indicators:

$$(12) \qquad \mathbf{v}_i = \mathbf{\Lambda}_v\boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_{vi}, \quad i = 1, 2, \ldots, n,$$

where $\mathbf{v}_i$ is a $p_3 \times 1$ vector of observed indicators, $\mathbf{\Lambda}_v$ is a $p_3 \times q_3$ factor loading matrix, and $\boldsymbol{\epsilon}_{vi}$ is a $p_3 \times 1$ vector of random errors with distribution $N(\mathbf{0}, \mathbf{\Psi}_v)$, where $\mathbf{\Psi}_v$ is diagonal. In the study of the NSLY79 data, $p_3 = 3$ and $q_3 = 1$. The observed indicators are household variable, home cognitive stimulation score, and emotional support score, which are used to measure the latent variable "home environment". We assume that $\boldsymbol{\zeta}_i$ is independent of $\boldsymbol{\epsilon}_{vi}$ and distributed as $N(\mathbf{0}, \mathbf{\Phi}_v)$, where $\mathbf{\Phi}_v$ is an unknown covariance matrix. Model (12) can be identified by fixing appropriate elements in $\mathbf{\Lambda}_v$ at known values.

Compared with the conventional multinomial logit model, Model (11) involves fixed covariates and latent variables as predictors. The main goal of this extension is to provide a more flexible model for better prediction of the unknown probability of class membership. In Equations (5) and (11), and (7) and (12), common variables between $\mathbf{w}_i$ and $\mathbf{b}_i$, $\mathbf{v}_i$ and $\mathbf{x}_i$, and $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ may exist, respectively.

In the study of the NLSY79 data, "home environment" and "gender" are of interest in predicting the probability of class membership. Model (11) with $p_3 = 3$ and $q_3 = r_2 = 1$ can be written as

(13)
$$\pi_{ik} = p(c_i = k|b_i, \zeta_i) = \frac{\exp(\varphi_{k0} + \varphi_{k1} b_i + \varphi_{k2}\zeta_i)}{\sum_{l=1}^{K} \exp(\varphi_{l0} + \varphi_{l1} b_i + \varphi_{l2}\zeta_i)},$$

where b_i indicates the observed covariate "gender", and ζ_i indicates the latent variable "home environment", which is measured by the indicators of household variable, home cognitive stimulation score, and emotional support score via Equation (12).

2.3.2 Within-class generalized latent curve model

In Section 2.3.1, we defined a mixture model, in which $\mathbf{u}_i$ is assumed to come from K latent trajectory classes. In this section, we further define the within-class model. Given that $c_i = k,\ k = 1, \ldots, K$, the within-class generalized LCM can be defined as:

$$\text{(14)} \qquad \mathbf{u}_i = \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i + \boldsymbol{\epsilon}_{ui}, \quad i = 1, 2, \ldots, n,$$
$$\text{(15)} \qquad \boldsymbol{\eta}_i = \boldsymbol{\mu}_k + \mathbf{A}_k \mathbf{w}_i + \boldsymbol{\Pi}_k \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k \mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i,$$

where all the components are defined in a similar manner as those in Equations (5) and (9) except that the unknown parameters are class-specific. The models defined in Equations (14) and (15) allow for heterogeneity both in the patterns of trajectories and in the effects of explanatory latent and observed variables on the latent growth factors. These models are characterized by a form of invariance [42], that is, the fixed parameters are fixed at the same locations in different trajectory classes. For instance, in Equation (8), we fix the element λ_{x11} at 1.0 to identify the model in each trajectory class.

2.3.3 Generalization from normal distribution to EFD

Model (14) is defined under the assumption that the conditional distributions of observed variables in $\mathbf{u}_i$, given latent variables, are normal. To handle non-normal data, we generalized the conditional distribution of $\mathbf{u}_i$ from normal to EFD. This family includes discrete distributions such as binomial and Poisson; it also includes continuous distributions such as normal and gamma.

We assume that the distribution of u_{ij}, $j = 1, \ldots, p$ conditional on the kth class and $\boldsymbol{\omega}_i$, is independent and comes from the EFD with canonical parameter ϑ_{kij} [1, 31, 33]:

$$\text{(16)} \quad p_{kj}(u_{ij}|c_i = k, \boldsymbol{\omega}_i) = \exp\{[u_{ij}\vartheta_{kij} - b_j(\vartheta_{kij})]/\psi_{\epsilon kj} + c_j(u_{ij}, \psi_{\epsilon kj})\},$$

where $b_j(\cdot)$ and $c_j(\cdot)$ are differentiable functions with the dots denoting the derivatives, the forms of which depend on the distributions of the response variables, $E(u_{ij}|c_i = k, \boldsymbol{\omega}_i) = \dot{b}_j(\vartheta_{kij})$, and $Var(u_{ij}|c_i = k, \boldsymbol{\omega}_i) = \psi_{\epsilon kj}\ddot{b}_j(\vartheta_{kij})$. Let $\boldsymbol{\vartheta}_{ki} = (\vartheta_{ki1}, \ldots, \vartheta_{kip})^T$, where $\vartheta_{kij} = g_j(E(u_{ij}|c_i = k, \boldsymbol{\omega}_i))$, and $g_j(\cdot)$ are link functions. Under the EFD framework, The expectation of Equation (14) can be rewritten as

$$\text{(17)} \qquad \boldsymbol{\vartheta}_{ki} = \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i.$$

Similarly, the distribution of $\mathbf{v}_i$ in Equation (12) can also be generalized from normal to EFD. Let $\boldsymbol{\vartheta}_{vi} = (\vartheta_{vi1}, \ldots, \vartheta_{vip_3})^T$, where $\vartheta_{vij} = h_j(E(v_{ij}|\boldsymbol{\zeta}_i))$, and $h_j(\cdot)$ are link functions. Under the EFD framework, the expectation of Equation (12) can be rewritten as

$$\text{(18)} \qquad \boldsymbol{\vartheta}_{vi} = \boldsymbol{\Lambda}_v \boldsymbol{\zeta}_i.$$

For instance, in the study of the NLSY79 data, the indicators $\{v_{i1}, v_{i2}, v_{i3}\}$ for measuring the latent variable "home environment, ζ_i" are discrete and continuous, in which v_{i1} is a binary variable coded with $\{0, 1\}$ representing "in household with father and other relatives" and "in household with mother," respectively, whereas v_{i2}, home cognitive stimulation score, and v_{i3}, emotional support score, are continuous and assumed to be normally distributed. Under the EFD framework, the corresponding CFA model can be written as

$$\text{(19)} \qquad \begin{pmatrix} \vartheta_{v1} \\ \vartheta_{v2} \\ \vartheta_{v3} \end{pmatrix} = \begin{pmatrix} 1.0 \\ \lambda_{v,21} \\ \lambda_{v,31} \end{pmatrix} \zeta_i,$$

where $\vartheta_{vj} = h_j(E(v_{ij}|\zeta_i))$, in which the link function h_1 is a logit link, and h_2 and h_3 are the identity links. The value of 1.0 is fixed to identify the model and to introduce a scale to the latent variable ζ_i.

In summary, the GLCMM proposed in this section integrates the following features: (i) continuous latent variables, including latent growth factors and explanatory latent variables that may influence both latent growth curve and latent class membership; (ii) a latent allocation variable that accounts for latent trajectory classes; (iii) fixed and latent effects in LCM and in the multinomial logit model; and (iv) discrete and continuous data from EFD. Hence, the proposed model is rather general and useful for revealing possible latent classes and class-specific change patterns in the analysis of heterogeneous longitudinal data.

3. BAYESIAN ANALYSIS OF GLCMM

3.1 Bayesian estimation with MCMC methods

3.1.1 Prior distributions

In a full Bayesian analysis, the unknown parameters are treated as random. An important issue is to specify prior distributions for unknown parameters. Diebolt and Roberts [9] and Roeder and Wasserman [30] pointed out that using fully non-informative prior distributions in a mixture model may lead to improper posterior distributions. Hence, following the common practice in Bayesian mixture modeling [see, for example, 18, 30, 33], we utilize the conjugate type prior distributions in this analysis.

Let $\mathbf{\Lambda}_{km}^T$ denote the m-th row of $\mathbf{\Lambda}_k$, $\psi_{\epsilon km}$ denote the m-th diagonal element of $\mathbf{\Psi}_{\epsilon k}$, for $m = 1, 2, \ldots, p$, and $\mathbf{\Lambda}_{\delta k} = (\mathbf{A}_k, \mathbf{\Pi}_k, \mathbf{\Gamma}_k)$; let $\mathbf{\Lambda}_{\delta kl}^T$ denote the l-th row of $\mathbf{\Lambda}_{\delta k}$, $\psi_{\delta kl}$ denote the l-th diagonal element of $\mathbf{\Psi}_{\delta k}$, for $l = 1, 2, \ldots, q_1$; $\mathbf{\Lambda}_{vl}^T$ denote the l-th row of $\mathbf{\Lambda}_v$, and ψ_{vl} denote the l-th diagonal element of $\mathbf{\Psi}_v$, for $l = 1, 2, \ldots, p_3$. The following conjugate prior distributions are specified:

$$
\begin{aligned}
&p(\mathbf{\Lambda}_{km}) \sim N(\mathbf{\Lambda}_{0\epsilon km}, \psi_{\epsilon km}\mathbf{H}_{0\epsilon km}),\\
&p(\psi_{\epsilon km}^{-1}) \sim Gamma(\alpha_{0\epsilon km}, \beta_{0\epsilon km}),\\
&p(\mathbf{\Lambda}_{\delta kl}) \sim N(\mathbf{\Lambda}_{0\delta kl}, \psi_{\delta kl}\mathbf{H}_{0\delta kl}),\\
&p(\psi_{\delta kl}^{-1}) \sim Gamma(\alpha_{0\delta kl}, \beta_{0\delta kl}),\\
&p(\mathbf{\Lambda}_{vl}) \sim N(\mathbf{\Lambda}_{0vl}, \psi_{vl}\mathbf{H}_{0vl}),\\
&p(\psi_{vl}^{-1}) \sim Gamma(\alpha_{0vl}, \beta_{0vl}),\\
&p(\boldsymbol{\mu}_k) \sim N(\boldsymbol{\mu}_{0k}, \mathbf{\Sigma}_{0k}),\\
&p(\mathbf{\Phi}_k^{-1}) \sim Wishart(\mathbf{R}_{0k}, \rho_0),\\
&p(\boldsymbol{\varphi}_k) \sim N(\boldsymbol{\varphi}_{0k}, \mathbf{\Sigma}_{\varphi 0k}),\\
&p(\mathbf{\Phi}_v^{-1}) \sim Wishart(\mathbf{R}_{0v}, \rho_{0v}),
\end{aligned} \tag{20}
$$

where $\mathbf{\Lambda}_{0\epsilon km}$, $\mathbf{\Lambda}_{0\delta kl}$, $\mathbf{\Lambda}_{0vl}$, $\boldsymbol{\mu}_{0k}$, $\boldsymbol{\varphi}_{0k}$, $\alpha_{0\epsilon km}$, $\beta_{0\epsilon km}$, $\alpha_{0\delta kl}$, $\beta_{0\delta kl}$, α_{0vl}, β_{0vl}, ρ_0 and ρ_{0v}, and the positive-definite matrices $\mathbf{H}_{0\epsilon km}$, $\mathbf{H}_{0\delta kl}$, $\mathbf{H}_{0vl}$, $\mathbf{\Sigma}_{0k}$, $\mathbf{\Sigma}_{\varphi 0k}$, $\mathbf{R}_{0k}$, and $\mathbf{R}_{0v}$ are hyperparameters. In practice, values of hyperparameters are adopted from previous studies or other sources. Given that assigning conjugate prior distributions for unknown parameters usually leads to the same forms of posterior distributions [11, 13, 18], the resulting posterior distributions for most of the unknown parameters in our proposed model are standard, and simulating observations from them is efficient.

3.1.2 Posterior inferences with MCMC sampling

Let $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ denote the observed data matrices, $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_n)$ and $\widetilde{\mathbf{\Omega}} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_n)$ denote the matrices of continuous latent variables, and $\mathbf{C} = (c_1, c_2, \ldots, c_n)^T$ denote the vector of latent allocation variables. Moreover, let $\boldsymbol{\theta}_k$ denote the parameter vector that contains all the unknown parameters in $\mathbf{\Lambda}_k$, $\boldsymbol{\mu}_k$, $\mathbf{A}_k$, $\mathbf{\Pi}_k$, $\mathbf{\Gamma}_k$, $\mathbf{\Phi}_k$, $\mathbf{\Psi}_{\epsilon k}$, $\mathbf{\Psi}_{\delta k}$, $\mathbf{\Lambda}_v$, $\mathbf{\Phi}_v$, $\mathbf{\Psi}_v$, and $\boldsymbol{\varphi}_k$, where $\boldsymbol{\varphi}_k = (\varphi_{k0}, \varphi_{k1}, \ldots, \varphi_{k,r_2+q_3})^T$, for $k = 1, 2, \ldots, K$. Denote $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K\}$. Considering that the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{U}$ and $\mathbf{V}$, $p(\boldsymbol{\theta}|\mathbf{U}, \mathbf{V})$, is intractable, we performed data augmentation [40] by augmenting the observed data $\mathbf{F}_o = \{\mathbf{U}, \mathbf{V}\}$ with the latent quantities $\mathbf{F}_l = \{\mathbf{\Omega}, \widetilde{\mathbf{\Omega}}, \mathbf{C}\}$ in the posterior analysis. The Bayesian estimate of $\boldsymbol{\theta}$ is obtained based on the observations drawn from $p(\boldsymbol{\theta}, \mathbf{F}_l|\mathbf{F}_o)$ using MCMC tools such as the Gibbs sampler and the MH algorithm. The Bayesian analysis of the proposed model is complex. In particular, the following issues should be addressed.

(1) Posterior distributions and sampling Given the complexity of the model and data structure, some full conditional distributions involved in the MCMC algorithm are nontrivial. For instance, the EFD framework, the nonlinear function $\mathbf{F}(\cdot)$, and the multinomial logit model lead to the posterior distributions that one cannot directly sample from (see details in the Appendix). Simulating observations from these distributions is a challenging task in implementing the MCMC algorithm. We employed the MH algorithm together with other MCMC techniques to achieve this purpose. The full conditional distributions of other parameters in $\boldsymbol{\theta}$ are the normal, Gamma, and inverted Wishart distributions. Simulating observations from these distributions is straightforward.

(2) Label switching and permutation sampler A second issue that needs to be addressed is the label switching problem in Bayesian mixture modeling. For the proposed GLCMM with K classes, the likelihood function is invariant with a permutation of the class labels $1, 2, \ldots, K$. So, the unconstrained posterior is also invariant to relabeling the states under a symmetric prior. This condition induces a multimodal posterior, and may give rise to misleading results in Bayesian estimation. Following Lee [18], the permutation sampler proposed by Frühwirth-Schnatter [14] is used to solve the label switching problem. The details of permutation sampler can be found in [14] and [18].

To conduct the posterior inference, a sufficiently large number of random observations were simulated from the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{F}_l|\mathbf{F}_o)$ using the Gibbs sampler [15] coupled with the MH algorithm [17, 25]. More specifically, $\boldsymbol{\theta}$ and $\mathbf{F}_l$ were simulated from their corresponding full conditional distributions iteratively. Under mild regularity conditions, the joint distribution of $(\boldsymbol{\theta}^{(j)}, \mathbf{F}_l^{(j)})$ converges at an exponential rate to the desired posterior distribution $p(\boldsymbol{\theta}, \mathbf{F}_l|\mathbf{F}_o)$, after a sufficiently large number of burn-in iterations [15]. The required number of burn-in iterations for achieving convergence can be determined by plotting the simulated sequences of individual parameters. At convergence, parallel sequences generated with different starting values should mix well. Another method for monitoring convergence is based on the estimated potential scale reduction (EPSR) values [16], in which convergence is achieved when the EPSR values are all less than 1.2. Based on a simulated sample of observations after convergence, the Bayesian estimates of $\boldsymbol{\theta}$ can be obtained through the sample mean and the standard error estimates can be obtained through the sample covariance matrix.

3.2 Bayesian model selection

For the proposed GLCMM, the number of latent classes, K, is usually unknown. A model selection procedure is needed to determine the most appropriate K. In this article, we adopted a modified DIC [38] for the model selection. DIC combines Bayesian measures of fit and model complexity. The model with the smallest DIC value is selected. However, DIC cannot be directly applied to the selection of mixture models [18, 39]. Recently, Celeux et al. [6] have proposed different DIC constructions for mixture models. In this paper, the most current modification of DIC for mixture model is used. This modified DIC is defined as follows:

$$DIC = -4E_{\boldsymbol{\theta},\mathbf{F}_l}\{\log p(\mathbf{F}_o,\mathbf{F}_l|\boldsymbol{\theta})|\mathbf{F}_o\} + 2E_{\mathbf{F}_l}\left\{\log p(\mathbf{F}_o,\mathbf{F}_l|E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{F}_o,\mathbf{F}_l])|\mathbf{F}_o\right\}, \tag{21}$$

where $\log p(\mathbf{F}_o,\mathbf{F}_l|\boldsymbol{\theta})$ is the complete data log-likelihood function. The modified DIC can be approximated by

$$DIC \approx -\frac{4}{T}\sum_{t=1}^{T}\log p(\mathbf{F}_o,\mathbf{F}_l^{(t)}|\boldsymbol{\theta}^{(t)}) + \frac{2}{T}\sum_{t=1}^{T}\log p(\mathbf{F}_o,\mathbf{F}_l^{(t)}|\bar{\boldsymbol{\theta}}^{(t)}), \tag{22}$$

where $\{(\boldsymbol{\theta}^{(t)},\mathbf{F}_l^{(t)}); t=1,\ldots,T\}$ are the simulated samples from $p(\boldsymbol{\theta},\mathbf{F}_l|\cdot)$, and $\bar{\boldsymbol{\theta}}^{(t)}$ is the sample mean calculated based on the simulated samples from $p(\boldsymbol{\theta}|\mathbf{F}_o,\mathbf{F}_l^{(t)})$.

4. SIMULATION STUDY

The objective of this simulation study is to examine the empirical performance of the Bayesian estimation and model selection for the proposed GLCMM. The data set was generated from a GLCMM with two trajectory classes ($K=2$). For $k=1,2$, and $i=1,2,\ldots,n$, the probability of the ith individual falling in the kth class was modeled through the following multinomial logit model with $r_2=1$, $q_3=2$:

$$\pi_{ik} = \frac{\exp(\varphi_{k0}+\varphi_{k1}b_i+\varphi_{k2}\zeta_{i1}+\varphi_{k3}\zeta_{i2})}{\sum_{l=1}^{2}\exp(\varphi_{l0}+\varphi_{l1}b_i+\varphi_{l2}\zeta_{i1}+\varphi_{l3}\zeta_{i2})}, \tag{23}$$

where $(\varphi_{10},\varphi_{11},\varphi_{12},\varphi_{13})^T = (-0.5,1.0,0.8,0.8)^T$, and $(\varphi_{20},\varphi_{21},\varphi_{22},\varphi_{23})^T$ is fixed to $(0.0,0.0,0.0,0.0)^T$ for identifying the model; the fixed covariate b_i was generated from $N(1.0,1.0)$; and the explanatory latent variables (ζ_{i1},ζ_{i2}) were measured by six observed indicators via the following CFA model:

$$\begin{pmatrix} v_{i1}\\ v_{i2}\\ v_{i3}\\ v_{i4}\\ v_{i5}\\ v_{i6}\end{pmatrix} = \begin{pmatrix} 1.0 & 0.0\\ \lambda_{v,21} & 0.0\\ \lambda_{v,31} & 0.0\\ 0.0 & 1.0\\ 0.0 & \lambda_{v,52}\\ 0.0 & \lambda_{v,62}\end{pmatrix}\begin{pmatrix}\zeta_{i1}\\ \zeta_{i2}\end{pmatrix} + \begin{pmatrix}\epsilon_{vi1}\\ \epsilon_{vi2}\\ \epsilon_{vi3}\\ \epsilon_{vi4}\\ \epsilon_{vi5}\\ \epsilon_{vi6}\end{pmatrix},$$

where 1's and 0's are fixed to identify the model, $(\zeta_{i1},\zeta_{i2})^T \sim N(\mathbf{0},\mathbf{\Phi}_v)$, and $\epsilon_{vij}\sim N(0,\psi_{vj})$. The true population values of unknown parameters are $\lambda_{v,21}=\lambda_{v,31}=\lambda_{v,52}=\lambda_{v,62}=0.8$, $\phi_{v,11}=\phi_{v,22}=1.0$, and $\phi_{v,21}=0.3$, and $\psi_{v1}=\cdots=\psi_{v6}=0.36$.

To define the within-class generalized LCM, we let $\mathbf{u}_i=(u_{i1},u_{i2},\ldots,u_{i,14})^T=(y_{i1},\ldots,y_{i8},x_{i1},\ldots,x_{i6})^T$. For $j=1,\ldots,4$, y_{ij} were generated from a binomial distribution $B(1,p_{kij})$, with $p_{kij}=\exp(\vartheta_{kij})/[1+\exp(\vartheta_{kij})]$, and ϑ_{kij} satisfy the LCM

$$\begin{pmatrix}\vartheta_{ki1}\\ \vartheta_{ki1}\\ \vartheta_{ki3}\\ \vartheta_{ki4}\end{pmatrix} = \begin{pmatrix}1.0 & 0.0\\ 1.0 & 1.0\\ 1.0 & 2.0\\ 1.0 & 3.0\end{pmatrix}\begin{pmatrix}\eta_{i1}\\ \eta_{i2}\end{pmatrix}.$$

For $j=5,\ldots,8$, y_{ij} are continuous, and satisfy the LCM

$$\begin{pmatrix}y_{i5}\\ y_{i6}\\ y_{i7}\\ y_{i8}\end{pmatrix} = \begin{pmatrix}1.0 & 0.0\\ 1.0 & 1.0\\ 1.0 & 2.0\\ 1.0 & 3.0\end{pmatrix}\begin{pmatrix}\eta_{i3}\\ \eta_{i4}\end{pmatrix} + \begin{pmatrix}\epsilon_{yi5}\\ \epsilon_{yi6}\\ \epsilon_{vi7}\\ \epsilon_{vi8}\end{pmatrix},$$

where $\epsilon_{yij}\sim N(0,\psi_{ykj})$.

For $j=1,\ldots,6$, x_{ij} relate to the explanatory latent variables (ξ_{i1},ξ_{i2}) via the CFA model

$$\begin{pmatrix} x_{i1}\\ x_{i2}\\ x_{i3}\\ x_{i4}\\ x_{i5}\\ x_{i6}\end{pmatrix} = \begin{pmatrix} 1.0 & 0.0\\ \lambda_{xk,21} & 0.0\\ \lambda_{xk,31} & 0.0\\ 0.0 & 1.0\\ 0.0 & \lambda_{xk,52}\\ 0.0 & \lambda_{xk,62}\end{pmatrix}\begin{pmatrix}\xi_{i1}\\ \xi_{i2}\end{pmatrix} + \begin{pmatrix}\epsilon_{xi1}\\ \epsilon_{xi2}\\ \epsilon_{xi3}\\ \epsilon_{xi4}\\ \epsilon_{xi5}\\ \epsilon_{xi6}\end{pmatrix},$$

where 1's and 0's are fixed for the purpose of identification, $(\xi_{i1},\xi_{i2})^T\sim N(\mathbf{0},\mathbf{\Phi}_k)$, and $\epsilon_{xij}\sim N(0,\psi_{xkj})$.

The effects of explanatory latent variables and covariates on the latent growth factors, $\eta_{ij}, j=1,\ldots,4$ were assessed through the structural equations with $r_1=1$, $q_2=2$, and $\mathbf{F}(\boldsymbol{\xi}_i)=(\xi_{i1},\xi_{i2},\xi_{i1}\xi_{i2})^T$ as follows:

$$\begin{aligned}
\eta_{i1} &= \mu_{k1}+a_{k,11}w_i+\gamma_{k,11}\xi_{i1}+\gamma_{k,12}\xi_{i2}+\gamma_{k,13}\xi_{i1}\xi_{i2}+\delta_{i1},\\
\eta_{i2} &= \mu_{k2}+a_{k,21}w_i+\gamma_{k,21}\xi_{i1}+\gamma_{k,22}\xi_{i2}+\gamma_{k,23}\xi_{i1}\xi_{i2}+\delta_{i2},\\
\eta_{i3} &= \mu_{k3}+a_{k,31}w_i+\gamma_{k,31}\xi_{i1}+\gamma_{k,32}\xi_{i2}+\gamma_{k,33}\xi_{i1}\xi_{i2}+\delta_{i3},\\
\eta_{i4} &= \mu_{k4}+a_{k,41}w_i+\gamma_{k,41}\xi_{i1}+\gamma_{k,42}\xi_{i2}+\gamma_{k,43}\xi_{i1}\xi_{i2}+\delta_{i4},
\end{aligned}$$

where the μ's, a's, and γ's were unknown parameters. The fixed covariate w_i $(i=1,2,\ldots,n)$ was generated from $N(0.0,1.0)$, and $\delta_{ij}\sim N(0,\psi_{\delta kj})$.

Figure 4 depicts the path diagram of the within-class generalized LCM defined above.

In Figure 4, two linear LCMs were considered simultaneously. One was defined with latent growth factors $\{\eta_{i1},\eta_{i2}\}$ and binary repeated measures $y_{i1},\ldots,y_{i4}$ under the EFD framework (see Equation (17)). The other was defined with latent growth factors $\{\eta_{i3},\eta_{i4}\}$ and continuous repeated measures $y_{i5},\ldots,y_{i8}$.

The true population values of the unknown parameters involved in the within-class LCMs are given as follows:

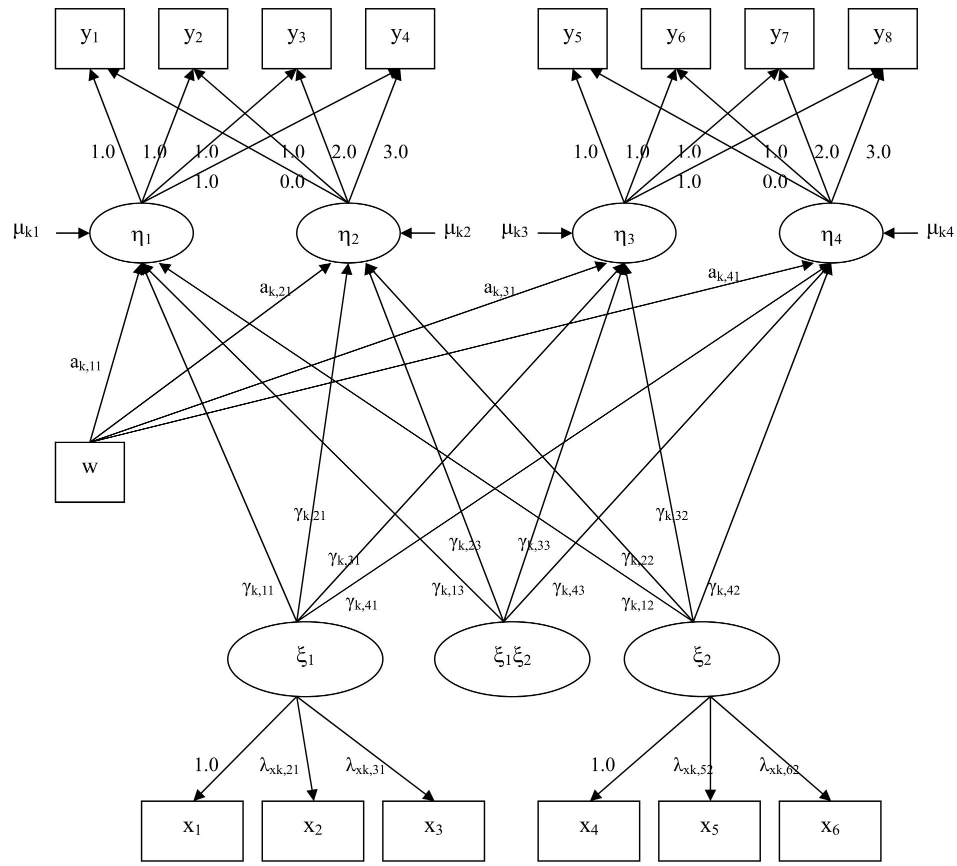

Figure 4. Path diagram of within-class generalized LCM for simulation study.

At the first class ($k = 1$), $\lambda_{x1,21} = \lambda_{x1,31} = \lambda_{x1,52} = \lambda_{x1,62} = 0.8$; $(\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14})^T = (2.0, 1.0, -2.0, -1.5)^T$; $(a_{1,11}, a_{1,21}, a_{1,31}, a_{1,41})^T = (0.5, -0.5, 0.5, -0.5)^T$; $\gamma_{1,jm} = 0.5$, for $j = 1, \ldots, 4$, $m = 1, \ldots, 3$; $\psi_{y1j} = 0.36$, for $j = 5, \ldots, 8$; $\psi_{x1j} = 0.36$, for $j = 1, \ldots, 6$; $\psi_{\delta 1l} = 0.36$, for $l = 1, \ldots, 4$; $\phi_{1,11} = \phi_{1,22} = 1.0$, and $\phi_{1,21} = 0.3$.

At the second class ($k = 2$), $\lambda_{x2,21} = \lambda_{x2,31} = \lambda_{x2,52} = \lambda_{x2,62} = 0.4$; $(\mu_{21}, \mu_{22}, \mu_{23}, \mu_{24})^T = (1.0, 1.5, 1.5, 0.5)^T$; $(a_{2,11}, a_{2,21}, a_{1,31}, a_{2,41})^T = (-0.5, 0.5, -0.5, 0.5)^T$; $\gamma_{2,jm} = -0.5$, for $j = 1, \ldots, 4$, $m = 1, \ldots, 3$; $\psi_{y2j} = 0.25$, for $j = 5, \ldots, 8$; $\psi_{x2j} = 0.25$, for $j = 1, \ldots, 6$; $\psi_{\delta 2l} = 0.25$, for $l = 1, \ldots, 4$; $\phi_{2,11} = \phi_{2,22} = 1.0$, and $\phi_{2,21} = 0.3$.

The total number of unknown parameters in the proposed model was 99. Given the number of parameters and the existence of binary data, a relatively large sample size was required to achieve accurate estimation result [18]. In this simulation study, sample sizes of $n = 800$ and $n = 1,600$ were used, and 100 replications were conducted for each sample size.

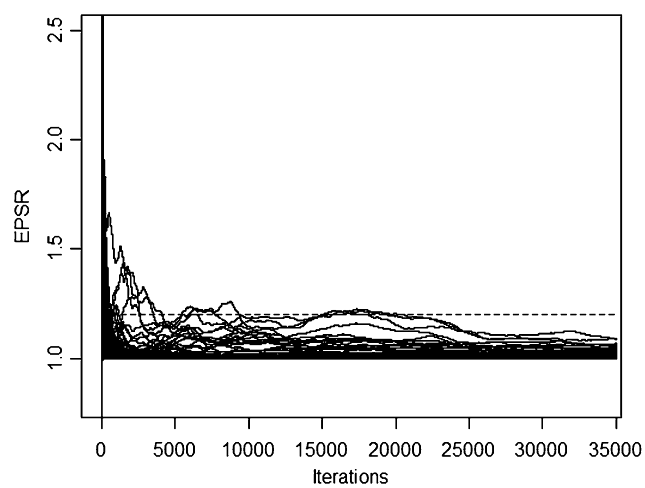

Figure 5. Plot of the estimated potential scale reduction values against the iteration numbers in simulation study.

To provide a sensitivity analysis of the Bayesian results to the prior specification, the prior input in (20) was perturbed as follows:

Type I: All elements in $\{\boldsymbol{\mu}_{0k}, \boldsymbol{\Lambda}_{0\epsilon km}, \boldsymbol{\Lambda}_{0\delta kl}, \boldsymbol{\varphi}_{0k}, \boldsymbol{\Lambda}_{0vl}\}$ were taken to be 1.0, $\alpha_{0\epsilon km} = \alpha_{0\delta kl} = 9$, $\beta_{0\epsilon km} = \beta_{0\delta kl} = 4$, $\boldsymbol{\Sigma}_{0k}$, $\mathbf{H}_{0\epsilon km}$, $\mathbf{H}_{0\delta kl}$, $\mathbf{H}_{0vl}$ and $\boldsymbol{\Sigma}_{\varphi 0k}$ were identity matrices, $\rho_0 = \rho_{0v} = 5$, $\mathbf{R}_{0k}^{-1}$ and $\mathbf{R}_{0v}^{-1}$ were taken as diagonal matrix with diagonal elements 2.0.

Type II: All elements in $\{\boldsymbol{\mu}_{0k}, \boldsymbol{\Lambda}_{0\epsilon km}, \boldsymbol{\Lambda}_{0\delta kl}, \boldsymbol{\varphi}_{0k}, \boldsymbol{\Lambda}_{0vl}\}$ were taken to be 0.0, $\alpha_{0\epsilon km} = \alpha_{0\delta kl} = 10$, $\beta_{0\epsilon km} = \beta_{0\delta kl} = 3$, $\boldsymbol{\Sigma}_{0k}$, $\mathbf{H}_{0\epsilon km}$, $\mathbf{H}_{0\delta kl}$, $\mathbf{H}_{0vl}$ and $\boldsymbol{\Sigma}_{\varphi 0k}$ were diagonal matrices

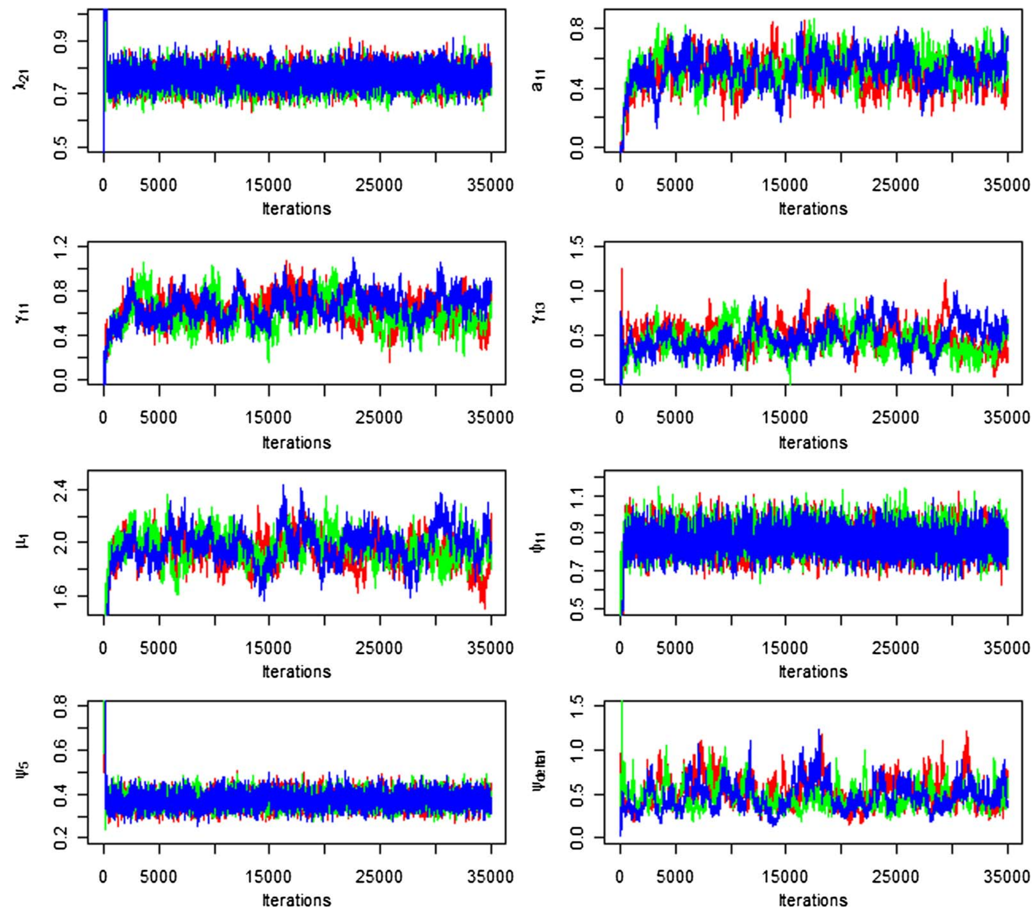

Figure 6. Three chains of observations corresponding to λ_{21}, a_{11}, γ_{11}, γ_{13}, μ_1, ϕ_{11}, ψ_5 and $\psi_{\delta 1}$ of Class 1 in simulation study generated by different initial values.

with diagonal elements 5.0, $\rho_0 = \rho_{0v} = 8$, $\mathbf{R}_{0k}^{-1}$ and $\mathbf{R}_{0v}^{-1}$ were taken as diagonal matrices with diagonal elements 8.0.

Initially, we ran the random permutation sampler [14] and found that $\mu_{13} < \mu_{23}$ was an appropriate identifiability constraint for the MCMC algorithm. A few test runs were conducted to provide information about the number of MCMC iterations required for achieving convergence. For $n = 800$, Figure 5 shows that the MCMC algorithm converged in less than 25,000 iterations, in which the EPSR values were stable below 1.2. Moreover, the plots of sequences of observations corresponding to the parameters randomly selected from Class 1 are displayed in Figure 6. Those corresponding to other parameters are very similar and are not presented here. These trace plots show that MCMC chains with different starting values mix very well. Hence, we collected 10,000 observations after discarding 25,000 burn-in iterations to obtain the Bayesian results at each replication. Based on 100 replications, the bias of the estimates (Bias) and the root mean squares (RMS) between the estimates and the true values were computed. Tables 1 to 4 report the results. The findings are summarized as follows: (1) Most of the Bias and the RMS values are reasonably small. (2) The estimates of the parameters associated with the binary observed variables are not as accurate as those associated with the continuous observed variables (i.e., their RMS values are relatively larger) because binary variables usually provide less information than continuous variables. (3) Under the given sample sizes, the Bayesian estimates obtained under the two different prior inputs have no substantial difference. (4) As expected, a larger sample size produces better estimates with smaller RMS values.

To access the performance of the modified DIC in determining the number of classes, K, we denote M_k as the GLCMM with k latent classes, $k = 1, 2, 3$. Here, M_2 is the true model defined above. For M_1 and M_3, the model setting and prior specification in each class are similar to those given in M_2. The modified DIC values for M_1 to M_3 were calculated from the simulated observations in the MCMC iterations. Table 5 presents a summary of the modified DIC values under different model settings and prior inputs. With $n = 800$, the true model M_2 was selected with the smallest DIC value in 98 of 100 replications under Prior I and 96 of 100 replications under Prior II. With $n = 1,600$, the true model M_2 is selected with the smallest DIC value under each

Table 1. Performance of the Bayesian estimates in the simulation study with Type I prior input under sample size $n = 800$

	Class 1			Class 2				Class 1			Class 2		
Par	True	Bias	RMS	True	Bias	RMS	Par	True	Bias	RMS	True	Bias	RMS
λ_{21}	0.8	0.012	0.044	0.4	0.024	0.044	ψ_{y5}	0.36	0.008	0.039	0.25	0.023	0.035
λ_{31}	0.8	0.019	0.049	0.4	0.024	0.044	ψ_{y6}	0.36	0.007	0.032	0.25	0.015	0.028
λ_{52}	0.8	0.013	0.049	0.4	0.015	0.048	ψ_{y7}	0.36	0.001	0.034	0.25	0.009	0.028
λ_{62}	0.8	0.012	0.045	0.4	0.020	0.045	ψ_{y8}	0.36	0.025	0.061	0.25	0.046	0.058
μ_1	2.0	−0.005	0.145	1.0	0.068	0.164	ψ_{x1}	0.36	0.059	0.093	0.25	0.063	0.078
μ_2	1.0	0.033	0.175	1.5	0.233	0.305	ψ_{x2}	0.36	0.007	0.030	0.25	0.008	0.022
μ_3	−2.0	−0.027	0.059	1.5	0.018	0.050	ψ_{x3}	0.36	0.009	0.032	0.25	0.010	0.025
μ_4	−1.5	−0.043	0.062	0.5	0.015	0.045	ψ_{x4}	0.36	0.055	0.086	0.25	0.049	0.067
a_{11}	0.5	−0.018	0.136	−0.5	−0.041	0.140	ψ_{x5}	0.36	0.012	0.031	0.25	0.012	0.023
a_{21}	−0.5	−0.014	0.107	0.5	0.059	0.142	ψ_{x6}	0.36	0.006	0.035	0.25	0.008	0.021
a_{31}	0.5	−0.004	0.043	−0.5	0.001	0.048	$\psi_{\delta 1}$	0.36	0.154	0.207	0.25	0.170	0.178
a_{41}	−0.5	0.002	0.031	0.5	−0.004	0.035	$\psi_{\delta 2}$	0.36	0.129	0.163	0.25	0.145	0.160
γ_{11}	0.5	0.007	0.161	−0.5	−0.042	0.192	$\psi_{\delta 3}$	0.36	0.015	0.050	0.25	0.019	0.039
γ_{12}	0.5	−0.002	0.193	−0.5	−0.066	0.182	$\psi_{\delta 4}$	0.36	0.026	0.048	0.25	0.009	0.028
γ_{13}	0.5	0.147	0.257	−0.5	−0.062	0.200	ϕ_{11}	1.0	−0.038	0.113	1.0	−0.084	0.129
γ_{21}	0.5	0.013	0.182	−0.5	−0.092	0.223	ϕ_{21}	0.3	−0.009	0.051	0.3	−0.039	0.073
γ_{22}	0.5	0.011	0.154	−0.5	−0.099	0.208	ϕ_{22}	1.0	−0.019	0.094	1.0	−0.076	0.134
γ_{23}	0.5	0.101	0.233	−0.5	-0.057	0.192	φ_0	−0.5	0.016	0.120	0.0*		
γ_{31}	0.5	0.015	0.063	−0.5	−0.011	0.059	φ_1	1.0	−0.007	0.102	0.0*		
γ_{32}	0.5	−0.007	0.058	−0.5	0.008	0.055	φ_2	0.8	0.032	0.124	0.0*		
γ_{33}	0.5	0.047	0.083	−0.5	−0.057	0.120	φ_3	0.8	0.001	0.115	0.0*		
γ_{41}	0.5	0.004	0.060	−0.5	0.000	0.050							
γ_{42}	0.5	0.001	0.065	−0.5	−0.004	0.050							
γ_{43}	0.5	0.068	0.095	−0.5	−0.047	0.099							

Par	True	Bias	RMS	Par	True	Bias	RMS
$\lambda_{v,21}$	0.8	0.005	0.034	ψ_{v1}	0.36	0.009	0.037
$\lambda_{v,31}$	0.8	0.004	0.036	ψ_{v2}	0.36	0.002	0.025
$\lambda_{v,52}$	0.8	0.012	0.035	ψ_{v3}	0.36	0.001	0.025
$\lambda_{v,62}$	0.8	0.003	0.035	ψ_{v4}	0.36	0.011	0.033
$\phi_{v,11}$	1.0	0.000	0.078	ψ_{v5}	0.36	−0.001	0.020
$\phi_{v,21}$	0.3	−0.003	0.041	ψ_{v6}	0.36	0.003	0.024
$\phi_{v,22}$	1.0	−0.001	0.075				

Note: The subscript k is suppressed for notational simplicity, and the 0's with asterisk are fixed for identification purpose.

of the 100 replications and each of the prior inputs. The above findings demonstrates that DIC is a valid Bayesian model selection criterion.

5. AN ILLUSTRATIVE EXAMPLE

In this section, we applied the GLCMM to a data set extracted from the NLSY79 data to illustrate our methodology. A subset of the NLSY79 data collected in 1990, 1992, and 1994 was used in this example. After excluding missing data, the sample size was $n = 1,674$. We consider the continuous longitudinal variables y_1, y_2, y_3, which represent the PIAT scores [10] in mathematics in 1990, 1992, and 1994, respectively. PIAT is one of the most commonly used age-appropriate assessments of academic achievement, and it is used as an indication of a child's cognitive development. This illustrative example has the following main objectives: (1) investigate whether the trajectories of PIAT scores in mathematics contain heterogeneous patterns, (2) examine how "gender" and "home environment" influence the latent class membership, and (3) examine how "behavior problems" influences the developmental trend of PIAT scores in each trajectory class. A path diagram of the proposed GLCMM is depicted in Figure 7. The covariate "gender, b", with "1" representing female, and an explanatory latent variable "home environment, ζ" were used to model the class probability via the multinomial logit model (13). Three observed indicators, "the household variable, v_1", "the score of home cognitive stimulation, v_2", and "emotional support, v_3", were used to measure the latent variable "home environment, ζ" through Model (19). Among these indicators, v_1 is a binary variable coded with $\{0, 1\}$ representing "in household with father and other relatives" and "in household with mother," respectively, whereas v_2 and v_3 are continuous variables. The within-class LCMs are defined by Models (3), (6), and (8), in which the explanatory latent variable "behavior problems, ξ" was measured by five BPI subscales for antisocial (x_1), anxious (x_2), peer conflict (x_3), headstrong (x_4),

Table 2. Performance of the Bayesian estimates in the simulation study with Type II prior input under sample size $n = 800$

	Class 1			Class 2				Class 1			Class 2		
Par	True	Bias	RMS	True	Bias	RMS	Par	True	Bias	RMS	True	Bias	RMS
λ_{21}	0.8	0.011	0.044	0.4	0.017	0.039	ψ_{y5}	0.36	0.007	0.035	0.25	0.025	0.036
λ_{31}	0.8	0.018	0.049	0.4	0.018	0.039	ψ_{y6}	0.36	0.011	0.032	0.25	0.014	0.027
λ_{52}	0.8	0.010	0.046	0.4	0.010	0.043	ψ_{y7}	0.36	0.001	0.037	0.25	0.009	0.028
λ_{62}	0.8	0.010	0.045	0.4	0.019	0.040	ψ_{y8}	0.36	0.023	0.059	0.25	0.045	0.054
μ_1	2.0	−0.032	0.158	1.0	0.064	0.162	ψ_{x1}	0.36	0.065	0.091	0.25	0.048	0.059
μ_2	1.0	0.017	0.167	1.5	0.250	0.324	ψ_{x2}	0.36	0.008	0.028	0.25	0.007	0.020
μ_3	−2.0	−0.024	0.053	1.5	0.009	0.043	ψ_{x3}	0.36	0.012	0.035	0.25	0.007	0.024
μ_4	−1.5	−0.047	0.062	0.5	0.008	0.044	ψ_{x4}	0.36	0.069	0.098	0.25	0.037	0.050
a_{11}	0.5	−0.034	0.135	−0.5	−0.044	0.147	ψ_{x5}	0.36	0.013	0.031	0.25	0.011	0.022
a_{21}	−0.5	−0.011	0.097	0.5	0.064	0.138	ψ_{x6}	0.36	0.005	0.030	0.25	0.006	0.019
a_{31}	0.5	−0.010	0.043	−0.5	0.001	0.046	$\psi_{\delta 1}$	0.36	0.176	0.215	0.25	0.175	0.180
a_{41}	−0.5	0.002	0.031	0.5	−0.001	0.036	$\psi_{\delta 2}$	0.36	0.172	0.210	0.25	0.147	0.162
γ_{11}	0.5	−0.022	0.186	−0.5	−0.051	0.229	$\psi_{\delta 3}$	0.36	0.020	0.054	0.25	0.022	0.040
γ_{12}	0.5	−0.016	0.203	−0.5	−0.054	0.190	$\psi_{\delta 4}$	0.36	0.035	0.052	0.25	0.006	0.024
γ_{13}	0.5	0.144	0.231	−0.5	−0.069	0.192	ϕ_{11}	1.0	−0.027	0.101	1.0	−0.084	0.123
γ_{21}	0.5	−0.014	0.174	−0.5	−0.103	0.253	ϕ_{21}	0.3	−0.004	0.048	0.3	−0.052	0.081
γ_{22}	0.5	−0.012	0.167	−0.5	−0.107	0.233	ϕ_{22}	1.0	−0.014	0.091	1.0	−0.079	0.125
γ_{23}	0.5	0.069	0.208	−0.5	−0.018	0.186	φ_0	−0.5	0.044	0.130	0.0*		
γ_{31}	0.5	0.011	0.062	−0.5	−0.004	0.058	φ_1	1.0	−0.014	0.099	0.0*		
γ_{32}	0.5	−0.004	0.060	−0.5	0.012	0.053	φ_2	0.8	0.026	0.119	0.0*		
γ_{33}	0.5	0.050	0.077	−0.5	−0.011	0.077	φ_3	0.8	0.000	0.105	0.0*		
γ_{41}	0.5	0.000	0.054	−0.5	0.006	0.047							
γ_{42}	0.5	0.005	0.069	−0.5	0.000	0.049							
γ_{43}	0.5	0.075	0.095	−0.5	−0.010	0.068							

Par	True	Bias	RMS	Par	True	Bias	RMS
$\lambda_{v,21}$	0.8	0.004	0.034	ψ_{v1}	0.36	0.004	0.034
$\lambda_{v,31}$	0.8	0.004	0.038	ψ_{v2}	0.36	0.003	0.024
$\lambda_{v,52}$	0.8	0.010	0.035	ψ_{v3}	0.36	0.000	0.024
$\lambda_{v,62}$	0.8	0.004	0.034	ψ_{v4}	0.36	0.010	0.033
$\phi_{v,11}$	1.0	0.004	0.080	ψ_{v5}	0.36	0.000	0.020
$\phi_{v,21}$	0.3	−0.002	0.041	ψ_{v6}	0.36	0.004	0.025
$\phi_{v,22}$	1.0	0.002	0.076				

Note: The subscript k is suppressed for notational simplicity, and the 0's with asterisk are fixed for identification purpose.

and hyperactive behaviors (x_5). Higher scores represent a greater level of behavior problems. The scores used in this study were normed; thus, they can be treated as continuous. Given that the values of the continuous variables were quite large, the longitudinal measurements y_1 to y_3 were divided by 10 for readability, and the continuous observed indicators v_2, v_3, and x_1 to x_5 were standardized before analysis. The proposed GLCMM was used to fit the data set that involves both continuous and binary data.

The modified DIC was used to determine the number of trajectory classes, K. Let M_k denote the k-class model, $k = 1, \ldots, 5$. Given that we had little information about the model and the parameters, we assigned the vague prior inputs in (20) as follows: elements in $\{\boldsymbol{\mu}_{0k}, \boldsymbol{\Lambda}_{0\epsilon km}, \boldsymbol{\Lambda}_{0\delta kl}, \boldsymbol{\varphi}_{0k}, \boldsymbol{\Lambda}_{0vl}\}$ were fixed at 0.0, $\alpha_{0\epsilon km} = \alpha_{0\delta kl} = 9$, $\beta_{0\epsilon km} = \beta_{0\delta kl} = 4$, $\boldsymbol{\Sigma}_{0k}$, $\mathbf{H}_{0\epsilon km}$, $\mathbf{H}_{0\delta kl}$, $\boldsymbol{\Sigma}_{\varphi 0k}$; $\mathbf{H}_{0vl}$ were identity matrices with appropriate orders, $\rho_0 = \rho_{0v} = 5$, $\mathbf{R}_{0k}^{-1}$ and $\mathbf{R}_{0v}^{-1}$ were diagonal matrices with the diagonal elements 2.0. The values of the modified DICs were $DIC_{M_1} = 51{,}883$, $DIC_{M_2} = 51{,}352$, $DIC_{M_3} = 52{,}088$, $DIC_{M_4} = 53{,}835$, and $DIC_{M_5} = 55{,}406$. Therefore, the 2-class model, M_2, was selected.

Based on the random permutation sampler, we found that $\mu_{11} < \mu_{21}$ is an appropriate identifiability constraint. Furthermore, our pilot study showed that the MCMC algorithm converged within 20,000 iterations. Figure 8 shows the plot of the EPSR values against the iteration numbers. After discarding 20,000 burn-in iterations, we collected additional 20,000 observation to obtain the Bayesian estimates (EST) of the unknown parameters and their corresponding standard error estimates (SE). Table 6 summarizes the results. The main findings are as follows. First, more than one latent trajectory class exist, each of which has a different trajectory pattern. On average, the children in Class 1 have low initial PIAT scores in mathematics but a high rate of change across time, whereas those in Class 2 present an opposite pattern–high initial scores but low rate of change. In practice, there might be more classes of students, for example,

Table 3. Performance of the Bayesian estimates in the simulation study with Type I prior input under sample size $n = 1{,}600$

	Class 1			Class 2		
Par	True	Bias	RMS	True	Bias	RMS
λ_{21}	0.8	0.012	0.037	0.4	0.015	0.031
λ_{31}	0.8	0.014	0.033	0.4	0.021	0.035
λ_{52}	0.8	0.013	0.032	0.4	0.015	0.030
λ_{62}	0.8	0.009	0.030	0.4	0.016	0.030
μ_1	2.0	0.004	0.131	1.0	0.047	0.118
μ_2	1.0	0.036	0.154	1.5	0.181	0.231
μ_3	−2.0	−0.015	0.038	1.5	0.017	0.043
μ_4	−1.5	−0.020	0.034	0.5	0.012	0.031
a_{11}	0.5	−0.003	0.107	−0.5	−0.023	0.105
a_{21}	−0.5	−0.015	0.070	0.5	0.041	0.099
a_{31}	0.5	−0.004	0.027	−0.5	0.002	0.027
a_{41}	−0.5	0.000	0.027	0.5	−0.005	0.026
γ_{11}	0.5	0.007	0.136	−0.5	−0.038	0.132
γ_{12}	0.5	0.010	0.141	−0.5	−0.045	0.130
γ_{13}	0.5	0.106	0.201	−0.5	−0.005	0.154
γ_{21}	0.5	−0.006	0.121	−0.5	−0.091	0.148
γ_{22}	0.5	−0.001	0.117	−0.5	−0.076	0.162
γ_{23}	0.5	0.042	0.139	−0.5	−0.047	0.155
γ_{31}	0.5	0.003	0.041	−0.5	−0.004	0.040
γ_{32}	0.5	0.000	0.040	−0.5	−0.003	0.041
γ_{33}	0.5	0.029	0.047	−0.5	−0.061	0.103
γ_{41}	0.5	−0.001	0.039	−0.5	−0.009	0.040
γ_{42}	0.5	−0.002	0.036	−0.5	−0.004	0.044
γ_{43}	0.5	0.029	0.045	−0.5	−0.052	0.083

	Class 1			Class 2		
Par	True	Bias	RMS	True	Bias	RMS
ψ_{y5}	0.36	0.008	0.028	0.25	0.012	0.027
ψ_{y6}	0.36	0.001	0.019	0.25	0.009	0.019
ψ_{y7}	0.36	0.002	0.024	0.25	0.006	0.022
ψ_{y8}	0.36	0.021	0.046	0.25	0.024	0.037
ψ_{x1}	0.36	0.034	0.048	0.25	0.045	0.059
ψ_{x2}	0.36	0.002	0.021	0.25	0.007	0.015
ψ_{x3}	0.36	0.007	0.022	0.25	0.003	0.015
ψ_{x4}	0.36	0.036	0.053	0.25	0.044	0.058
ψ_{x5}	0.36	0.000	0.021	0.25	0.005	0.016
ψ_{x6}	0.36	0.004	0.020	0.25	0.007	0.016
$\psi_{\delta 1}$	0.36	0.172	0.212	0.25	0.165	0.179
$\psi_{\delta 2}$	0.36	0.119	0.171	0.25	0.125	0.140
$\psi_{\delta 3}$	0.36	0.006	0.033	0.25	0.013	0.030
$\psi_{\delta 4}$	0.36	0.015	0.032	0.25	0.009	0.023
ϕ_{11}	1.0	−0.022	0.078	1.0	−0.062	0.098
ϕ_{21}	0.3	−0.008	0.042	0.3	−0.021	0.051
ϕ_{22}	1.0	−0.014	0.064	1.0	−0.064	0.103
φ_0	−0.5	0.015	0.094	0.0*		
φ_1	1.0	0.003	0.078	0.0*		
φ_2	0.8	0.010	0.090	0.0*		
φ_3	0.8	0.002	0.089	0.0*		

Par	True	Bias	RMS	Par	True	Bias	RMS
$\lambda_{v,21}$	0.8	−0.003	0.023	ψ_{v1}	0.36	0.004	0.026
$\lambda_{v,31}$	0.8	0.000	0.024	ψ_{v2}	0.36	0.006	0.018
$\lambda_{v,52}$	0.8	0.001	0.023	ψ_{v3}	0.36	0.001	0.020
$\lambda_{v,62}$	0.8	−0.001	0.025	ψ_{v4}	0.36	0.004	0.023
$\phi_{v,11}$	1.0	0.010	0.049	ψ_{v5}	0.36	0.004	0.020
$\phi_{v,21}$	0.3	0.005	0.027	ψ_{v6}	0.36	0.003	0.016
$\phi_{v,22}$	1.0	0.002	0.052				

Note: The subscript k is suppressed for notational simplicity, and the 0's with asterisk are fixed for identification purpose.

students with low (high) initial PIAT scores in mathematics also have low (high) rates of change across time. However, we did not identify such classes of students in the NLSY data. Second, the influence of "behavior problems" on the latent growth factors is significant in Class 2 but not in Class 1. Third, in the multinomial logit model, φ_1 and φ_2 are significantly positive, indicating that girls and children with a better "home environment" are more likely to belong to Class 1. Finally, some of the other parameters, including λ's, ψ_ϵ's, ϕ's, and ψ_δ's, vary substantially across classes. This finding confirms the existence of heterogeneity within the NLSY data.

Using a single PC with an Intel Core i3 CPU 550@3.20 GHz and 4.00 GB RAM, the computing time for obtaining the Bayesian estimates of parameters and the modified DIC value under M_2 was approximately 100 minutes. Our program is written in C language and is available upon request.

To assess the sensitivity of the Bayesian estimation and model selection to different prior inputs, the above analysis was repeated with some disturbances to the current prior input. Still, the 2-class model was selected, and the Bayesian estimates of the unknown parameters under the selected model were close to those presented in Table 6. The results are not reported here.

6. DISCUSSION

In this article, an integrated generalized mixture LCM was established under the Bayesian framework. Unlike traditional LCMs, this integrated model can simultaneously handle different kinds of latent and observed variables, such as (1) continuous latent variables, including latent growth factors and explanatory latent variables; (2) a latent allocation variable representing latent trajectory classes; (3) fixed and latent effects in predicting latent growth factors and probabilities of subjects' class memberships; and (4) mixed continuous and discrete observed variables that follow the EFD. The proposed GLCMM represents a broad class of statistical models flexible enough to detect the heterogeneity of a longitudinal trajectory pattern, explore fixed and latent effects that influence both longitudinal change patterns

Table 4. Performance of the Bayesian estimates in the simulation study with Type II prior input under sample size $n = 1{,}600$

	Class 1			Class 2				Class 1			Class 2		
Par	True	Bias	RMS	True	Bias	RMS	Par	True	Bias	RMS	True	Bias	RMS
λ_{21}	0.8	0.007	0.036	0.4	0.011	0.028	ψ_{y5}	0.36	0.004	0.032	0.25	0.004	0.022
λ_{31}	0.8	0.008	0.030	0.4	0.012	0.027	ψ_{y6}	0.36	−0.003	0.021	0.25	0.002	0.015
λ_{52}	0.8	0.013	0.036	0.4	0.005	0.030	ψ_{y7}	0.36	0.003	0.024	0.25	0.001	0.018
λ_{62}	0.8	0.013	0.032	0.4	0.004	0.027	ψ_{y8}	0.36	−0.002	0.043	0.25	0.008	0.036
μ_1	2.0	−0.028	0.112	1.0	0.030	0.109	ψ_{x1}	0.36	0.036	0.056	0.25	0.019	0.033
μ_2	1.0	−0.029	0.138	1.5	0.123	0.185	ψ_{x2}	0.36	0.004	0.022	0.25	−0.001	0.014
μ_3	−2.0	−0.010	0.030	1.5	0.008	0.031	ψ_{x3}	0.36	0.003	0.022	0.25	0.002	0.013
μ_4	−1.5	−0.025	0.037	0.5	−0.001	0.026	ψ_{x4}	0.36	0.035	0.060	0.25	0.013	0.034
a_{11}	0.5	−0.035	0.102	−0.5	−0.011	0.114	ψ_{x5}	0.36	0.002	0.026	0.25	0.000	0.017
a_{21}	−0.5	0.005	0.066	0.5	0.006	0.072	ψ_{x6}	0.36	0.004	0.022	0.25	0.002	0.017
a_{31}	0.5	−0.005	0.028	−0.5	−0.001	0.030	$\psi_{\delta 1}$	0.36	0.049	0.156	0.25	0.046	0.060
a_{41}	−0.5	0.006	0.029	0.5	0.002	0.022	$\psi_{\delta 2}$	0.36	0.052	0.123	0.25	0.047	0.077
γ_{11}	0.5	−0.011	0.139	−0.5	−0.002	0.128	$\psi_{\delta 3}$	0.36	0.004	0.036	0.25	0.006	0.027
γ_{12}	0.5	−0.007	0.143	−0.5	−0.001	0.143	$\psi_{\delta 4}$	0.36	0.018	0.035	0.25	0.005	0.023
γ_{13}	0.5	0.077	0.174	−0.5	−0.013	0.134	ϕ_{11}	1.0	−0.024	0.075	1.0	−0.041	0.089
γ_{21}	0.5	−0.028	0.114	−0.5	−0.074	0.145	ϕ_{21}	0.3	−0.022	0.048	0.3	−0.029	0.055
γ_{22}	0.5	−0.027	0.116	−0.5	−0.051	0.142	ϕ_{22}	1.0	−0.025	0.074	1.0	−0.031	0.085
γ_{23}	0.5	0.027	0.140	−0.5	−0.030	0.153	φ_0	−0.5	0.013	0.091	0.0*		
γ_{31}	0.5	0.003	0.041	−0.5	−0.008	0.039	φ_1	1.0	0.006	0.075	0.0*		
γ_{32}	0.5	0.003	0.037	−0.5	0.009	0.037	φ_2	0.8	0.007	0.082	0.0*		
γ_{33}	0.5	0.024	0.051	−0.5	−0.002	0.041	φ_3	0.8	−0.010	0.087	0.0*		
γ_{41}	0.5	−0.008	0.039	−0.5	0.001	0.039							
γ_{42}	0.5	0.006	0.041	−0.5	0.006	0.036							
γ_{43}	0.5	0.044	0.071	−0.5	−0.001	0.035							

Par	True	Bias	RMS	Par	True	Bias	RMS
$\lambda_{v,21}$	0.8	−0.002	0.022	ψ_{v1}	0.36	0.002	0.024
$\lambda_{v,31}$	0.8	−0.001	0.022	ψ_{v2}	0.36	0.000	0.017
$\lambda_{v,52}$	0.8	−0.001	0.024	ψ_{v3}	0.36	0.005	0.019
$\lambda_{v,62}$	0.8	0.001	0.022	ψ_{v4}	0.36	−0.003	0.024
$\phi_{v,11}$	1.0	0.004	0.050	ψ_{v5}	0.36	0.001	0.019
$\phi_{v,21}$	0.3	−0.007	0.030	ψ_{v6}	0.36	0.001	0.017
$\phi_{v,22}$	1.0	0.004	0.049				

Note: The subscript k is suppressed for notational simplicity, and the 0's with asterisk are fixed for identification purpose.

Table 5. Performance of the modified DIC in model selection with Type I and Type II prior inputs under sample sizes $n = 800$ and $n = 1{,}600$ (the presented values are means and standard deviations)

	$n = 800$		$n = 1{,}600$	
	Prior I	Prior II	Prior I	Prior II
M_1	50410.99 (340.27)	50029.04 (415.41)	101261.54 (545.98)	100995.04 (584.94)
M_2	45979.87 (267.76)	45131.99 (315.20)	91574.05 (470.69)	90331.27 (490.70)
M_3	47182.20 (1022.62)	48237.50 (2642.64)	100469.55 (6369.87)	99088.84 (7086.21)

and latent class membership, and accommodate mixed-data types. Our proposed Bayesian approach, which includes the Gibbs sampler, the MH algorithm, the permutation sampler, and the modified DIC, solves computational challenges that arise from the GLCMM modeling efforts. The NLSY example demonstrates how the proposed GLCMM can be applied to longitudinal data and how it can test hypothesis outcomes that are generally not achievable using conventional methods.

The proposed GLCMM can be extended to several directions. First, our model assumes that data are collected with the same number of waves and the same spacing of waves among individuals (i.e., balanced data). However, unbalanced setting is common in longitudinal data. Hence, extending the current model to deal with unbalanced data is worthy of further consideration. Second, the proposed model can be extended to include longitudinal latent effects to predict latent growth factors and the probability of class membership. Third, given that ordered and unordered categorical data are very common in practical research, the idea of probit (logit) regression or cumulative probit (logit) regression methods [1, 4, 33, 36] could be applied to incorporate these

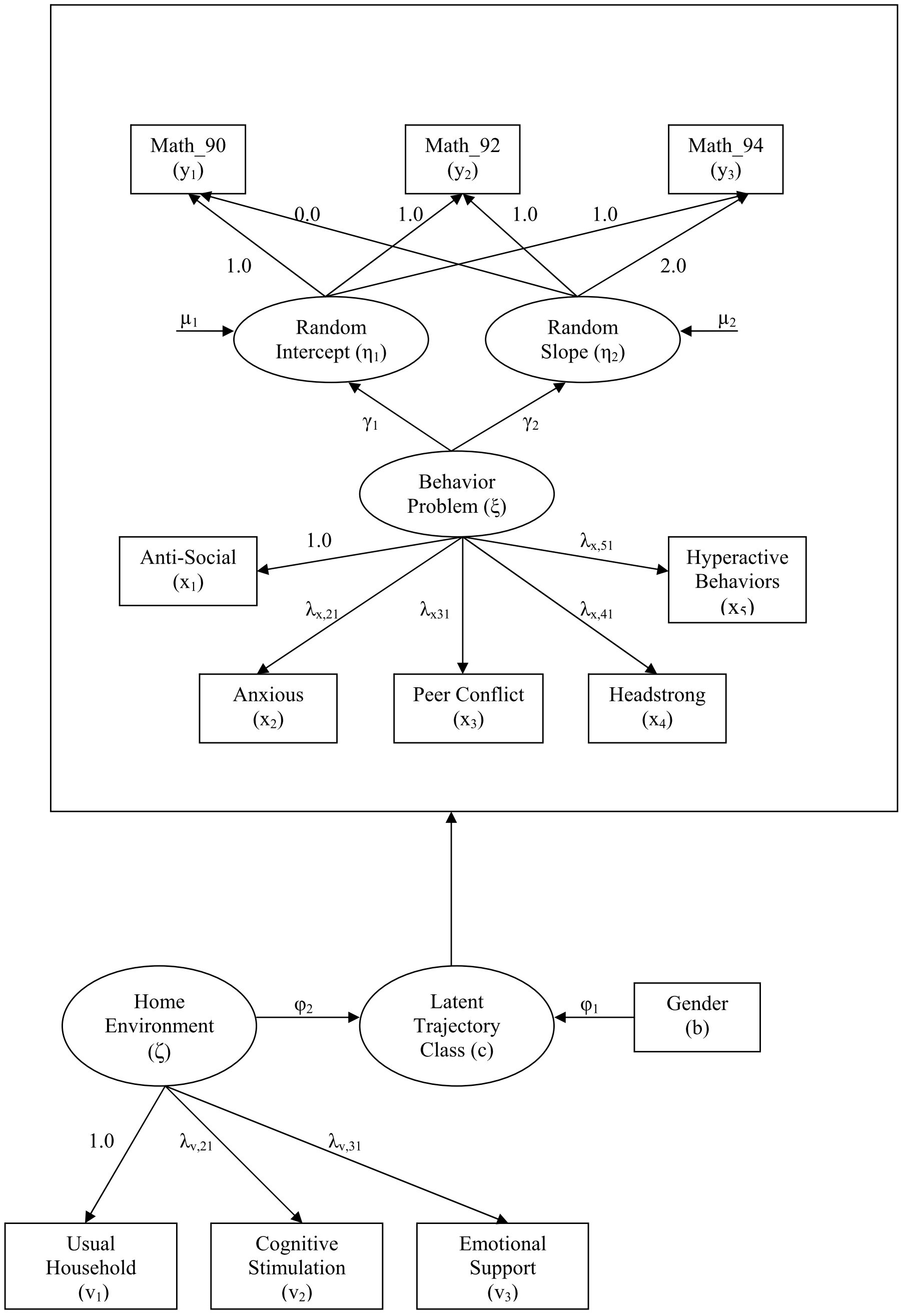

Figure 7. Path diagram of proposed model in the illustrative example.

types of data in the current model. Finally, except for determining the number of latent trajectory classes, the modified DIC can also be used to select appropriate linear or nonlinear parametric functions $f_1, \ldots, f_h$. Moreover, the parametric trajectory equation and parametric functions $f_1, \ldots, f_h$ in the current model can be further extended to nonparametric forms. The Bayesian P-splines approach [13, 34, 35] is a promising candidate for the nonparametric modeling of LCMs.

APPENDIX A. FULL CONDITIONAL DISTRIBUTIONS

(a) The full conditional distribution of $\boldsymbol{\varphi}_k$:

$$p(\boldsymbol{\varphi}_k|\mathbf{U}, \mathbf{V}, \boldsymbol{\Omega}, \widetilde{\boldsymbol{\Omega}}, \mathbf{C}, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} p(c_i|\boldsymbol{\varphi}_k, \boldsymbol{\theta}, \mathbf{u}_i, \mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\zeta}_i)p(\boldsymbol{\varphi}_k)$$

Table 6. Bayesian estimates and their corresponding standard error estimates in the illustrative example

	Class 1		Class 2	
Par	EST	SE	EST	SE
μ_1	2.613	0.036	4.321	0.209
μ_2	1.076	0.013	0.462	0.104
γ_1	−0.058	0.061	−0.687	0.147
γ_2	−0.042	0.026	0.192	0.068
λ_{x21}	0.926	0.051	0.889	0.095
λ_{x31}	1.119	0.054	0.706	0.071
λ_{x41}	0.988	0.049	0.738	0.071
λ_{x51}	0.761	0.046	1.219	0.128
ϕ_x	0.435	0.034	1.987	0.456
$\psi_{\epsilon 1}$	0.419	0.027	0.400	0.103
$\psi_{\epsilon 2}$	0.336	0.021	0.659	0.171
$\psi_{\epsilon 3}$	0.288	0.025	0.374	0.097
$\psi_{\epsilon 4}$	0.457	0.026	0.766	0.187
$\psi_{\epsilon 5}$	0.567	0.025	0.418	0.115
$\psi_{\epsilon 6}$	0.439	0.024	0.337	0.069
$\psi_{\epsilon 7}$	0.554	0.025	0.307	0.061
$\psi_{\epsilon 8}$	0.574	0.026	0.824	0.217
$\psi_{\delta 1}$	1.035	0.049	0.674	0.170
$\psi_{\delta 2}$	0.067	0.006	0.218	0.045
φ_0	2.118	0.284	0.0*	
φ_1	4.321	0.209	0.0*	
φ_2	0.462	0.104	0.0*	
Par. related to ζ				
λ_{v21}	1.525	0.259		
λ_{v31}	1.273	0.243		
ϕ_v	0.114	0.025		
ψ_{v1}	0.735	0.077		
ψ_{v2}	0.816	0.056		

Note: The subscript k is suppressed for notational simplicity, and the 0's with asterisk are fixed for identification purpose.

$$\propto \exp\Big\{ \sum_{i:c_i=k} \boldsymbol{\varphi}_k^T \mathbf{G}(\boldsymbol{\zeta}_i) - \sum_{i=1}^n \log\big[\sum_{j=1}^K \exp(\boldsymbol{\varphi}_j^T \mathbf{G}(\boldsymbol{\zeta}_i))\big] - \frac{1}{2}(\boldsymbol{\varphi}_k - \boldsymbol{\varphi}_{0k})^T \boldsymbol{\Sigma}_{\varphi 0k}^{-1} (\boldsymbol{\varphi}_k - \boldsymbol{\varphi}_{0k}) \Big\},$$

where $\mathbf{G}(\boldsymbol{\zeta}_i) = (1, \mathbf{b}_i^T, \boldsymbol{\zeta}_i^T)^T$.

(b) The full conditional distribution of $\widetilde{\boldsymbol{\Omega}}$:

$$\begin{aligned} p(\boldsymbol{\zeta}_i|\mathbf{u}_i, \mathbf{v}_i, \boldsymbol{\omega}_i, \mathbf{c}_i = k, \boldsymbol{\theta}) &\propto p(\mathbf{v}_i|\boldsymbol{\zeta}_i, \boldsymbol{\theta}) p(c_i = k|\boldsymbol{\zeta}_i, \boldsymbol{\theta}) p(\boldsymbol{\zeta}_i|\boldsymbol{\theta}) \\ &\propto \exp\Big\{ -\frac{1}{2}(\mathbf{v}_i - \boldsymbol{\Lambda}_v \boldsymbol{\zeta}_i)^T \boldsymbol{\Psi}_v^{-1} (\mathbf{v}_i - \boldsymbol{\Lambda}_v \boldsymbol{\zeta}_i) \\ &\quad - \frac{1}{2}\boldsymbol{\zeta}_i^T \boldsymbol{\Phi}_v^{-1} \boldsymbol{\zeta}_i + \boldsymbol{\varphi}_k^T \mathbf{G}(\boldsymbol{\zeta}_i) - \log\big[\sum_{j=1}^K \exp(\boldsymbol{\varphi}_j^T \mathbf{G}(\boldsymbol{\zeta}_i))\big] \Big\}. \end{aligned}$$

(c) The full conditional distribution of $\boldsymbol{\Omega}$:

$$p(\boldsymbol{\Omega}|\mathbf{U}, \mathbf{V}, \mathbf{C}, \widetilde{\boldsymbol{\Omega}}, \boldsymbol{\theta}) = \prod_{i=1}^n p(\boldsymbol{\omega}_i|\mathbf{u}_i, \mathbf{v}_i, \mathbf{c}_i, \boldsymbol{\zeta}_i, \boldsymbol{\theta}), \quad \text{and}$$

$$\begin{aligned} p(\boldsymbol{\omega}_i|\mathbf{u}_i, \mathbf{v}_i, \mathbf{c}_i, \boldsymbol{\zeta}_i, \boldsymbol{\theta}) &\propto \exp\Big\{ \sum_{j=1}^p [u_{ij}\vartheta_{kij} - b_j(\vartheta_{kij})]/\psi_{\epsilon kj} - \frac{1}{2}\boldsymbol{\xi}_i^T \boldsymbol{\Phi}_k^{-1} \boldsymbol{\xi}_i \\ &\quad - \frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_{\delta k}\mathbf{G}(\boldsymbol{\omega}_i))^T \\ &\quad \times \boldsymbol{\Psi}_{\delta k}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_{\delta k}\mathbf{G}(\boldsymbol{\omega}_i)) \Big\}, \end{aligned}$$

where $\boldsymbol{\Lambda}_{\delta k} = (\mathbf{A}_k, \boldsymbol{\Pi}_k, \boldsymbol{\Gamma}_k)$, $\mathbf{G}(\boldsymbol{\omega}_i) = (\mathbf{w}_i^T, \boldsymbol{\eta}_i^T, \mathbf{F}(\boldsymbol{\xi}_i)^T)^T$.

(d) The full conditional distribution of $\mathbf{C}$:

$$p(\mathbf{C}|\mathbf{U}, \mathbf{V}, \boldsymbol{\Omega}, \widetilde{\boldsymbol{\Omega}}, \boldsymbol{\theta}) = \prod_{i=1}^n p(c_i|\mathbf{u}_i, \mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\zeta}_i, \boldsymbol{\theta}), \quad \text{and}$$

$$p(c_i = k|\mathbf{u}_i, \mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\zeta}_i, \boldsymbol{\theta}) = \frac{p(\mathbf{u}_i|c_i = k, \boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|c_i = k, \boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|c_i = k, \boldsymbol{\theta}) p(c_i = k|\boldsymbol{\zeta}_i, \boldsymbol{\theta})}{\sum_{j=1}^K p(\mathbf{u}_i|c_i = j, \boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|c_i = j, \boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|c_i = j, \boldsymbol{\theta}) p(c_i = j|\boldsymbol{\zeta}_i, \boldsymbol{\theta})}.$$

(e) The full conditional distribution of $\boldsymbol{\Lambda}_{kj}$: for $j = 1, \ldots, p$,

$$\begin{aligned} p(\boldsymbol{\Lambda}_{kj}|\mathbf{U}, \mathbf{V}, \boldsymbol{\Omega}, \widetilde{\boldsymbol{\Omega}}, \mathbf{C}, \boldsymbol{\theta}) &\propto \exp\Big\{ -\frac{1}{2}\psi_{\epsilon kj}^{-1}(\boldsymbol{\Lambda}_{kj} - \boldsymbol{\Lambda}_{0\epsilon kj})^T \mathbf{H}_{0\epsilon kj}^{-1}(\boldsymbol{\Lambda}_{kj} - \boldsymbol{\Lambda}_{0\epsilon kj}) \\ &\quad + \sum_{i:c_i=k} [u_{ij}\vartheta_{kij} - b_j(\vartheta_{kij})]/\psi_{\epsilon kj} \Big\}. \end{aligned}$$

The full conditional distributions of other parameters in $\boldsymbol{\theta}$ are the normal, Gamma, and inverted Wishart distributions. They are not presented here.

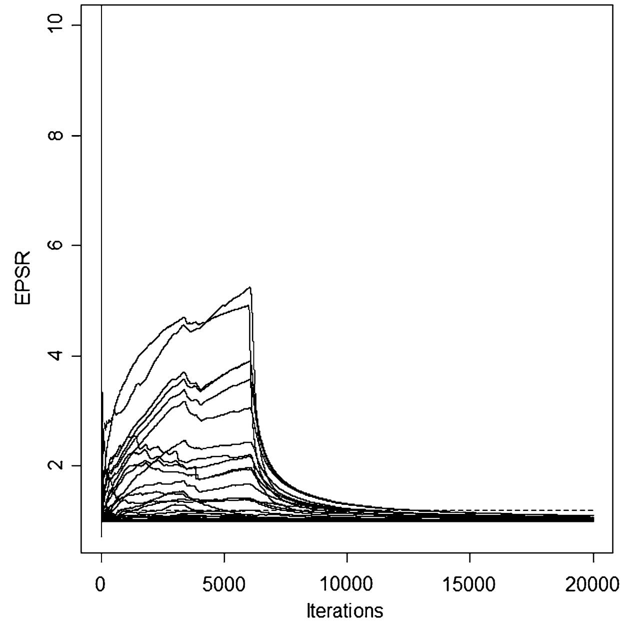

Figure 8. Plot of the estimated potential scale reduction values against the iteration numbers in the illustrative example.

Received 27 April 2012

REFERENCES

[1] Agresti, A. (2002). *Categorical Data Analysis*. New Jersey: John Wiley & Sons, Inc. MR1914507

[2] Bollen, K. A. (1989). *Structural Equation Models with Latent Variables*. New Jersey: John Wiley & Sons, Inc. MR0996025

[3] Bollen, K. A., and Curran, P. J. (2006). *Latent Curve Models – A Structural Equation Perspective*. New Jersey: John Wiley & Sons, Inc. MR2184502

[4] Cai, J. H., Song, X. Y., Lam, K. H., and Ip, H. S. (2011). A mixture of generalized latent variable models for mixed mode and heterogeneous data. *Computational Statistics and Data Analysis* **55** 2889–2907. MR2813054

[5] Caldwell, B. M., and Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas at Little Rock, Center for Child Development and Education.

[6] Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1** 651–674. MR2282197

[7] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39** 1–38. MR0501537

[8] Duncan, T. E., Duncan, S. C., and Strycker, L. A. (2006). *An Introduction to Latent Variable Growth Curve Modeling*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

[9] Diebolt, J., and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B* **56** 363–375. MR1281940

[10] Dunn, L. M., and Markwardt, F. C. (1970). *Peabody Individual Achievement Test Manual*. Circle Pines, NM: American Guidance Services.

[11] Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62** 355–366. MR1749544

[12] Elliott, M. R., Gallo, J. J., Gonger, H. R., and Katz, I. R. (2005). Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics* **6** 119–143.

[13] Fahrmeir, L. and Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika* **72** 327–346. MR2361960

[14] Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96** 194–208. MR1952732

[15] Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

[16] Gelman, A. (1996). Inference and monitoring convergence. In: *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds. London: Chapman and Hall, pp. 131–144. MR1397966

[17] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika* **57** 97–109.

[18] Lee, S. Y. (2007). *Structural Equation Modeling – A Bayesian Approach*. New York: John Wiley. MR2290938

[19] Lee, S. Y. and Song, X. Y. (2004). Evaluation of Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* **39** 653–686.

[20] Lee, S. Y., Song, X. Y., and Poon, W.Y. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research* **39** 37–67.

[21] Li, Y., and Wang, H. Z. (2010). Bayesian analysis for finite mixture in non-recursive non-linear structural equation models. *British Journal of Mathematical and Statistical Psychology* **63** 361–377 MR2756883

[22] McArdle, J. J., and Bell, R. Q. (2000). Recent trends in modeling longitudinal data by latent growth curve methods. In: *Modeling Longitudinal and Multiple-group Data: Practical Issues, Applied Approaches, and Scientific Examples*, T. D. Little, K. U. Schnabel, & J. Baumert, eds. Mahwah, NJ: Lawrence Erlbaum, pp. 69–108.

[23] Meredith, W., and Tisak, J. (1984). *On "Tuckerizing" curves*. Presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

[24] Meredith, W., and Tisak, J. (1990). Latent curve analysis. *Psychometrika* **55**(1) 107–122.

[25] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics* **21** 1087–1091.

[26] Muthén, B. O. (2001). Latent variable mixture modeling. In *New Developments and Techniques in Structural Equation Modeling*, G. A. Marcoulides & R. E. Schumacker, eds. Lawrence Erlbaum Associates, pp. 1–33.

[27] Muthén, B. O., and Curran, J. P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods* **2** 371–402.

[28] Muthén, B. O., and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55** 463–469.

[29] Pan, J. H., Song, X. Y., Lee, S. Y., and Kwok, T. (2008). Longitudinal analysis of quality of life for stroke survivors using latent curve models. *Stroke* **39** 2795–2802.

[30] Roeder, C., and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association* **92** 894–902. MR1482121

[31] Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B* **59** 667–678.

[32] Scheines, R., Hoijtink, H., and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika* **64** 37–52.

[33] Song, X. Y. and Lee, S. Y. (2012). *Basic and Advanced Structural Equation Modeling: With Applications in Medical and Behavioural Sciences*. John Wiley.

[34] Song, X. Y., and Lu, Z. H. (2010). Semiparametric latent variable models with Bayesian P-splines. *Journal of Computational and Graphical Statistics* **19** 590–608. MR2732494

[35] Song, X. Y., and Lu, Z. H. (2012). Semiparametric transformation models with Bayesian P-splines. *Statistics and Computing* **22** 1085–1098. MR2950087

[36] Song, X. Y., Lu, Z. H., Cai, J. H., and Ip, H. S. (2012). A Bayesian modeling approach for generalized semiparametric structural equation models. *Psychometrika*, accepted.

[37] Song, X. Y., Lee, S. Y., and Hser, Y.I. (2009). Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects. *Structural Equation Modeling: A Multidisciplinary Journal* **16(2)** 245–266. MR2649696

[38] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measure of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64** 583–639. MR1979380

[39] Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2003). *WinBUGS User Manual. Version 1.4*. Cambridge, England: MRC Biostatistics Unit.

[40] Tanner, M. A., and Wong, W. H. (1987). The calculation of posterier distribution by data augmentation (with discussion). *Journal of the American Statistical Association* **86** 79–86. MR0898357

[41] Willett, J. B., and Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of change. *Psychological Bulletin* **116** 363–381.

[42] Yung, Y. F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika* **62** 297–330.

[43] Zill, N. (1985). *Behavior Problem Scales Developed From the 1981 Child Health Supplement to the National Health Interview Survey*. Washington DC: Child Trends, Inc.

Jun-Hao Pan
Department of Psychology
Sun Yat-Sen University
Guangzhou
China
E-mail address: panjunh@mail.sysu.edu.cn

Xin-Yuan Song
Department of Statistics
The Chinese University of Hong Kong
Hong Kong
China
E-mail address: xysong@sta.cuhk.edu.hk

Edward H. Ip
Department of Biostatistical Sciences
Division of Public Health Sciences
Wake Forest University Health Sciences, NC
United States
E-mail address: eip@wakehealth.edu