Computational Statistics and Data Analysis 71 (2014) 220–240

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application☆

Zhenqiu (Laura) Lu[a,*], Zhiyong Zhang[b]

[a] University of Georgia, United States

[b] University of Notre Dame, United States

HIGHLIGHTS

- Four non-ignorable missingness models are proposed.
- Three robust models to deal with outliers are proposed.
- A full Bayesian method is implemented.
- Model selection criteria are proposed in a Bayesian context.
- Three simulation studies and one real data case study are conducted.

ARTICLE INFO

Article history:
Received 3 July 2012
Received in revised form 25 July 2013
Accepted 26 July 2013
Available online 7 August 2013

Keywords:
Growth mixture models
Non-ignorable missing data
Robust methods
Bayesian method
Model selecting criteria

ABSTRACT

Challenges in the analyses of growth mixture models include missing data, outliers, estimation, and model selection. Four non-ignorable missingness models to recover the information due to missing data, and three robust models to reduce the effect of non-normality are proposed. A full Bayesian method is implemented by means of data augmentation algorithm and Gibbs sampling procedure. Model selection criteria are also proposed in the Bayesian context. Simulation studies are then conducted to evaluate the performances of the models, the Bayesian estimation method, and selection criteria under different situations. The application of the models is demonstrated through the analysis of education data on children's mathematical ability development. The models can be widely applied to longitudinal analyses in medical, psychological, educational, and social research.

1. Introduction

Mixture models offer natural models for unobserved population heterogeneity. The importance of mixture models, their enormous developments, and their frequent applications are not only remarked by a number of recent books but also by a diversity of journal publications. For example, Computational Statistics & Data Analysis has published two special issues on mixture models (Bohning and Seidel, 2003; Bohning et al., 2007) and the current issue is a new one. Latent growth models are used to study individuals' latent growth trajectories by analyzing the variables of interest on the same individuals repeatedly through time (e.g., Bollen and Curran, 2006; McArdle and Bell, 1999; Meredith and Tisak, 1990). These models are very popular in biological, psychological, educational, and social sciences (e.g., Collins, 1991; Fitzmaurice et al., 2004; Singer and Willett, 2003). By combining latent growth models and finite mixture models (e.g., McLachlan and Peel, 2000),

☆ Supplementary tables summarized from simulation results can be accessed from the web site of http://nd.psychstat.org/research/csda2013.

* Correspondence to: 325 Aderhold Hall, University of Georgia, Athens, GA 30602, United States. Tel.: +1 706 542 4540; fax: +1 706 542 4240.
E-mail address: zlu@uga.edu (Z. Lu).

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.07.036

growth mixture models (GMMs, see, e.g., Lubke and Muthén, 2005; Muthén, 2004; Muthén et al., 2011), therefore, provide researchers with a flexible set of models for growth data with latent population heterogeneity.

However, with the increase in complexity of model specification comes an increase in difficulties estimating GMMs. First, missing data are almost inevitable (e.g., Little and Rubin, 2002; Yuan and Lu, 2008), especially in longitudinal studies (e.g., Jelicic et al., 2009; Roth, 1994). Little and Rubin (2002) distinguished *ignorable* and *non-ignorable* missingness mechanisms. Non-ignorable missingness is a crucial and serious concern, because not attending to it may result in severely biased statistical estimates, standard errors, and associated confidence intervals (e.g., Little and Rubin, 2002; Schafer, 1997; Zhang and Wang, 2012). However, most of the literature on the problems of missing data focuses on ignorable missingness (e.g., Schafer and Graham, 2002). Second, data may have outliers (e.g., Hoaglin et al., 1983), particularly in social and behavioral sciences (e.g., Micceri, 1989). The consequences of applying a normal distribution assumption to such data include unreliable parameter estimates (e.g., Pan and Fang, 2002), unreliable standard errors and confidence intervals, and misleading statistical tests and inference (e.g., Yuan and Bentler, 1998). Third, for complex models such as GMMs with missing data and outliers, maximum likelihood methods might fail or provide biased estimates (e.g., Yuan and Zhang, 2012). Most of the previous estimations have relied on maximum likelihood methods for parameter estimation and have carried out inferences through conventional likelihood procedures (e.g., Song et al., 2014). Fourth, even with effective estimation methods, model selection in such complex situations becomes extremely difficult. Traditional criteria for model selection, including Akaike's Information Criterion (AIC, Akaike, 1974), Bayesian Information Criterion (BIC, Schwarz, 1978), consistent Akaike's Information Criterion (CAIC, Bozdogan, 1987), sample-size adjusted Bayesian Information Criterion (ssBIC, Sclove, 1987), and Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), are not uniformly effective due to latent effects and missing data (e.g., Celeux et al., 2006).

Few studies have discussed how to address these common problems in longitudinal research in the framework of GMMs. Lu et al. (2011) discussed GMMs with non-ignorable missing data using Bayesian methods. However, they (1) considered only one type of non-ignorable missingness, (2) assumed data are normally distributed without any outlier, and (3) did not propose any model selection criterion.

This article extends the study of Lu et al. (2011) and addresses these challenges in GMMs: missing data, outliers, estimation, and model selection. Regarding missing data, we propose new types of non-ignorable missingness in GMMs and investigate their influences on model estimation under different situations. Regarding outliers, we use robust models (e.g., Lange et al., 1989) to minimize the effects of contaminated data. Because convenient robust methods often lead to other problems such as under-estimation of standard errors (e.g., Poon and Poon, 2002), we adopt t-distributions to deal with heavy-tailed data (Lin et al., 2004; Zhang et al., 2013). Regarding estimation methods, as Bayesian methods provide many advantages of estimating complex models (e.g., Dunson, 2000), we propose a full Bayesian approach, which is flexible enough to estimate a variety of models with different missing data mechanisms, contaminated data, and mixture structure. Regarding model selection, we propose several selection criteria in the Bayesian context. The performances of these criteria are investigated under different situations.

In the next section of this article, Section 2, we propose GMMs with different types of missing data and outliers. In Section 3, we present Bayesian estimation methods. In Section 4, we propose Bayesian model selection criteria. In Section 5, we conduct three simulation studies on Bayesian GMMs under different conditions. In Section 6, we demonstrate the application of the GMMs and the Bayesian method by analyzing real education data on children's mathematical ability development. In Section 7, we draw conclusions. The Appendices present the technical details of our analyses.

2. Models

The density function of a growth mixture model is

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i), \tag{1}$$

where π_k is the invariant class probability (or weight) for class k, ($k = 1, \ldots, K$), satisfying $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$ (e.g., McLachlan and Peel, 2000), and $f_k(\mathbf{y}_i)$ is the density for the kth class, in which $\mathbf{y}_i$ is a $T \times 1$ vector of outcomes for participant i ($i = 1, \ldots, N$) following a latent growth model

$$\begin{cases} \mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{e}_i, \\ \boldsymbol{\eta}_i = \boldsymbol{\beta} + \boldsymbol{\xi}_i, \end{cases} \tag{2}$$

where $\boldsymbol{\eta}_i$ is a $q \times 1$ vector of latent effects, $\boldsymbol{\Lambda}$ is a $T \times q$ matrix of factor loadings for $\boldsymbol{\eta}_i$, $\mathbf{e}_i$ is a $T \times 1$ vector of residual or measurement errors, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fix-effects, and $\boldsymbol{\xi}_i$ captures the variation of $\boldsymbol{\eta}_i$.

In the Extended Growth Mixture Models (EGMMs, Muthén and Shedden, 1999), π_k is not invariant any more for all individuals in class k. It is allowed to vary individually depending on covariates, so it is expressed as $\pi_{ik}(\mathbf{x}_i)$. In this study, a probit link function is used

$$\begin{cases} \pi_{i1}(\mathbf{x}_i) = \Phi(X_i' \boldsymbol{\varphi}_1), \\ \pi_{ik}(\mathbf{x}_i) = \Phi(X_i' \boldsymbol{\varphi}_k) - \Phi(X_i' \boldsymbol{\varphi}_{k-1}), \quad (k = 2, 3, \ldots, K-1) \\ \pi_{iK}(\mathbf{x}_i) = 1 - \Phi(X_i' \boldsymbol{\varphi}_{K-1}), \end{cases} \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution, and $X_i = (1, \mathbf{x}_i')'$ with an $r \times 1$ vector of observed covariates $\mathbf{x}_i$. Note that $\Phi(X_i' \boldsymbol{\varphi}_k) = \sum_{j=1}^{k} \pi_{ij}(\mathbf{x}_i)$ and $\Phi(X_i' \boldsymbol{\varphi}_K) \equiv 1$.

In both cases, a dummy variable $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iK})'$ is used to indicate the class membership. If individual i comes from group k, $z_{ik} = 1$ and $z_{ij} = 0$ $(\forall j \neq k)$. $\mathbf{z}_i$ is multinomially distributed (McLachlan and Peel, 2000, p. 7).

2.1. Non-ignorable missingness

To build models with non-ignorable missingness, we use selection models (Glynn et al., 1986; Little, 1993, 1995) instead of pattern mixture models (Little and Rubin, 2002), in part, because substantively it is more natural to consider the behavior of the response variable in the full target population of interests, rather than in sub-populations defined by missing data pattern (e.g., Fitzmaurice et al., 2008). For individual i, the complete-data likelihood function (see, Celeux et al., 2006) of a selection model with auxiliary latent variables is expressed as

$$L_i = \prod_{k=1}^{K} \left[\pi_{ik}(\mathbf{x}_i)\, f_k(\boldsymbol{\eta}_i)\, f_k(\mathbf{y}_i|\boldsymbol{\eta}_i)\, f_k(\mathbf{m}_i|\mathbf{z}_i, \boldsymbol{\eta}_i, \mathbf{y}_i, \mathbf{x}_i)\right]^{z_{ik}}, \tag{4}$$

where $\mathbf{m}_i = (m_{i1}, m_{i2}, \ldots, m_{iT})'$ is a missing data indicator for $\mathbf{y}_i$, with $m_{it} = 1$ when y_{it} is missing and 0 when observed. Let $\tau_{it} = p(m_{it} = 1)$ be the probability that y_{it} is missing, then $m_{it} \sim$ Bernoulli(τ_{it}). τ_{it} depends on the non-ignorable missingness mechanisms. Lu et al. (2011) proposed Latent-Class-Dependent (LCD) missingness (see Fig. 1 panel (a)) in which τ_{it} is assumed to depend on latent class membership and observed covariates,

$$\tau_{it} = \Phi(\mathbf{z}_i' \boldsymbol{\gamma}_{zt}^* + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}), \tag{5}$$

where $\boldsymbol{\gamma}_{zt}^* = (\gamma_{zt_1}^*, \gamma_{zt_2}^*, \ldots, \gamma_{zt_K}^*)'$ is the coefficient vector for the class membership $\mathbf{z}_i$, and $\boldsymbol{\gamma}_{xt} = (\gamma_{xt_1}, \gamma_{xt_2}, \ldots, \gamma_{xt_r})'$ is the $r \times 1$ coefficient vector for covariates. For LCD, the missingness is ignorable within each latent class.

In reality, however, the missingness mechanism in each class may also be non-ignorable. Lu et al. (submitted for publication) illustrated some possible missingness. For example, in a given latent class, students may miss a test when they have few prior knowledge of that course (i.e., low latent initial level), or when their scores did not get much improved during the semester (i.e., small latent slope). In these cases, the missingness within a class actually depends on some random effects. We may call it Latent-Class-Random-Effect-Dependent (LCRED) missingness. According to different situations under consideration, LCRED can be further divided into more specific sub-types: Latent-Class-Intercept-Dependent (LCID) missingness, and Latent-Class-Slope-Dependent (LCSD) missingness. Strictly speaking, LCD is another special case of LCRED when the dependency on random effect is not significant. In addition to random effects, the missingness may also depend on potential outcomes that may be missing. For example, a student who feels he is not doing well in a test may be more likely to give up the test. By considering all these cases, therefore, we build three more non-ignorable missingness models in the framework of EGMMs: LCID, LCSD, and the Latent-Class-Outcome-Dependent (LCOD) missingness. They are illustrated in Fig. 1: (b)–(d).

(1) For LCID, τ_{it} is a function of latent class, covariates, and latent individual initial levels, so we model it as follows.

$$\tau_{Iit} = \Phi(\mathbf{z}_i' \boldsymbol{\gamma}_{zt}^* + I_i \gamma_{It} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}), \tag{6}$$

where I_i is the latent initial levels for individual i, γ_{It} is the coefficient for I_i, and $\boldsymbol{\gamma}_{zt}$ and $\boldsymbol{\gamma}_{xt}$ are the same as in (5). A special case of the LCID is the Latent-Intercept-Dependent (LID) missingness in which τ_{it} does not depend on the latent class.

$$\tau_{Iit} = \Phi(\gamma_{0t} + I_i \gamma_{It} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}). \tag{7}$$

(2) For LCSD, τ_{it} is a function of latent class, covariates, and latent individual slopes of growth, so it can be modeled as

$$\tau_{Sit} = \Phi(\mathbf{z}_i' \boldsymbol{\gamma}_{zt}^* + S_i \gamma_{St} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}), \tag{8}$$

where S_i is the latent slope for individual i, and γ_{St} is the coefficient for S_i. Similarly, a special case is the Latent-Slope-Dependent (LSD) missingness.

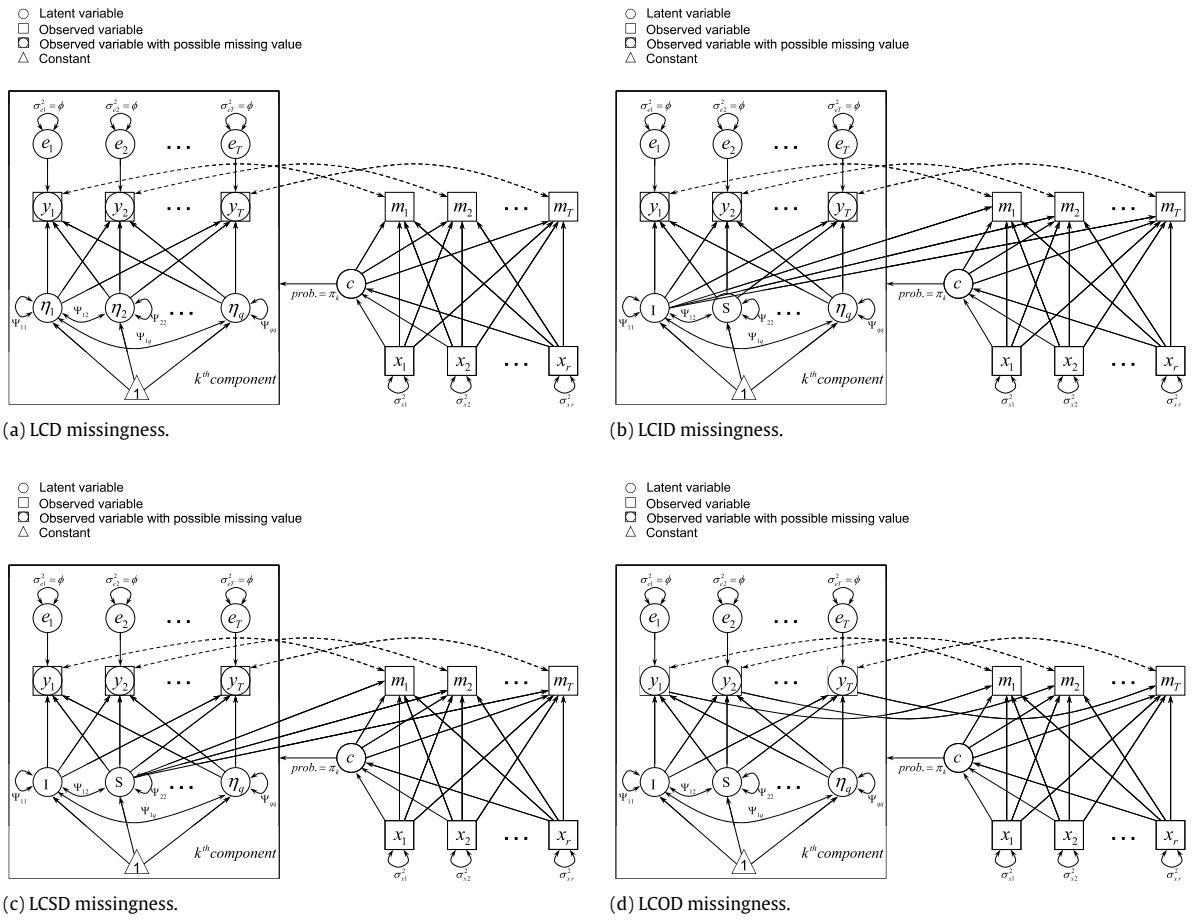
$$\tau_{Sit} = \Phi(\gamma_{0t} + S_i \gamma_{St} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}). \tag{9}$$

(3) For LCOD, τ_{it} is a function of latent class, covariates, and potential outcomes that may be missing. We express τ_{it} as

$$\tau_{Yit} = \Phi(\mathbf{z}_i' \boldsymbol{\gamma}_{zt}^* + y_{it} \gamma_{yt} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}), \tag{10}$$

where y_{it} is the potential outcomes for individual i at time t, and γ_{yt} is the coefficient for y_{it}. And a special case of the LCOD is the Latent-Outcome-Dependent (LOD) missingness.

$$\tau_{Yit} = \Phi(\gamma_{0t} + y_{it} \gamma_{yt} + \mathbf{x}_i' \boldsymbol{\gamma}_{xt}). \tag{11}$$



(a) LCD missingness.

(b) LCID missingness.

(c) LCSD missingness.

(d) LCOD missingness.

Fig. 1. Path diagrams of EGMMs with different missingness.

2.2. Robust GMMs

The effects of outliers can be reduced by using robust components. In this study we adopt the t-distribution, which is a natural replacement of normal distribution especially when data have outliers and heavy-tails (Zhang et al., 2013). As the model in Eqs. (1) and (2) has two levels, we propose three robust models as in Fig. 2.

(1) t-Normal (TN) model in which the measurement errors are t-distributed and the latent random effects are normally distributed,

$$\begin{cases} \mathbf{e}_i \sim Mt_T(\mathbf{0}, \boldsymbol{\Theta}, \nu), \\ \boldsymbol{\xi}_i \sim MN_q(\mathbf{0}, \boldsymbol{\Psi}), \end{cases} \tag{12}$$

where $Mt_T(\mathbf{0}, \boldsymbol{\Theta}, \nu)$ is a T-dimensional multivariate t-distribution with a scale matrix $\boldsymbol{\Theta}$ and degrees of freedom ν, and $MN_q(\mathbf{0}, \boldsymbol{\Psi})$ is a q-dimensional multivariate Normal distribution with a variance $\boldsymbol{\Psi}$.

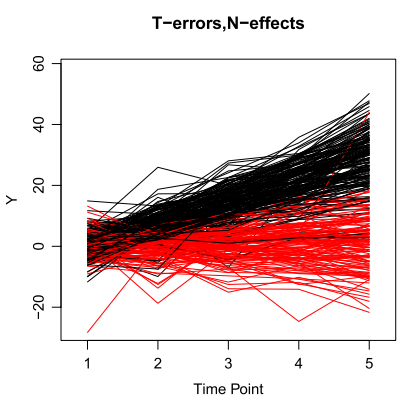
(2) Normal-t (NT) model in which the measurement errors are normally distributed but the latent random effects are t-distributed,

$$\begin{cases} \mathbf{e}_i \sim MN_T(\mathbf{0}, \boldsymbol{\Theta}), \\ \boldsymbol{\xi}_i \sim Mt_q(\mathbf{0}, \boldsymbol{\Psi}, u). \end{cases} \tag{13}$$

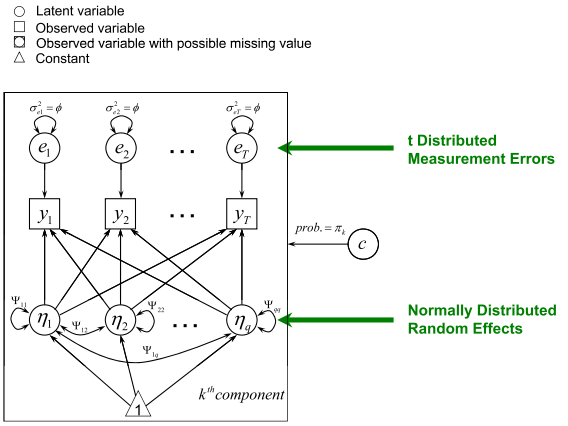
(3) t–t (TT) model in which both the measurement errors and the latent random effects are t-distributed,

$$\begin{cases} \mathbf{e}_i \sim Mt_T(\mathbf{0}, \boldsymbol{\Theta}, \nu), \\ \boldsymbol{\xi}_i \sim Mt_q(\mathbf{0}, \boldsymbol{\Psi}, u). \end{cases} \tag{14}$$

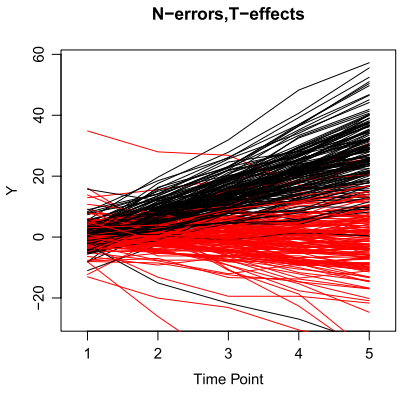
By combining different missingness and distributions, the proposed models are flexible enough to cover a series of GMMs with a variety of missing data mechanisms and contaminated data.



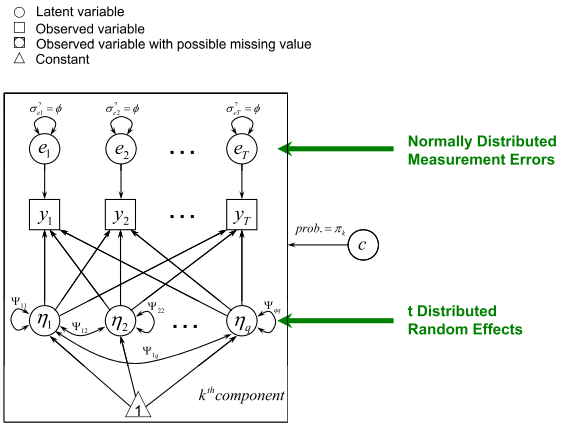
Data plots



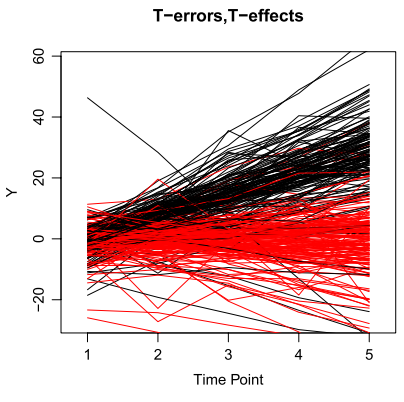
t-Normal (TN) Model: Path Diagram



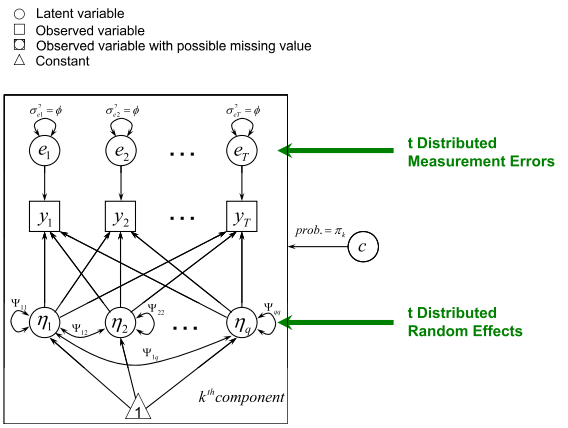
Data plots



Normal-t (NT) Model: Path Diagram



Data plots



t-t (TT) Model: Path Diagram

Fig. 2. Different robust GMMs.

3. Bayesian approach

In this section we describe a full Bayesian approach. First, we utilize the data augmentation method (Tanner and Wong, 1987) to obtain the joint likelihood function of the selection model. The observed data $\mathbf{y}_i^{\text{obs}}$ are augmented with the missing data $\mathbf{y}_i^{\text{mis}}$, the latent random effects $\boldsymbol{\eta}_i$, and the class membership $\mathbf{z}_i$. The detailed likelihood functions are shown in

Table 1
Prior distributions.

Prior distribution	Hyper-parameters
$\phi_k \sim IG(v_{0k}/2, s_{0k}/2)$	Scalars v_{0k} and s_{0k}
$\boldsymbol{\beta}_k \sim MN_q(\boldsymbol{\beta}_{k0}, \boldsymbol{\Sigma}_{k0})$	A $q \times q$ matrix $\boldsymbol{\Sigma}_{k0}$, a q-dimensional vector $\boldsymbol{\beta}_{k0}$
$\boldsymbol{\Psi}_k \sim IW(m_{k0}, \mathbf{V}_{k0})$	A scalar m_{k0}, a $q \times q$ matrix $\mathbf{V}_{k0}$
$\boldsymbol{\varphi}_k \sim MN_{r+1}(\boldsymbol{\mu}_{\varphi_k}, \boldsymbol{\Sigma}_{\varphi_k})$	An $(r+1)$-dimensional vector $\boldsymbol{\mu}_{\varphi k}$, an $(r+1) \times (r+1)$ matrix $\boldsymbol{\Sigma}_{\varphi_k}$
$\nu_k \sim \text{Uniform}(a, b)$	A scalar $a > 2$
$u_k \sim \text{Uniform}(c, d)$	A scalar $c > 2$
$\boldsymbol{\gamma}_t \sim MN_{K+r}(\boldsymbol{\gamma}_{t0}, \mathbf{D}_{t0})$	A $(K+r)$-dimensional vector $\boldsymbol{\gamma}_{t0}$, a $(K+r) \times (K+r)$ matrix $\mathbf{D}_{t0}$

Appendix A. Second, prior distributions are adopted. Table 1 lists the priors and their hyper-parameters. We use uninformative priors so that all the hyper-parameters carry little information. Third, posterior distributions for unknown parameters are calculated. We use conditional posterior distributions instead of the joint posterior because the integrations of marginal posterior distributions of the parameters are hard to obtain explicitly for high-dimensional data. Appendix B lists the detailed posterior distributions. Fourth, with conditional posterior distributions, Markov chains for unknown parameters are generated by implementing a Gibbs sampling algorithm (Geman and Geman, 1984). Fifth, after burn-in periods, convergence tests (e.g., Geweke's z statistics, Geweke, 1992) are conducted to test the convergence of generated Markov chains. Sixth, if the chains pass convergence tests, they are viewed as from the joint distribution, and then statistical inference is conducted. Let θ denote an unknown parameter in the model, and $(\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)})$ be the converged Markov chains. Then across S Markov iterations the posterior estimates $\hat{\theta} = \sum_{s=1}^{S} \theta^{(s)}/S$ with a standard error $s.e.(\hat{\theta}) = \sqrt{\sum_{s=1}^{S}(\theta^{(s)} - \hat{\theta})^2/(S-1)}$. Both percentile intervals and the highest posterior density intervals (HPD, Box and Tiao, 1973) are provided. Seventh, model selection criterion is used to compare competing models and identify the best-fit model. The details are described in Section 4. Finally, the results obtained from the selected model are interpreted.

4. Model selection

In the Bayesian context, there are two versions of deviance, which are the Monte Carlo estimation of the deviance, $\overline{D(\theta)} = E_{\theta|\mathbf{y}}[-2\log(p(\mathbf{y}|\theta))] + C$, and the estimate plugged-in deviance, $D(\hat{\theta}) = -2\log(p(\mathbf{y}|E_{\theta|\mathbf{y}}[\theta])) + C$ for some constant C. The difference between $\overline{D(\theta)}$ and $D(\hat{\theta})$ comes from Jensen's inequality (Casella and George, 1992). When $D(\theta)$ is convex, then $\overline{D(\theta)} \geq D(\hat{\theta})$, and when $D(\theta)$ is concave, then $\overline{D(\theta)} \leq D(\hat{\theta})$. In the detailed framework of Bayesian GMMs with missing data, these two versions can be approximated by

$$\overline{D(\theta)} = -2\left[\frac{1}{S}\sum_{s=1}^{S}\left\{\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}^{(s)} \sum_{1=t}^{T}\left[(1-m_{it})l_{ikt}^{(s)}(y) + l_{ikt}^{(s)}(m)\right]\right\}\right] \quad \text{and}$$

$$D(\hat{\theta}) = -2\left\{\sum_{i=1}^{N}\sum_{k=1}^{K}\hat{z}_{ik}\sum_{t=1}^{T}\left[(1-m_{it})l_{ikt}(y|\hat{\theta}) + l_{ikt}(m|\hat{\theta})\right]\right\},$$

where S is the number of iterations for converged Markov chains, and $z_{ik}^{(s)}$ is the class membership estimated at the sth iteration, $\hat{z}_{ik}$ is the posterior mode of class membership, $l_{ikt}^{(s)}(y)$ and $l_{ikt}^{(s)}(m)$ are conditional loglikelihood functions (see, Celeux et al., 2006) of y_{it} and m_{it}, respectively. In particular, the loglikelihood function for the missing data indicator m_{it} is $l_{ikt}(m) = m_{it}\log(\tau_{it}) + (1-m_{it})\log(1-\tau_{it})$, where τ_{it} is defined by Eqs. (5)–(10).

Based on the two versions of deviance and Lu et al. (submitted for publication)'s article, new model selection criteria are proposed in the framework of Bayesian GMMs with missing data, and their definitions are shown in Table 2.

5. Simulation studies

In this section, three simulation studies are conducted to (1) demonstrate the accuracy of the estimates (including point estimates, standard error estimates, and confidence intervals) of the proposed robust GMMs with missing data using Bayesian methods, and (2) evaluate the performance of the proposed model selection criteria under different situations.

5.1. Simulation design

The detailed design is presented in Table 3. Three simulation studies are designed such that the model complexity increases from study 1 to study 3. Within each study, complete data are generated first and then missing data are created on each occasion. Simulation factors include sample size, missingness pattern, measurement errors distribution, random effect distribution, and class separation (or distance, Anderson and Bahadur, 1962). Study 1 focuses on robust GMMs in which class probabilities are fixed and the non-ignorable missingness do not depend on latent class membership. The true

Table 2
Model selection criteria.

Criterion (index) =	Deviance +	Penalty
Dbar.AIC	$\overline{D(\theta)}$	$2p$
Dbar.BIC	$\overline{D(\theta)}$	$\log(N)\,p$
Dbar.CAIC	$\overline{D(\theta)}$	$(\log(N)+1)\,p$
Dbar.ssBIC	$\overline{D(\theta)}$	$\log((N+2)/24)\,p$
RDIC	$\overline{D(\theta)}$	$\mathrm{var}(D_i)/2$
Dhat.AIC	$D(\hat{\theta})$	$2p$
Dhat.BIC	$D(\hat{\theta})$	$\log(N)\,p$
Dhat.CAIC	$D(\hat{\theta})$	$(\log(N)+1)\,p$
Dhat.ssBIC	$D(\hat{\theta})$	$\log((N+2)/24)\,p$
DIC	$D(\hat{\theta})$	$2pD$

Note:
p is the number of parameters in the model;
N is the sample size;
$pD = \overline{D(\theta)} - D(\hat{\theta})$.

Table 3
Simulation study design.

Sim. study	Model	Data distribution				Missingness depends on					Sample size		Class separation[k]	
		$\mathbf{e}_i$[b]		$\boldsymbol{\eta}_i$[c]										
		N[d]	t[e]	N	t	C[f]	X[g]	I[h]	S[i]	Y[j]	1000	1500	S	M
Study 1	Robust GMMs: Use relative large sample sizes due to multiple classes data, and small class separation due to fixed class probabilities.													
	TN-ignorable		✓	✓			✓				✓	✓		✓
	TN-XI		✓	✓			✓	✓			✓	✓		✓
	TN-XS[a]		✓	✓			✓		✓		✓	✓		✓
	TN-XY		✓	✓			✓			✓	✓	✓		✓
	TT-ignorable		✓		✓		✓				✓	✓		✓
	TT-XI		✓		✓		✓	✓			✓	✓		✓
	TT-XS		✓		✓		✓		✓		✓	✓		✓
	TT-XY		✓		✓		✓			✓	✓	✓		✓
	NT-ignorable	✓			✓		✓				✓	✓		✓
	NT-XI	✓			✓		✓	✓			✓	✓		✓
	NT-XS	✓			✓		✓		✓		✓	✓		✓
	NT-XY	✓			✓		✓			✓	✓	✓		✓
	NN-ignorable	✓		✓			✓				✓	✓		✓
	NN-XI	✓		✓			✓	✓			✓	✓		✓
	NN-XS	✓		✓			✓		✓		✓	✓		✓
	NN-XY	✓		✓			✓			✓	✓	✓		✓
	In total, there are 16 models × 2 levels of sample size = 32 cases.													
Study 2	Robust EGMMs: Select 5 competing models based on the performance in study 1. Use relative large sample sizes due to multiple-class data and varied class probabilities.													
	TN-CXS		✓	✓		✓	✓		✓		✓	✓	✓	✓
	TN-CX		✓	✓		✓	✓				✓	✓	✓	✓
	TT-CXS		✓		✓	✓	✓		✓		✓	✓	✓	✓
	NN-CXS	✓		✓		✓	✓		✓		✓	✓	✓	✓
	NN-CX	✓		✓		✓	✓				✓	✓	✓	✓
	In total, there are 5 models × 2 levels of sample size × 2 levels of class separation = 20 cases.													

(*continued on next page*)

model is the "TN-XS" GMM with t-distributed measurement errors and missingness depending on both covariates and latent slope. In total, there are $16 \times 2 = 32$ conditions. Study 2 is designed for the robust Extended GMMs (EGMMs) in which class probabilities are not fixed and may depend on values of covariates. Also, the missingness may depend on latent class membership. The true model is the "TN-CXS" EGMM with t-distributed measurement errors and missingness depending on observed covariates, the latent slope, and latent class membership. Based on the findings in study 1, 5 competing models (TN-CXS, TT-CXS, NN-CXS, TN-CX, NN-CX) are selected to fit the data. In addition to the factors in study 1, two levels of class

Table 3 (*continued*)

Sim. study	Model	Data distribution				Missingness depends on					Sample size		Class separation[k]	
		$\mathbf{e}_i$[b]		η_i[c]										
		N[d]	t[e]	N	t	C[f]	X[g]	I[h]	S[i]	Y[j]	1000	1500	S	M
Study 3	Number of classes: Select 3 competing models and focus on distributions. Single-class LGCMs vs. multiple-class GMMs.													
	1 class LGCMs													
	TN-XS		✓	✓			✓		✓		✓	✓		
	TT-XS		✓		✓		✓		✓		✓	✓		
	NN-XS	✓		✓			✓		✓		✓	✓		
	2 classes GMMs													
	TN-XS		✓	✓			✓		✓		✓	✓		✓
	TT-XS		✓		✓		✓		✓		✓	✓		✓
	NN-XS	✓		✓			✓		✓		✓	✓		✓
	3 classes GMMs													
	TN-XS		✓	✓			✓		✓		✓	✓		
	TT-XS		✓		✓		✓		✓		✓	✓		
	NN-XS	✓		✓			✓		✓		✓	✓		
	4 classes GMMs													
	TN-XS		✓	✓			✓		✓		✓	✓		
	TT-XS		✓		✓		✓		✓		✓	✓		
	NN-XS	✓		✓			✓		✓		✓	✓		
	In total, there are 3 models × 4 different numbers of classes = 12 cases.													

[a] The shaded model is the true model.
[b] Measurement errors.
[c] Random effects.
[d] Normal distribution.
[e] t distribution.
[f] Latent class dependent (non-ignorable).
[g] Observed covariates.
[h] Latent intercept dependent (non-ignorable).
[i] Latent slope dependent (non-ignorable).
[j] Potential outcome y dependent (Non-ignorable).
[k] Class separation (Anderson and Bahadur, 1962) when generating data (S: small = 1.7, M: medium = 2.7).

separation are considered, which are 2.7 (medium) and 1.7 (small). In total, there are $5 \times 2 \times 2 = 20$ conditions. Study 3 focuses on the number of classes. GMMs with different classes are compared. Based on the performance in the previous two studies, we focus on the robust part with correct missingness. In total, there are 3 models × 4 numbers of classes = 12 conditions.

5.2. Simulation results

In each of the simulation cases, parameter estimates are summarized based on 100 converged replications. We have also tried 1000 replications for a small set of conditions and found no noticeable differences in the results. Each replication generates at least 10,000 iterations for a burn-in period and Markov chains with 50,000 iterations for convergence testing and statistical inference.

5.2.1. Results from study 1

All the 32 summary tables in study 1 (Tables 1–32) are uploaded to the website of http://nd.psychstat.org/research/csda2013. As an example, the summary table of the true mode "TN-XS" when $N = 1500$ is shown in Table 4, from which one can see that, except for the estimates of both degrees of freedom which are inflated due to mis-classification, the true model recovers parameters accurately. For true models, we further summarize all results in Table 5.

Misspecified models perform not as good as the true model (shown in Tables 3–32). With estimates of slope (S) significantly lower than the true value 3, misspecified models may reject the true value with a high chance, and then conclusions will be severely misleading. Also, the misspecified measurement errors distribution leads to low coverages for var(e), var(I), and class proportion. Note that among misspecified models, the model TT-XS performs similarly to the true model TN-XS because the only difference is between a normal distribution and a t distribution with $df \geq 50$. This is because the t distribution approach the normal distribution with the increase of degrees of freedom.

Next, model selection proportions for 10 indices are listed in Table 6. The performances of the criteria in study 1 are ranked, from high to low, as CAIC, BIC, ssBIC, AIC, and DIC.

Table 4
Summary of TN-XS GMM (the true model) with $N = 1500$ and class separation $= 2.7$.

	Par.[a]	True[b]	est.[c]	BIAS		SE		MSE[h]	CI ($\alpha = 0.05$)[i]			HPD ($\alpha = 0.05$)[j]		
				smp.[d]	rel.[e]	emp.[f]	avg.[g]		lower	upper	cover (0.95)	lower	upper	cover (0.95)
Growth curve parameters	Class 1	I	5	4.968	−0.032	−0.006	0.182	0.158	0.06	4.643	5.265	0.93	4.656	5.274	0.93
		S	3	3.007	0.007	0.002	0.119	0.116	0.028	2.78	3.234	0.95	2.78	3.233	0.97
		var(I)	1	1.065	0.065	0.065	0.352	0.292	0.217	0.567	1.708	0.90	0.526	1.641	0.89
		var(S)	4	4.022	0.022	0.006	0.294	0.308	0.182	3.453	4.663	0.97	3.431	4.634	0.97
		cov(IS)	0	0.009	0.009	0.009	0.201	0.195	0.079	−0.381	0.383	0.95	−0.375	0.387	0.95
		var(e)	1	1.06	0.06	0.06	0.107	0.106	0.027	0.863	1.276	0.94	0.857	1.269	0.94
	Class 2	I	1	1.001	0.001	0.001	0.183	0.155	0.058	0.711	1.319	0.91	0.702	1.307	0.91
		S	3	3.004	0.004	0.001	0.113	0.118	0.027	2.773	3.235	0.96	2.773	3.235	0.96
		var(I)	1	0.999	−0.001	−0.001	0.275	0.283	0.158	0.517	1.619	0.95	0.479	1.555	0.94
		var(S)	4	3.965	−0.035	−0.009	0.318	0.316	0.202	3.38	4.62	0.92	3.36	4.593	0.91
		cov(IS)	0	0.026	0.026	0.026	0.211	0.194	0.083	−0.361	0.401	0.95	−0.357	0.404	0.95
		var(e)	1	1.057	0.057	0.057	0.127	0.107	0.031	0.857	1.274	0.85	0.852	1.267	0.85
Probit parameters		CP_1[k]	0.5	0.508	0.008	0.017	0.041	0.04	0.003	0.431	0.587	0.93	0.432	0.586	0.94
		CP_2	0.5	0.492	−0.008	−0.017	0.041	0.04	0.003	0.413	0.569	0.93	0.414	0.568	0.94
	Wave 1	γ_{01}[l]	−1	−1.037	−0.037	0.037	0.148	0.149	0.045	−1.339	−0.753	0.96	−1.33	−0.75	0.96
		γ_{x1}[m]	−1.5	−1.535	−0.035	0.023	0.091	0.103	0.02	−1.748	−1.343	0.96	−1.74	−1.338	0.96
		γ_{S1}[n]	0.5	0.516	0.016	0.033	0.05	0.053	0.006	0.417	0.625	0.97	0.416	0.621	0.98
	Wave 2	γ_{02}	−1	−1.031	−0.031	0.031	0.147	0.146	0.044	−1.327	−0.754	0.93	−1.32	−0.753	0.94
		γ_{x2}	−1.5	−1.53	−0.03	0.02	0.095	0.1	0.02	−1.736	−1.344	0.95	−1.726	−1.337	0.94
		γ_{S2}	0.5	0.513	0.013	0.026	0.053	0.051	0.006	0.417	0.618	0.95	0.415	0.613	0.96
	Wave 3	γ_{03}	−1	−1.008	−0.008	0.008	0.139	0.142	0.04	−1.295	−0.738	0.96	−1.287	−0.734	0.96
		γ_{x3}	−1.5	−1.516	−0.016	0.011	0.098	0.095	0.019	−1.711	−1.338	0.94	−1.703	−1.333	0.94
		γ_{S3}	0.5	0.506	0.006	0.011	0.045	0.049	0.004	0.415	0.605	0.98	0.413	0.601	0.98
	Wave 4	γ_{04}	−1	−1.034	−0.034	0.034	0.155	0.149	0.047	−1.339	−0.753	0.94	−1.327	−0.748	0.91
		γ_{x4}	−1.5	−1.545	−0.045	0.03	0.1	0.098	0.022	−1.747	−1.362	0.94	−1.74	−1.357	0.94
		γ_{S4}	0.5	0.518	0.018	0.036	0.052	0.05	0.006	0.424	0.621	0.93	0.422	0.616	0.91
	df	df_{y1}[o]	5	6.164	1.164	0.233	1.913	1.551	8.308	3.999	9.944	0.93	3.797	9.315	0.96
		df_{y2}	5	6.598	1.598	0.32	2.634	1.717	13.517	4.139	10.656	0.81	3.96	10.09	0.83

Note: The results are summarized based on 100 converged replications with a convergence rate of $100/101 \approx 99\%$.

[a] Parameters. Growth model parameters for both classes are "I": Intercept, "S": Slope, "var(I)": the variance of intercept, "var(S)": the variance of slope, "cov(IS)": the covariance of intercept and slope, and "var(e)": the variance of errors.

[b] The true value of the corresponding parameter.

[c] The parameter estimate, calculated by the Bayesian posterior mean $\text{est.}_j = \bar{\hat{\theta}}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}/100$.

[d] The simple bias, defined by $\text{BIAS.smp}_j = \bar{\hat{\theta}}_j - \theta_j$.

[e] The relative bias, defined by $\text{BIAS.rel}_j = \bar{\hat{\theta}}_j - \theta_j/\theta_j$ if $\theta_j \neq 0$, and $\bar{\hat{\theta}}_j - \theta_j$ if $\theta_j = 0$.

[f] The empirical standard errors, defined by $\text{SE.emp}_j = \sqrt{\sum_{i=1}^{100}(\hat{\theta}_{ij} - \bar{\hat{\theta}}_j)^2/99}$.

[g] The average standard errors, defined by $\text{SE.avg}_j = \sum_{i=1}^{100} \hat{s}_{ij}/100$.

[h] The mean square error, defined by $\text{MSE}_j = \sum_{i=1}^{100} \text{MSE}_{ij}/100$ where MSE_{ij} is the mean square error for the jth parameter in the ith simulation replication, $\text{MSE}_{ij} = (\text{Bias}_{ij})^2 + (\hat{s}_{ij})^2$.

[i] The lower, upper limits, and coverage of percentile confidence interval, defined by $\text{CI.low}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}^l/100$, $\text{CI.upper}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}^u/100$, and $\text{CI.cover}_j = \#(\hat{\theta}_{ij}^l \leq \theta_j \text{ and } \theta_j \leq \hat{\theta}_{ij}^u)/100$.

[j] The lower, upper limits, and coverage of HPD interval.

[k] The fixed class probability for class 1.

[l] The probit intercept of the missing data probability at the 1th wave of data, defined in Eq. (9).

[m] The probit coefficient of the covariate at the 1th wave of data, defined in Eq. (9).

[n] The probit coefficient of the latent slope S at the 1th wave of data, defined in Eq. (9).

[o] The degrees of freedom of the multivariate-t of measurement errors for both classes.

5.2.2. Results from study 2

First, the 20 summary tables in study 2 (Tables 33–52) are shown on http://nd.psychstat.org/research/csda2013. Again, for the true model the results are further summarized in Table 7, from which one can tell that the true model performs well in study 2. With the increase of the sample size or the distance of classes, both the point estimates and standard errors get more accurately estimated, and the statistics across all parameters are improved.

Misspecified models cause biased estimates, part of which is shown in Fig. 3. The true value of the slope is 3, but when models are incorrectly assumed as CX missingness, our classification analysis shows the slopes of more than 50% individuals are incorrectly estimate as less than 1.5 for class separation $= 1.7$ and less than 2 for class separation $= 2.7$. When the missingness is misspecified, the slope coverages are very low, with the largest value 6% for class 2 in Table 50. Also, when the distribution is not correctly modeled, the coverages of var(e) are very low for both classes. For most misspecified models, the convergence rates are low. As in study 1, the model TT-CXS performs almost identically to the true model.

Table 5
Summary results of true model in study 1.

		|Bias.rel|[a]	|SE.diff|[b]	MSE[c]	CI.cover[d]	HPD.cover[e]	CVG.rate[f]
Average across all model parameters, except for *df*							
N	1500	0.020	0.009	0.055	0.940	0.940	99(%)
	1000	0.031	0.016	0.089	0.932	0.931	99(%)

[a] The average simple relative bias across all model parameters, defined by $|\text{Bias.rel}| = \sum_{j=1}^{p} |\text{Bias.rel}_j|/p$. The smaller, the better.

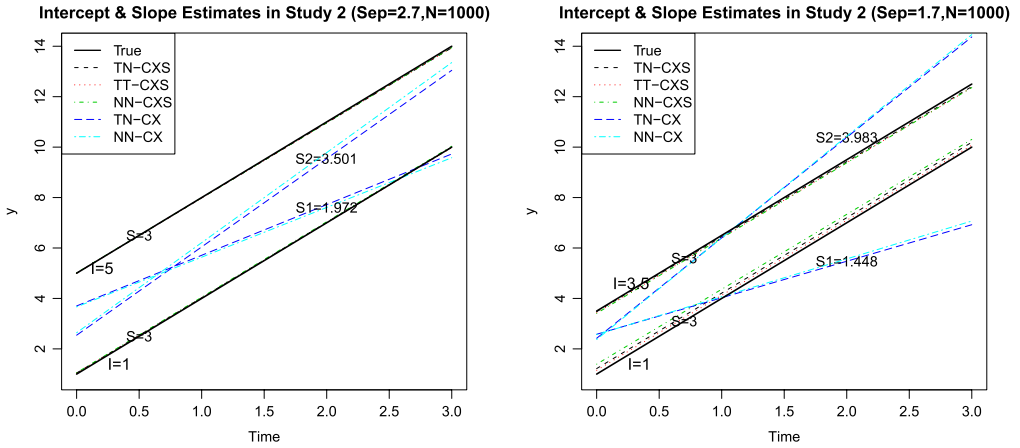
[b] The average absolute difference between the empirical SEs and the average Bayesian SEs across all model parameters, defined by $|\text{SE.diff}| = \sum_{j=1}^{p} |\text{SE.emp}_j - \text{SE.avg}_j|/p$. The smaller, the better.

[c] The Mean Square Errors (MSE) across all model parameters, defined by $\text{MSE} = \sum_{j=1}^{p}[(\text{Bias}_j)^2 + (\hat{s}_j)^2]/p$. The smaller, the better.

[d] The average percentile coverage probability across all model parameters, defined by $\text{CI.cover} = \sum_{j=1}^{p} \text{CI.cover}_j/p$, with a theoretical value of 0.95.

[e] The average highest posterior density (HPD) coverage probability across all model parameters, defined by $\text{HPD.cover} = \sum_{j=1}^{p} \text{HPD.cover}_j/p$, with a theoretical value of 0.95.

[f] The convergence rate.



(a) Class separation = 2.7 (true $\text{Int}_1 = 5$, $\text{Int}_2 = 1$, Slope = 3).

(b) Class separation = 1.7 (true $\text{Int}_1 = 3.5$, $\text{Int}_2 = 1$, Slope = 3).

Fig. 3. Comparison of 5 models in study 2.

Second, the selection proportions for each criterion are listed in Table 8. Most criteria can correctly identify the true model with high certainty.

5.2.3. *Results from study 3*

As all misspecified models with 4 classes do not converge well in study 3, we did not summarize the results. In addition to the cases with 2 classes (Tables 1, 3, and 7), the remaining 6 summary tables (Tables 53–58) are uploaded to the website. For the misspecified models with a single class, the intercept is estimated to be around the average of the true intercepts of two classes, and the variance of intercept is estimated much bigger than the true value. Their convergence rates are high, though, because the number of parameters are less than that of other GMMs with multi-class. For the misspecified GMMs with 3 classes, intercepts are mistakenly estimated to be around 5, 4, and 1. And the convergence rates for are very low, especially for TN-XS or TT-XS the rate is only 4%.

Next, the selection proportions of criteria to pick the best-fit model are listed as in Table 9. According to Table 9, the Dhat group performs better than the Dbar group. Within the Dhat group, the order of correct selection proportions is, from high to low, CAIC, BIC, ssBIC, AIC, and DIC.

5.3. *Simulation conclusions*

The following conclusions can be drawn from the simulation studies. First, the Bayesian method can recover model parameters well as indicated by the small relative biases and the close to 95% average coverage probabilities. Second, with the increase of the sample size or distance between classes, (1) the relative biases get smaller, which means that estimates get closer to their true values, and (2) the average Bayesian SEs get closer to the empirical SEs, which means that standard errors become more accurate. Third, the small difference between the empirical SE and the average Bayesian SE demonstrates the Bayesian method used in the study can estimate the standard errors accurately. Fourth, the estimates of degrees of freedom are inflated due to the mis-classification in estimation. Fifth, incorrectly modeling the distribution will also lead to incorrect conclusions. Sixth, with the correct number of classes, almost all the criteria can correctly identify the true model with

Table 6
Model selection proportion in study 1.

Criterion[a]		$N = 1500$				$N = 1000$			
		Non-ignorable			Ignorable	Non-ignorable			Ignorable
		XS[b]	XY	XI		XS	XY	XI	
Dbar.AIC	TN[b]	0.621[c]	0.000	0.000	0.000	0.593	0.000	0.000	0.000
	TT	0.357	0.000	0.000	0.000	0.314	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000
	NN	0.021	0.000	0.000	0.000	0.071	0.000	0.000	0.000
Dbar.BIC	TN	0.864	0.000	0.000	0.000	0.843	0.000	0.000	0.000
	TT	0.114	0.000	0.000	0.000	0.064	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000
	NN	0.021	0.007	0.000	0.000	0.079	0.000	0.000	0.000
Dbar.CAIC	TN	0.893	0.000	0.000	0.000	0.857	0.000	0.000	0.000
	TT	0.079	0.000	0.000	0.000	0.043	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.007	0.007	0.000	0.000
	NN	0.021	0.007	0.000	0.000	0.086	0.000	0.000	0.000
Dbar.ssBIC	TN	0.729	0.000	0.000	0.000	0.750	0.000	0.000	0.000
	TT	0.250	0.000	0.000	0.000	0.157	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000
	NN	0.021	0.007	0.000	0.000	0.079	0.000	0.000	0.000
RDIC	TN	0.071	0.000	0.000	0.000	0.143	0.000	0.000	0.000
	TT	0.086	0.000	0.000	0.000	0.071	0.000	0.000	0.000
	NT	0.450	0.000	0.000	0.000	0.393	0.007	0.000	0.000
	NN	0.393	0.000	0.000	0.000	0.379	0.007	0.000	0.000
Dhat.AIC	TN	0.586	0.000	0.000	0.000	0.621	0.000	0.000	0.000
	TT	0.379	0.000	0.000	0.000	0.329	0.000	0.000	0.000
	NT	0.014	0.000	0.000	0.000	0.014	0.007	0.000	0.000
	NN	0.014	0.007	0.000	0.000	0.057	0.000	0.000	0.000
Dhat.BIC	TN	0.757	0.000	0.000	0.000	0.793	0.000	0.000	0.000
	TT	0.207	0.000	0.000	0.000	0.121	0.000	0.000	0.000
	NT	0.007	0.000	0.000	0.000	0.007	0.007	0.000	0.000
	NN	0.021	0.007	0.000	0.000	0.071	0.000	0.000	0.000
Dhat.CAIC	TN	0.757	0.000	0.000	0.000	0.814	0.000	0.000	0.000
	TT	0.207	0.000	0.000	0.000	0.100	0.000	0.000	0.000
	NT	0.007	0.000	0.000	0.000	0.007	0.007	0.000	0.000
	NN	0.021	0.007	0.000	0.000	0.071	0.000	0.000	0.000
Dhat.ssBIC	TN	0.586	0.000	0.000	0.000	0.664	0.000	0.000	0.000
	TT	0.379	0.000	0.000	0.000	0.250	0.000	0.000	0.000
	NT	0.014	0.000	0.000	0.000	0.014	0.007	0.000	0.000
	NN	0.014	0.007	0.000	0.000	0.064	0.000	0.000	0.000
DIC	TN	0.507	0.000	0.000	0.000	0.364	0.007	0.000	0.000
	TT	0.371	0.000	0.000	0.000	0.286	0.000	0.000	0.000
	NT	0.043	0.036	0.000	0.000	0.129	0.029	0.007	0.000
	NN	0.043	0.000	0.000	0.000	0.150	0.029	0.000	0.000

[a] The definition of each criterion is given in Table 2.
[b] The shaded model is the true model.
[c] The shaded cell has the largest proportion.

high certainty, with an order of correct selection proportion, CAIC, BIC, ssBIC, AIC, from high to low. When models having different numbers of classes, Dhat.CAIC, Dhat.BIC, Dhat.ssBIC, and Dhat.AIC can be good model selection criteria. Seventh, non-convergent Markov chains might be a sign of a misspecified model. The simulation studies also verify that ignoring the non-ignorable missingness will cause severely misleading conclusions which has been illustrated by previous literature (e.g., Little and Rubin, 2002; Zhang and Wang, 2012).

6. Real data analysis

In this section, we illustrate the application of the Bayesian robust GMMs with missing data through real data analysis. The same sample that has been analyzed in Lu et al. (2011) is used here. It is a mathematical ability growth sample from the NLSY97 survey (Bureau of Labor Statistics, US Department of Labor, 1997), including data collected from $N = 1510$

Table 7
Summary results of TN-CXS EGMM (true model) in study 2.

		|Bias.rel|	|SE.diff|	MSE	CI.cover	HPD.cover	CVG.rate
Average across all parameters, except for *df*							
Class separation = 2.7 (medium)							
N	1500	0.045	0.012	0.069	0.922	0.919	98(%)
	1000	0.043	0.016	0.106	0.931	0.934	96.2(%)
Class separation = 1.7 (small)							
N	1500	0.122	0.055	0.177	0.892	0.894	76(%)
	1000	0.242	0.137	0.504	0.873	0.873	69(%)

Notes: Abbreviations are as given in Table 5.

Table 8
Model selection proportion in study 2.

	TN-CXS	TT-CXS	NN-CXS	TN-CX	NN-CX	TN-CXS	TT-CXS	NN-CXS	TN-CX	NN-CX
	Class separation = 2.7, $N = 1500$					Class separation = 2.7, $N = 1000$				
Dbar.AIC	0.567	0.425	0.000	0.008	0.000	0.558	0.375	0.000	0.067	0.000
Dbar.BIC	0.808	0.158	0.000	0.033	0.000	0.750	0.125	0.000	0.125	0.000
Dbar.CAIC	0.850	0.108	0.000	0.0042	0.000	0.767	0.100	0.008	0.125	0.000
Dbar.ssBIC	0.667	0.300	0.000	0.033	0.000	0.633	0.292	0.000	0.075	0.000
RDIC	0.042	0.042	0.908	0.000	0.008	0.092	0.075	0.808	0.000	0.025
Dhat.AIC	0.475	0.392	0.000	0.133	0.000	0.350	0.358	0.000	0.292	0.000
Dhat.BIC	0.550	0.233	0.000	0.217	0.000	0.450	0.175	0.000	0.375	0.000
Dhat.CAIC	0.525	0.233	0.000	0.242	0.000	0.442	0.150	0.000	0.4	0.008
Dhat.ssBIC	0.467	0.367	0.000	0.167	0.000	0.392	0.300	0.000	0.308	0.000
DIC	0.467	0.500	0.033	0.000	0.000	0.417	0.450	0.108	0.008	0.017
	Class separation = 1.7, $N = 1500$					Class separation = 1.7, $N = 1000$				
Dbar.AIC	0.512	0.444	0.044	0.000	0.00	0.550	0.400	0.050	0.000	0.000
Dbar.BIC	0.744	0.212	0.044	0.000	0.00	0.719	0.194	0.081	0.006	0.000
Dbar.CAIC	0.781	0.175	0.044	0.000	0.00	0.750	0.162	0.081	0.006	0.000
Dbar.ssBIC	0.612	0.344	0.044	0.000	0.00	0.638	0.300	0.062	0.000	0.000
RDIC	0.306	0.238	0.350	0.006	0.10	0.244	0.256	0.362	0.000	0.138
Dhat.AIC	0.475	0.475	0.031	0.019	0.00	0.694	0.231	0.012	0.062	0.000
Dhat.BIC	0.712	0.238	0.031	0.019	0.00	0.644	0.294	0.012	0.050	0.000
Dhat.CAIC	0.712	0.238	0.031	0.019	0.00	0.694	0.231	0.012	0.062	0.000
Dhat.ssBIC	0.475	0.475	0.031	0.019	0.00	0.575	0.388	0.012	0.025	0.000
DIC	0.381	0.450	0.169	0.000	0.00	0.344	0.331	0.319	0.000	0.006

Note: Abbreviations are as given in Table 6.

Table 9
Model selection proportion in study 3.

Criterion	2 classes			1 class			3 classes		
	TN-XS	TT-XS	NN-XS	TN-XS	TT-XS	NN-XS	TN-XS	TT-XS	NN-XS
Dbar.AIC	0.000	0.000	0.057	0.393	0.129	0.000	0.021	0.007	0.393
Dbar.BIC	0.000	0.000	0.036	0.821	0.064	0.000	0.000	0.000	0.079
Dbar.CAIC	0.000	0.000	0.036	0.864	0.043	0.000	0.000	0.000	0.057
Dbar.ssBIC	0.000	0.000	0.057	0.593	0.100	0.000	0.000	0.000	0.25
RDIC	0.036	0.014	0.2	0.014	0.014	0.679	0.014	0.014	0.014
Dhat.AIC	0.621	0.343	0.064	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.BIC	0.793	0.136	0.071	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.CAIC	0.814	0.114	0.071	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.ssBIC	0.664	0.264	0.071	0.000	0.000	0.000	0.000	0.000	0.000
DIC	0.000	0.000	0.000	0.164	0.193	0.121	0.000	0.000	0.521

Note: Abbreviations are as given in Table 6.

Table 10
Summary statistics for PIAT math dataset.

	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
Mean	6.799	7.272	7.555	7.877	8.023
S.D.	1.665	1.712	1.695	1.604	1.659
Missing data (count)	83	69	120	115	143
Missing data (%)	5.497	4.570	7.947	7.616	9.470
	Male	$N = 763$		Female	$N = 747$

Note: The summary statistics are different from those reported in Lu et al. (2011), which conducted a power transformation to normalize the sample.

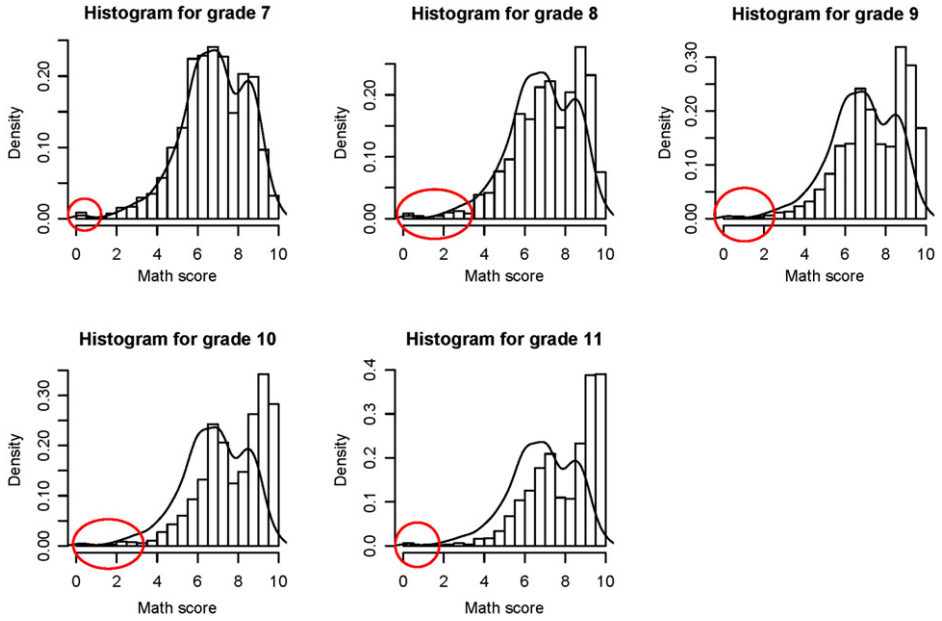


Fig. 4. Histograms of PIAT math scores for 5 grades.

adolescents yearly from 1997 to 2001 when each adolescent was administered the Peabody Individual Achievement Test (PIAT) Mathematics Assessment to measure their mathematical ability. Table 10 shows the summary statistics for the data. Overall, the means of mathematical ability increased over time with a roughly linear trend. The missing data rates range from 4.57% to 9.47%, and the raw data show the missing pattern is intermittent. About half of the sample is female.

Lu et al. (2011) assumed the data are normally distributed without any outliers. But the histograms drawn in Fig. 4 indicate that there are outliers (marked by red circles) at all five grades. So robust methods are used in this study. Also, different non-ignorable missingness mechanisms are considered.

The analysis is conducted following the steps in Table 11. In step 1, a tentative model (the TT-ignorable model) is fitted to the data. Gender is a covariate. The estimates of degrees of freedom of t for both classes are 2.342 and 3.263 for measurement errors and 75.65 and 50.96 for random effects, which indicates that measurement errors are t distributed while random effects are approximately normally distributed (i.e., a TN model). And then in step 2, 8 TN models with different missingness are fitted to the data. During estimation we use uninformative priors which carry little information for model parameters. A burn-in period is run first to make sure all the Markov chains are converged. For testing convergence, the history plot is examined and the Geweke's z statistic (Geweke, 1992) is checked for each parameter. Two selected history plots are presented in Fig. 5. The Geweke's z statistics for all the parameters are smaller than 1.96, which indicates converged Markov chains. To make sure all the parameters are estimated accurately, the next 50,000 iterations are then saved for data analysis. The ratio of Monte Carlo error (MCerror) to standard deviation (S.D.) for each parameter is smaller than or close to 0.05, which indicates parameter estimates are accurate (Spiegelhalter et al., 2003). The results are given in Tables 59–66 on the web site. Step 3, to compare models with different number of classes, two more models are fitted to the data, 3-class NN-X and 4-class NN-X. The results are shown in Tables 67 and 68 on the web site. Step 4, model selection criterion are used to compare the ten models. The indices are listed in Table 12. Suggested by Dhat.CAIC, Dhat.ssBIC, Dhat.BIC, and Dhat.AIC, without further substantive information, the TN-CXY model can be a good candidate of the best-fit model.

Table 13 provides the results of TN-CXY GMM model. It can be interpreted that (1) class 1 has a higher average initial level but a smaller average slope; (2) class 2 has larger variations for initial levels and slope; (3) the residual variance of class 2

Table 11
Steps and fitting models in real data analysis.

Model	$\mathbf{e}_i$		η_i		Missingness				
	N	T	N	T	C	X	I	S	Y
Step 1: Fit a tentative 2 classes model									
TT-ignorable		✓		✓					
Step 2: Fix TN and try different missingness									
2 classes GMMs									
TN-X		✓	✓			✓			
TN-XI		✓	✓			✓	✓		
TN-XS		✓	✓			✓		✓	
TN-XY		✓	✓			✓			✓
2 classes EGMMs									
TN-CX		✓	✓		✓	✓			
TN-CXI		✓	✓		✓	✓	✓		
TN-CXS		✓	✓		✓	✓		✓	
TN-CXY		✓	✓		✓	✓			✓
Step 3: Fit other models with different number of classes									
3 classes GMMs									
NN-X	✓		✓			✓			
4 classes GMMs									
NN-X	✓		✓			✓			
Step 4: Compare selection criteria									
Step 5: Interpret the results of the selected model									

Note: Abbreviations are as given in Table 3.

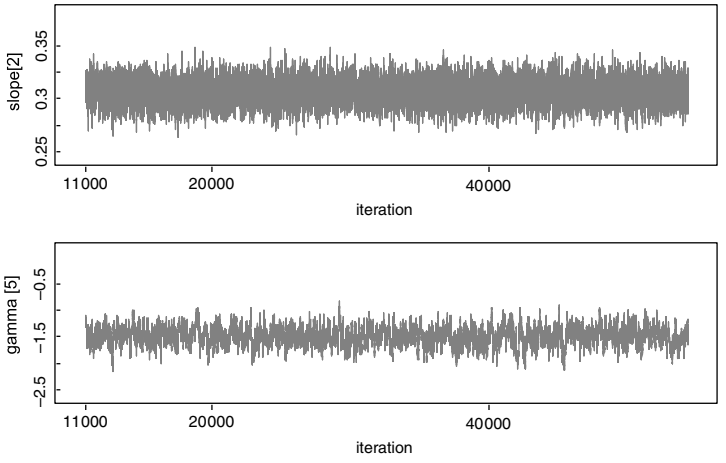


Fig. 5. The history plots for parameters slope[2] and gamma[5].

Table 12
Model selection in real data analysis.

Criterion[a]	2 classes								3 classes	4 classes
	TN-CXS	TN-CXY	TN-CXI	TN-CX	TN-XS	TN-XY	TN-XI	TN-X	NN-X	NN-X
Dbar.AIC	17 392	17 472	17 502	17 502	17 392	17 482	17 502	17 512	17 372	17 126
Dbar.BIC	17583.52	17663.52	17693.52	17666.92	17556.92	17646.92	17666.92	17650.32	17536.92	17328.15
Dbar.CAIC	17619.52	17699.52	17729.52	17697.92	17587.92	17677.92	17697.92	17676.32	17567.92	17366.15
Dbar.ssBIC	17469.15	17549.15	17579.15	17568.44	17458.44	17548.44	17568.44	17567.72	17438.44	17207.44
RDIC	22759.24	22704.5	22378.14	22601.28	22562.65	22755.44	22973.52	22520.18	22843.52	23333.2
Dhat.AIC	15 192	14 942[b]	17 482	19 822	21 922	23 622	25 722	27 352	15 872	15 716
Dhat.BIC	15383.52	15133.52	17673.52	19986.92	22086.92	23786.92	25886.92	27490.32	16036.92	15918.15
Dhat.CAIC	15419.52	15169.52	17709.52	20017.92	22117.92	23817.92	25917.92	27516.32	16067.92	15956.15
Dhat.ssBIC	15269.15	15019.15	17559.15	19888.44	21988.44	23688.44	25788.44	27407.72	15938.44	15797.44
DIC	19 520	19 930	17 450	15 120	12 800	11 280	9220	7620	18 810	18 460

[a] The definition of each criterion is given in Table 2.
[b] The shaded cell has the smallest value.

Table 13
Estimates of TN-CXY GMM in real data analysis.

		Parameter	Mean	S.D.	MCs.e. / S.D. [a]	lower [b]	upper [c]	Geweke t [d]
Growth curve parameters	Class 1	Intercept	8.647	0.037	0.026	8.572	8.717	0.007
		Slope	0.229	0.009	0.023	0.211	0.247	0.014
		Var(I)	0.234	0.028	0.024	0.183	0.293	−0.009
		Var(S)	0.014	0.002	0.018	0.011	0.017	0.004
		Cov(I, S)	−0.036	0.006	0.022	−0.049	−0.026	−0.005
		Var(e)	0.044	0.004	0.031	0.037	0.053	0.024
		df_y [e]	2.386	0.205	0.043	2.118	2.9	0.05
	Class 2	Intercept	6.196	0.047	0.02	6.103	6.287	0.054
		Slope	0.315	0.011	0.022	0.295	0.336	0.036
		Var(I)	1.326	0.084	0.017	1.167	1.497	0.02
		Var(S)	0.034	0.004	0.022	0.027	0.042	0.01
		Cov(I, S)	0.01	0.014	0.021	−0.018	0.037	−0.023
		Var(e)	0.372	0.02	0.033	0.336	0.412	−0.061
		df_y	3.2	0.195	0.04	2.85	3.6	−0.042
Probit parameters	Class	φ_{10} [f]	−0.214	0.119	0.051	−0.438	0.018	−0.039
		φ_{11}	−0.223	0.077	0.051	−0.372	−0.076	0.026
	Grade 7	γ^*_{01} [g]	−0.711	0.532	0.066	−1.843	0.204	−0.255
		γ^*_{11} [h]	−0.132	0.216	0.058	−0.527	0.31	0.231
		γ_{x1} [i]	−0.154	0.108	0.046	−0.368	0.058	0.008
		γ_{Y1} [j]	−0.087	0.059	0.065	−0.19	0.038	0.251
	Grade 8	γ^*_{02}	−1.157	0.446	0.064	−2.097	−0.447	−0.373
		γ^*_{12}	0.046	0.217	0.055	−0.345	0.489	0.347
		γ_{x2}	0.113	0.114	0.046	−0.109	0.334	0.032
		γ_{Y2}	−0.108	0.045	0.062	−0.188	−0.021	0.33
	Grade 9	γ^*_{03}	−0.613	0.454	0.065	−1.519	0.163	−0.462
		γ^*_{13}	−0.057	0.181	0.056	−0.403	0.292	0.381
		γ_{x3}	−0.147	0.094	0.046	−0.332	0.038	0.045
		γ_{Y3}	−0.074	0.045	0.064	−0.155	0.022	0.459
	Grade 10	γ^*_{04}	−0.032	0.512	0.066	−0.861	0.985	−0.426
		γ^*_{14}	−0.324	0.204	0.059	−0.732	0.029	0.362
		γ_{x4}	0.059	0.101	0.047	−0.142	0.251	0.128
		γ_{Y4}	−0.166	0.05	0.065	−0.266	−0.084	0.378
	Grade 11	γ^*_{05}	−1.298	0.421	0.065	−2.13	−0.442	−0.192
		γ^*_{15}	0.341	0.176	0.055	0.015	0.708	0.159
		γ_{x5}	−0.087	0.091	0.045	−0.263	0.083	0.001
		γ_{Y5}	−0.019	0.04	0.064	−0.092	0.062	0.189

[a] Ratio of MC error to standard deviation. A value around or less than 0.05 indicates that the corresponding estimate is accurate (Spiegelhalter et al., 2003).
[b] 2.5 percentile.
[c] 97.5 percentile.
[d] Geweke test t value. An absolute value less than 1.96 indicates that the corresponding chain has passed the convergence test.
[e] The degrees of freedom of the multivariate-t.
[f] The probit coefficient of the class probability for class 1, defined in Eq. (3).
[g] The probit coefficient of the class membership 1 at Grade 7, defined in Eq. (10).
[h] The probit coefficient of the class membership 2 at Grade 7, defined in Eq. (10).
[i] The probit coefficient of the covariate at Grade 7, defined in Eq. (10).
[j] The probit coefficient of the potential output Y at Grade 7, defined in Eq. (10).

is much larger than that of class 1; (4) in class 1 the initial level and the slope are significantly negatively correlated at the confidence level of 95%; (5) the missingness is not related to gender because none of the coefficients of gender are significant at the α level of 0.05; (6) at grade 11 adolescents in class 2 are more likely to miss tests than those in class 1 because the probit coefficient of class membership for grade 11 is significantly positive; and (7) at grades 8 and 10 students with higher potential scores are more likely to miss tests than the students having lower scores because the probit coefficients of the potential outcomes y at the two grades are significantly negative.

7. Conclusion

In this article, the proposed robust growth mixture models with missing data, the Bayesian estimation method, and the model selection criteria were demonstrated using both simulation studies and real data analysis. Simulation studies showed (1) the models can accurately recover parameters using the proposed Bayesian method, and (2) almost all the criteria can correctly identify the true model with high certainty. The real data analysis demonstrated the feasibility of the model, the method, and the criteria in typical longitudinal growth studies such as medical, psychological, educational, and social research.

Acknowledgments

The authors thank Erricos Kontoghiorghes, the CSDA Editor, an anonymous Associate Editor, and two anonymous reviewers for their very helpful comments and suggestions, which greatly improved the quality of the article.

Appendix A. Complete-data likelihood functions

For NT robust GMMs, as in case (13), the likelihood function for the whole sample is

$$L_{(n,t)}(\mathbf{y},\boldsymbol{\eta},\mathbf{m},\mathbf{z}) \propto \prod_{i=1}^{N}\prod_{k=1}^{K}\left\{\pi_{ik}(\mathbf{x}_i)\times|\phi_k|^{-T/2}\exp\left[-\frac{1}{2\phi_k}(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]\frac{\Gamma\left(\frac{u_k+q_k}{2}\right)}{(u_k\pi)^{\frac{q_k}{2}}\Gamma\left(\frac{u_k}{2}\right)}|\boldsymbol{\Psi}_k|^{-\frac{1}{2}}\right.$$
$$\left.\times\left[1+\frac{1}{u_k}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}}\prod_{t=1}^{T}\left[\tau_{ikt}^{m_{it}}(1-\tau_{ikt})^{1-m_{it}}\right]\right\}^{z_{ik}} \triangleq \prod_{i=1}^{N}\prod_{k=1}^{K}(v_{ik})^{z_{ik}}, \tag{A.1}$$

where "$\triangleq$" means "is defined as", $\pi_{ik}(\mathbf{x}_i)$ is defined by Eq. (3), τ_{ikt} is defined by Eq. (5) for the LCD missingness, by Eq. (6) for the LCID missingness, by Eq. (8) for the LCSD missingness, and by Eq. (10) for the LCOD missingness.

For TN robust GMMs, as in case (12), the likelihood function for the whole sample is

$$L_{(t,n)}(\mathbf{y},\boldsymbol{\eta},\mathbf{m},\mathbf{z}) \propto \prod_{i=1}^{N}\prod_{k=1}^{K}\left\{\pi_{ik}(\mathbf{x}_i)\times|\boldsymbol{\Psi}_k|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)\right]\right.$$
$$\times\frac{\Gamma\left(\frac{v_k+T}{2}\right)}{(v_k\pi)^{\frac{T}{2}}\Gamma\left(\frac{v_k}{2}\right)}|\phi_k|^{-\frac{T}{2}}\left[1+\frac{1}{v_k\phi_k}(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{v_k+T}{2}}$$
$$\left.\times\prod_{t=1}^{T}\left[\tau_{ikt}^{m_{it}}(1-\tau_{ikt})^{1-m_{it}}\right]\right\}^{z_{ik}} \triangleq \prod_{i=1}^{N}\prod_{k=1}^{K}(v_{ik})^{z_{ik}}, \tag{A.2}$$

with the same notations as in Eq. (A.1).

For TT robust GMMs, as in case (14), the likelihood function for the whole sample is

$$L_{(t,t)}(\mathbf{y},\boldsymbol{\eta},\mathbf{m},\mathbf{z}) \propto \prod_{i=1}^{N}\prod_{k=1}^{K}\left\{\pi_{ik}(\mathbf{x}_i)\times\frac{\Gamma\left(\frac{v_k+T}{2}\right)}{(v_k\pi)^{\frac{T}{2}}\Gamma\left(\frac{v_k}{2}\right)}|\phi_k|^{-\frac{T}{2}}\left[1+\frac{1}{v_k\phi_k}(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{v_k+T}{2}}\right.$$
$$\times\frac{\Gamma\left(\frac{u_k+q_k}{2}\right)}{(u_k\pi)^{\frac{q_k}{2}}\Gamma\left(\frac{u_k}{2}\right)}|\boldsymbol{\Psi}_k|^{-\frac{1}{2}}\left[1+\frac{1}{u_k}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i-\boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}}$$
$$\left.\times\prod_{t=1}^{T}\left[\tau_{ikt}^{m_{it}}(1-\tau_{ikt})^{1-m_{it}}\right]\right\}^{z_{ik}} \triangleq \prod_{i=1}^{N}\prod_{k=1}^{K}(v_{ik})^{z_{ik}}, \tag{A.3}$$

with the same notations as in Eq. (A.1).

Appendix B. Posterior distributions

B.1. For robust GMMs with t-distributed measurement errors

Let $n_k=\sum_{i=1}^{N}z_{ik}$ be the number of individuals who are in the kth class, and notate the set $(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_N)$ as $\boldsymbol{\eta}$. The conditional posterior distribution for ϕ_k $(k=1,2,\ldots,K)$ is

$$p(\phi_k|\boldsymbol{\Lambda}_k,\boldsymbol{\eta},\mathbf{y},\mathbf{z},v_k) \propto \phi_k^{-\frac{v_{0k}+n_kT}{2}-1}\exp\left(-\frac{s_{0k}}{2\phi_k}\right)\prod_{i=1}^{N}\left[1+\frac{1}{v_k\phi_k}(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{v_k+T}{2}z_{ik}}.$$

The conditional posterior distribution for v_k $(k=1,2,\ldots,K)$ is

$$p(v_k|\mathbf{y},\boldsymbol{\eta},\boldsymbol{\Lambda},\mathbf{z},\phi_k) \propto \left[\frac{\Gamma\left(\frac{v_k+T}{2}\right)}{(v_k)^{\frac{T}{2}}\Gamma\left(\frac{v_k}{2}\right)}\right]^{n_k}\prod_{i=1}^{N}\left[1+\frac{1}{v_k\phi_k}(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i-\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{v_k+T}{2}z_{ik}}I_{[a,b]},$$

where $a>2$ and $I_{[a,b]}$ is an indicator function with a value of 1 inside the compact set $[a,b]$ and 0 outside $[a,b]$.

The conditional posterior distribution for $\boldsymbol{\Psi}_k$ ($k = 1, 2, \ldots, K$) is an inverse Wishart distribution,

$$\boldsymbol{\Psi}_k|\boldsymbol{\beta}_k, \boldsymbol{\eta}, \mathbf{z} \sim IW\ (m_{k1}, \mathbf{V}_{k1})\ ,$$

where $m_{k1} = m_{k0} + n_k$ and $\mathbf{V}_{k1} = \mathbf{V}_{k0} + \sum_{i=1}^{N} z_{ik}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'$.

The conditional posterior distribution for $\boldsymbol{\beta}_k$ ($k = 1, 2, \ldots, K$) is a multivariate normal distribution,

$$\boldsymbol{\beta}_k|\boldsymbol{\Psi}_k, \boldsymbol{\eta}, \mathbf{z} \sim MN(\boldsymbol{\beta}_{k1}, \boldsymbol{\Sigma}_{k1}),$$

where $\boldsymbol{\beta}_{k1} = \left(n_k\boldsymbol{\Psi}_k^{-1} + \boldsymbol{\Sigma}_{k0}^{-1}\right)^{-1}\left(\boldsymbol{\Psi}_k^{-1}\sum_{i=1}^{N} z_{ik}\boldsymbol{\eta}_i + \boldsymbol{\Sigma}_{k0}^{-1}\boldsymbol{\beta}_{k0}\right)$ and $\boldsymbol{\Sigma}_{k1} = \left(n_k\boldsymbol{\Psi}_k^{-1} + \boldsymbol{\Sigma}_{k0}^{-1}\right)^{-1}$.

For $\boldsymbol{\varphi}_k$, when $k = 1$ the conditional posterior distribution for $\boldsymbol{\varphi}_1$ is

$$p(\boldsymbol{\varphi}_1|\boldsymbol{\varphi}_2, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi 1}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1})'\boldsymbol{\Sigma}_{\varphi 1}^{-1}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1}) + \sum_{i=1}^{N}\left\{z_{i1}\log[\Phi(X_i'\boldsymbol{\varphi}_1)] + z_{i2}\log[\Phi(X_i'\boldsymbol{\varphi}_2) - \Phi(X_i'\boldsymbol{\varphi}_1)]\right\}\right];$$

when $2 \le k \le K - 2$, the conditional posterior distribution of $\boldsymbol{\varphi}_k$ is

$$p(\boldsymbol{\varphi}_k|\boldsymbol{\varphi}_{k-1}, \boldsymbol{\varphi}_{k+1}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi k}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k})'\boldsymbol{\Sigma}_{\varphi k}^{-1}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k}) + \sum_{i=1}^{N}\left\{z_{ik}\log[\Phi(X_i'\boldsymbol{\varphi}_k) - \Phi(X_i'\boldsymbol{\varphi}_{k-1})] + z_{i,k+1}\log[\Phi(X_i'\boldsymbol{\varphi}_{k+1}) - \Phi(X_i'\boldsymbol{\varphi}_k)]\right\}\right];$$

and when $k = K - 1$, the conditional posterior distribution of $\boldsymbol{\varphi}_{K-1}$ is

$$p(\boldsymbol{\varphi}_{K-1}|\boldsymbol{\varphi}_{K-2}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi_{K-1}}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}})'\boldsymbol{\Sigma}_{\varphi_{K-1}}^{-1}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}}) + \sum_{i=1}^{N}\left\{z_{i,K-1}\log[\Phi(X_i'\boldsymbol{\varphi}_{K-1}) - \Phi(X_i'\boldsymbol{\varphi}_{K-2})] + z_{iK}\log[1 - \Phi(X_i'\boldsymbol{\varphi}_{K-1})]\right\}\right].$$

The conditional posterior distribution for $\boldsymbol{\gamma}_t$ ($t = 1, 2, \ldots, T$) is

$$p(\boldsymbol{\gamma}_t|\boldsymbol{\omega}, \mathbf{x}, \mathbf{z}, \mathbf{m}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0})'\mathbf{D}_{t0}^{-1}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0}) + \sum_{i=1}^{N}\left\{m_{it}\log\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t) + (1 - m_{it})\log[1 - \Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)]\right\}\right],$$

where $\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)$ is defined by Eq. (5).

The conditional posterior distribution for $\mathbf{z}_i$ ($i = 1, 2, \ldots, N$) is a multinomial distribution,

$$\mathbf{z}_i|\phi, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}, \mathbf{y}, \boldsymbol{\varphi}, \mathbf{x}, \mathbf{m}, \boldsymbol{\eta} \sim M\ \text{nomial}(1, \pi_{i1}^*, \pi_{i2}^*, \ldots, \pi_{iK}^*),$$

where $\pi_{ik}^* = v_{ik}/\sum_{i=1}^{K} v_{ik}$ with v_{ik} defined in Eq. (A.2).

The conditional posterior distribution for $\boldsymbol{\eta}_i$ ($i = 1, 2, \ldots, N$) is a multivariate normal distribution,

$$\boldsymbol{\eta}_i|\boldsymbol{\phi}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}_i, \mathbf{y}_i \sim MN(\boldsymbol{\mu}_{\eta i}, \boldsymbol{\Sigma}_{\eta i}),$$

where $\boldsymbol{\mu}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left[\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\mathbf{y}_i + \boldsymbol{\Psi}_k^{-1}\boldsymbol{\beta}_k\right)\right]$ and $\boldsymbol{\Sigma}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}$.

The conditional posterior distribution for the missing data $\mathbf{y}_i^{\text{mis}}$ ($i = 1, 2, \ldots, N$) is a normal distribution,

$$\mathbf{y}_i^{\text{mis}}|\mathbf{z}_i, \boldsymbol{\eta}_i, \boldsymbol{\phi} \sim MN\left[\sum_{k=1}^{K} z_{ik}(\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i), \sum_{k=1}^{K} z_{ik}(\mathbf{I}_T\phi_k)\right],$$

and its dimension and location depend on the corresponding $\mathbf{m}_i$ value.

B.2. For robust GMMs with t-distributed random effects

The conditional posterior distribution for ϕ_k $(k = 1, 2, \ldots, K)$ is an inverse gamma distribution,

$$\phi_k|\boldsymbol{\Lambda}_k, \boldsymbol{\eta}, \mathbf{y}, \mathbf{z} \sim IG\,(a_{k1}/2, b_{k1}/2)\,,$$

where $a_{k1} = v_{0k} + n_k T$ and $b_{k1} = s_{0k} + \sum_{i=1}^{N} z_{ik}(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)$.

The conditional posterior distribution for $\boldsymbol{\Psi}_k$ $(k = 1, 2, \ldots, K)$ is an inverse Wishart distribution,

$$\boldsymbol{\Psi}_k|\boldsymbol{\beta}_k, \boldsymbol{\eta}, \mathbf{z} \propto |\boldsymbol{\Psi}_k|^{-\frac{n_k}{2}} \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}}.$$

The conditional posterior distribution for $\boldsymbol{\beta}_k$ $(k = 1, 2, \ldots, K)$ is a multivariate normal distribution,

$$\boldsymbol{\beta}_k|\boldsymbol{\Psi}_k, \boldsymbol{\eta}, \mathbf{z} \propto \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}}.$$

The conditional posterior distribution for u_k $(k = 1, 2, \ldots, K)$ is

$$p(u_k|\boldsymbol{\eta}, \boldsymbol{\Psi}_k, \boldsymbol{\beta}_k, \mathbf{z}) \propto \left[\frac{\Gamma\left(\frac{u_k+q_k}{2}\right)}{(u_k\pi)^{\frac{q_k}{2}}\Gamma\left(\frac{u_k}{2}\right)}\right]^{n_k} \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}} I_{[c,d]},$$

where $c > 2$ and $I_{[c,d]}$ is an indicator function with a value of 1 inside the compact set $[c, d]$ and 0 outside $[c, d]$.

For $\boldsymbol{\varphi}_k$, when $k = 1$ the conditional posterior distribution for $\boldsymbol{\varphi}_1$ is

$$p(\boldsymbol{\varphi}_1|\boldsymbol{\varphi}_2, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi 1}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1})'\boldsymbol{\Sigma}_{\varphi 1}^{-1}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1}) + \sum_{i=1}^{N}\left\{z_{i1}\log[\Phi(X_i'\boldsymbol{\varphi}_1)] + z_{i2}\log[\Phi(X_i'\boldsymbol{\varphi}_2) - \Phi(X_i'\boldsymbol{\varphi}_1)]\right\}\right];$$

when $2 \le k \le K - 2$, the conditional posterior distribution of $\boldsymbol{\varphi}_k$ is

$$p(\boldsymbol{\varphi}_k|\boldsymbol{\varphi}_{k-1}, \boldsymbol{\varphi}_{k+1}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi k}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k})'\boldsymbol{\Sigma}_{\varphi k}^{-1}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k}) + \sum_{i=1}^{N}\left\{z_{ik}\log[\Phi(X_i'\boldsymbol{\varphi}_k) - \Phi(X_i'\boldsymbol{\varphi}_{k-1})] + z_{i,k+1}\log[\Phi(X_i'\boldsymbol{\varphi}_{k+1}) - \Phi(X_i'\boldsymbol{\varphi}_k)]\right\}\right];$$

and when $k = K - 1$, the conditional posterior distribution of $\boldsymbol{\varphi}_{K-1}$ is

$$p(\boldsymbol{\varphi}_{K-1}|\boldsymbol{\varphi}_{K-2}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi_{K-1}}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}})'\boldsymbol{\Sigma}_{\varphi_{K-1}}^{-1}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}}) + \sum_{i=1}^{N}\left\{z_{i,K-1}\log[\Phi(X_i'\boldsymbol{\varphi}_{K-1}) - \Phi(X_i'\boldsymbol{\varphi}_{K-2})] + z_{iK}\log[1 - \Phi(X_i'\boldsymbol{\varphi}_{K-1})]\right\}\right].$$

The conditional posterior distribution for $\boldsymbol{\gamma}_t$ $(t = 1, 2, \ldots, T)$ is

$$p(\boldsymbol{\gamma}_t|\boldsymbol{\omega}, \mathbf{x}, \mathbf{z}, \mathbf{m}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0})'\mathbf{D}_{t0}^{-1}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0}) + \sum_{i=1}^{N}\left\{m_{it}\log\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t) + (1 - m_{it})\log[1 - \Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)]\right\}\right],$$

where $\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)$ is defined by Eq. (5).

The conditional posterior distribution for $\mathbf{z}_i$ $(i = 1, 2, \ldots, N)$ is a multinomial distribution,

$$\mathbf{z}_i|\phi, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}, \mathbf{y}, \boldsymbol{\varphi}, \mathbf{x}, \mathbf{m}, \boldsymbol{\eta} \sim M\,\text{nomial}(1, \pi_{i1}^*, \pi_{i2}^*, \ldots, \pi_{iK}^*),$$

where $\pi_{ik}^* = v_{ik}/\sum_{i=1}^{K} v_{ik}$ with v_{ik} defined in Eq. (A.1).

The conditional posterior distribution for $\boldsymbol{\eta}_i$ $(i = 1, 2, \ldots, N)$ is a multivariate normal distribution,

$$\boldsymbol{\eta}_i|\boldsymbol{\phi}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}_i, \mathbf{y}_i \sim MN(\boldsymbol{\mu}_{\eta i}, \boldsymbol{\Sigma}_{\eta i}),$$

where $\boldsymbol{\mu}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left[\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\mathbf{y}_i + \boldsymbol{\Psi}_k^{-1}\boldsymbol{\beta}_k\right)\right]$ and $\boldsymbol{\Sigma}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}$.

The conditional posterior distribution for the missing data $\mathbf{y}_i^{\text{mis}}$ $(i = 1, 2, \ldots, N)$ is a normal distribution,

$$\mathbf{y}_i^{\text{mis}}|\mathbf{z}_i, \boldsymbol{\eta}_i, \boldsymbol{\phi} \sim MN\left[\sum_{k=1}^{K} z_{ik}(\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i), \sum_{k=1}^{K} z_{ik}(\mathbf{I}_T\phi_k)\right],$$

and its dimension and location depend on the corresponding $\mathbf{m}_i$ value.

B.3. For robust GMMs with t-distributed measurement errors and random effects

Let $n_k = \sum_{i=1}^{N} z_{ik}$ be the number of individuals who are in the kth class, and notate the set $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_N)$ as $\boldsymbol{\eta}$. The conditional posterior distribution for ϕ_k $(k = 1, 2, \ldots, K)$ is

$$p(\phi_k|\boldsymbol{\Lambda}_k, \boldsymbol{\eta}, \mathbf{y}, \mathbf{z}, \nu_k) \propto \phi_k^{-\frac{\nu_{0k}+n_kT}{2}-1} \exp\left(-\frac{s_{0k}}{2\phi_k}\right) \prod_{i=1}^{N}\left[1 + \frac{1}{\nu_k\phi_k}(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{\nu_k+T}{2}z_{ik}}.$$

The conditional posterior distribution for ν_k $(k = 1, 2, \ldots, K)$ is

$$p(\nu_k|\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\Lambda}, \mathbf{z}, \phi_k) \propto \left[\frac{\Gamma\left(\frac{\nu_k+T}{2}\right)}{(\nu_k)^{\frac{T}{2}}\Gamma\left(\frac{\nu_k}{2}\right)}\right]^{n_k} \prod_{i=1}^{N}\left[1 + \frac{1}{\nu_k\phi_k}(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}_k\boldsymbol{\eta}_i)\right]^{-\frac{\nu_k+T}{2}z_{ik}} I_{[a,b]},$$

where $a > 2$ and $I_{[a,b]}$ is an indicator function with a value of 1 inside the set $[a, b]$ and 0 outside $[a, b]$.

The conditional posterior distribution for $\boldsymbol{\Psi}_k$ $(k = 1, 2, \ldots, K)$ is an inverse Wishart distribution,

$$\boldsymbol{\Psi}_k|\boldsymbol{\beta}_k, \boldsymbol{\eta}, \mathbf{z} \propto |\boldsymbol{\Psi}_k|^{-\frac{n_k}{2}} \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}}.$$

The conditional posterior distribution for $\boldsymbol{\beta}_k$ $(k = 1, 2, \ldots, K)$ is a multivariate normal distribution,

$$\boldsymbol{\beta}_k|\boldsymbol{\Psi}_k, \boldsymbol{\eta}, \mathbf{z} \propto \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}}.$$

The conditional posterior distribution for u_k $(k = 1, 2, \ldots, K)$ is

$$p(u_k|\boldsymbol{\eta}, \boldsymbol{\Psi}_k, \boldsymbol{\beta}_k, \mathbf{z}) \propto \left[\frac{\Gamma\left(\frac{u_k+q_k}{2}\right)}{(u_k\pi)^{\frac{q_k}{2}}\Gamma\left(\frac{u_k}{2}\right)}\right]^{n_k} \prod_{i=1}^{N}\left[1 + \frac{1}{u_k}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)'\boldsymbol{\Psi}_k^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\beta}_k)\right]^{-\frac{u_k+q_k}{2}z_{ik}} I_{[c,d]},$$

where $c > 2$ and $I_{[c,d]}$ is an indicator function with a value of 1 inside the compact set $[c, d]$ and 0 outside $[c, d]$.

For $\boldsymbol{\varphi}_k$, when $k = 1$ the conditional posterior distribution for $\boldsymbol{\varphi}_1$ is

$$p(\boldsymbol{\varphi}_1|\boldsymbol{\varphi}_2, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi 1}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1})'\boldsymbol{\Sigma}_{\varphi 1}^{-1}(\boldsymbol{\varphi}_1 - \boldsymbol{\mu}_{\varphi_1}) + \sum_{i=1}^{N}\left\{z_{i1}\log[\Phi(X_i'\boldsymbol{\varphi}_1)] + z_{i2}\log[\Phi(X_i'\boldsymbol{\varphi}_2) - \Phi(X_i'\boldsymbol{\varphi}_1)]\right\}\right];$$

when $2 \le k \le K - 2$, the conditional posterior distribution of $\boldsymbol{\varphi}_k$ is

$$p(\boldsymbol{\varphi}_k|\boldsymbol{\varphi}_{k-1}, \boldsymbol{\varphi}_{k+1}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi k}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k})'\boldsymbol{\Sigma}_{\varphi k}^{-1}(\boldsymbol{\varphi}_k - \boldsymbol{\mu}_{\varphi_k}) + \sum_{i=1}^{N}\left\{z_{ik}\log[\Phi(X_i'\boldsymbol{\varphi}_k) - \Phi(X_i'\boldsymbol{\varphi}_{k-1})] + z_{i,k+1}\log[\Phi(X_i'\boldsymbol{\varphi}_{k+1}) - \Phi(X_i'\boldsymbol{\varphi}_k)]\right\}\right];$$

and when $k = K - 1$, the conditional posterior distribution of $\boldsymbol{\varphi}_{K-1}$ is

$$p(\boldsymbol{\varphi}_{K-1}|\boldsymbol{\varphi}_{K-2}, \mathbf{z}, X) \propto |\boldsymbol{\Sigma}_{\varphi_{K-1}}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}})'\boldsymbol{\Sigma}_{\varphi_{K-1}}^{-1}(\boldsymbol{\varphi}_{K-1} - \boldsymbol{\mu}_{\varphi_{K-1}}) + \sum_{i=1}^{N}\left\{z_{i,K-1}\log[\Phi(X_i'\boldsymbol{\varphi}_{K-1}) - \Phi(X_i'\boldsymbol{\varphi}_{K-2})] + z_{iK}\log[1 - \Phi(X_i'\boldsymbol{\varphi}_{K-1})]\right\}\right].$$

The conditional posterior distribution for $\boldsymbol{\gamma}_t$ $(t = 1, 2, \ldots, T)$ is

$$p(\boldsymbol{\gamma}_t|\boldsymbol{\omega}, \mathbf{x}, \mathbf{z}, \mathbf{m}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0})'\mathbf{D}_{t0}^{-1}(\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_{t0}) + \sum_{i=1}^{N}\left\{m_{it}\log\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t) + (1 - m_{it})\log[1 - \Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)]\right\}\right],$$

where $\Phi(\boldsymbol{\omega}_i'\boldsymbol{\gamma}_t)$ is defined by Eq. (5).

The conditional posterior distribution for $\mathbf{z}_i$ $(i = 1, 2, \ldots, N)$ is a multinomial distribution,

$$\mathbf{z}_i|\phi, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}, \mathbf{y}, \boldsymbol{\varphi}, \mathbf{x}, \mathbf{m}, \boldsymbol{\eta} \sim M\ \text{nomial}(1, \pi_{i1}^*, \pi_{i2}^*, \ldots, \pi_{iK}^*),$$

where $\pi_{ik}^* = v_{ik}/\sum_{i=1}^{K} v_{ik}$ with v_{ik} defined in Eq. (A.2).

The conditional posterior distribution for $\boldsymbol{\eta}_i$ $(i = 1, 2, \ldots, N)$ is a multivariate normal distribution,

$$\boldsymbol{\eta}_i|\boldsymbol{\phi}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \mathbf{z}_i, \mathbf{y}_i \sim MN(\boldsymbol{\mu}_{\eta i}, \boldsymbol{\Sigma}_{\eta i}),$$

where $\boldsymbol{\mu}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left[\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\mathbf{y}_i + \boldsymbol{\Psi}_k^{-1}\boldsymbol{\beta}_k\right)\right]$ and $\boldsymbol{\Sigma}_{\eta i} = \sum_{k=1}^{K} z_{ik}\left(\frac{1}{\phi_k}\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k + \boldsymbol{\Psi}_k^{-1}\right)^{-1}$.

The conditional posterior distribution for the missing data $\mathbf{y}_i^{\text{mis}}$ $(i = 1, 2, \ldots, N)$ is a normal distribution,

$$\mathbf{y}_i^{\text{mis}}|\mathbf{z}_i, \boldsymbol{\eta}_i, \boldsymbol{\phi} \sim MN\left[\sum_{k=1}^{K} z_{ik}(\boldsymbol{\Lambda}_k\boldsymbol{\eta}_i), \sum_{k=1}^{K} z_{ik}(\mathbf{I}_T\phi_k)\right],$$

and its dimension and location depend on the corresponding $\mathbf{m}_i$ value.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1919 (6), 716–723.

Anderson, T.W., Bahadur, R.R., 1962. Classification into two multivariate normal distributions with different covariance matrices. The Annals of Mathematical Statistics 33, 420–431.

Bohning, D., Seidel, W., 2003. Editorial: recent developments in mixture models. Computational Statistics and Data Analysis 41, 349–357.

Bohning, D., Seidel, W., Alfo, M., Garel, B., Patilea, V., Walther, G., 2007. Advances in mixture models. Computational Statistics and Data Analysis 51 (11), 5205–5210.

Bollen, K., Curran, P., 2006. Latent Curve Models: A Structural Equation Perspective. Wiley, New Jersey.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. John Wiley & Sons, Hoboken, NJ.

Bozdogan, H., 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika 52, 345–370.

Bureau of Labor Statistics, US Department of Labor, 1997. National Longitudinal Survey of Youth 1997 cohort, 1997–2003 (rounds 1–7). (Computer file). Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2005. URL http://www.bls.gov/nls/nlsy97.htm.

Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. The American Statistician 46, 167–174.

Celeux, G., Forbes, F., Robert, C., Titterington, D., 2006. Deviance information criteria for missing data models. Bayesian Analysis 4, 651–674.

Collins, L.M., 1991. Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions. American Psychological Association.

Dunson, D.B., 2000. Bayesian latent variable models for clustered mixed outcomes. Journal of the Royal Statistical Society B 62, 355–366.

Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (Eds.), 2008. Longitudinal Data Analysis. Chapman & Hall/CRC Press, Boca Raton, FL.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2004. Applied Longitudinal Analysis. In: Wiley Series in Probability and Statistics, John Wiley & Sons Inc., Hoboken, New Jersey.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernado, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 4. Clarendon Press, Oxford, UK, pp. 169–193.

Glynn, R.J., Laird, N.M., Rubin, D.B., 1986. Mixture Modeling Versus Selection Modeling With Nonignorable Nonresponse. In: Drawing Inferences from Self-Selected Samples. Springer Verlag, New York, pp. 115–142 (Chapter).

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons, New York.

Jelicic, H., Phelps, E., Lerner, R.M., 2009. Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. Developmental Psychology 45, 1195–1199.

Lange, K.L., Little, R.J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84, 881–896.

Lin, T., Lee, J., Ni, H., 2004. Bayesian analysis of mixture modelling using the multivariate t distribution. Statistics and Computing 14, 119–130.

Little, R.J.A., 1993. Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association 88, 125–134.

Little, R.J.A., 1995. Modelling the drop-out mechanism in repeated-measures studies. Journal of the American Statistical Association 90, 1112–1121.

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, second ed., Wiley-Interscience, New York, NY.

Lu, Z., Zhang, Z., Cohen, A., 2013. Bayesian inference for latent growth curve models with non-ignorable missing data. Multivariate Behavioral Research (submitted for publication).

Lu, Z., Zhang, Z., Lubke, G., 2011. Bayesian inference for growth mixture models with latent-class-dependent missing data. Multivariate Behavioral Research 46, 567–597.

Lubke, G.H., Muthén, B., 2005. Investigating population heterogeneity with factor mixture models. Psychological Methods 10, 21–39.

McArdle, J.J., Bell, R.Q., 1999. An introduction to latent growth models for developmental data analysis. In: Little, T.D., Schnabel, K.U., Baumert, J. (Eds.), Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples. Lawrence Erlbaum and Associates, Mahwah, NJ, pp. 69–107.

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons, New York, NY.

Meredith, W., Tisak, J., 1990. Latent curve analysis. Psychometrika 55, 107–122.

Micceri, T., 1989. The unicorn, the normal curve and the other improbable creatures. Psychological Bulletin 105, 156–166.

Muthén, B., 2004. Handbook of Quantitative Methodology for the Social Sciences. In: Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data, Sage Publications, Newbury Park, CA, pp. 345–368 (Chapter).

Muthén, B., Asparouhov, T., Hunter, A.M., Leuchter, A.F., 2011. Growth modeling with nonignorable dropout: alternative analyses of the STAR*D antidepressant trial. Psychological Methods 16, 17–33.

Muthén, B., Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55, 463–469.
Pan, J.X., Fang, K.T., 2002. Growth Curve Models and Statistical Diagnostics. Springer, New York.
Poon, W.Y., Poon, Y.S., 2002. Influential observations in the estimation of mean vector and covariance matrix. British Journal of Mathematical and Statistical Psychology 55, 177–192.
Roth, P.L., 1994. Missing data: a conceptual review for applied psychologists. Personnel Psychology 47, 537–560.
Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall/CRC, Boca Raton, FL.
Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychological Methods 7, 147–177.
Schwarz, G.E., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.
Sclove, L.S., 1987. Application of mode-selection criteria to some problems in multivariate analysis. Psychometrika 52, 333–343.
Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, Inc., New York, NY.
Song, W., Yao, W., Xing, Y., 2014. Robust mixture regression model fitting by Laplace distribution. Computational Statistics and Data Analysis 71, 128–137. (Available online 4 July 2013, Corrected Proof).
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.v.d., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 583–639.
Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS Manual Version 1.4 MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK. http://www.mrc-bsu.cam.ac.uk/bugs.
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.
Yuan, K.H., Bentler, P.M., 1998. Structural equation modeling with robust covariances. Sociological Methodology 28, 363–396.
Yuan, K.H., Lu, Z., 2008. SEM with missing data and unknown population using two-stage ML: theory and its application. Multivariate Behavioral Research 43, 621–652.
Yuan, K.H., Zhang, Z., 2012. Robust structural equation modeling with missing data and auxiliary variables. Psychometrika 77 (4), 803–826.
Zhang, Z., Lai, K., Lu, Z., Tong, X., 2013. Bayesian inference and application of robust growth curve models using student's t distribution. Structural Equation Modeling 20 (1), 47–78.
Zhang, Z., Wang, L., 2012. A note on the robustness of a full Bayesian method for non-ignorable missing data analysis. Brazilian Journal of Probability and Statistics 26 (3), 244–264.