Chapter 21
Model Selection Criteria for Latent Growth Models Using Bayesian Methods

Zhenqiu (Laura) Lu, Zhiyong Zhang, and Allan Cohen

Abstract Research in applied areas, such as statistical, psychological, behavioral, and educational areas, often involves the selection of the best available model from among a large set of candidate models. Considering that there is no well-defined model selection criterion in a Bayesian context and that latent growth mixture models are becoming popular in many areas, the goal of this study is to investigate the performance of a series of model selection criteria in the framework of latent growth mixture models with missing data and outliers in a Bayesian context. This study conducted five simulation studies to cover different cases, including latent growth curve models with missing data, latent growth curve models with missing data and outliers, growth mixture models with missing data and outliers, extended growth mixture models with missing data and outliers, and latent growth models with different classes. Simulation results show that almost all the proposed criteria can effectively identify the true models. This study also illustrated the application of these model selection criteria in real data analysis. The results will help inform the selection of growth models by researchers seeking to provide states with accurate estimates of the growth of their students.

21.1 Introduction

Traditional criteria are available for researchers to select the best-fit model from among a large set of candidate models. Akaike (1974) proposed the Akaike's information criterion (AIC), which offers a relative measure of the information lost. For Bayesian models the Bayes factor, which is the ratio of posterior odds to prior odds, can work for both hypothesis testing and model comparison. But the Bayes factor is often difficult or impossible to calculate, especially for models that involve random effects, large numbers of unknowns or improper priors. To approximate

Z. Lu (✉) • A. Cohen
University of Georgia, Athens, GA 30602, USA
e-mail: zlu@uga.edu; acohen@uga.edu

Z. Zhang
University of Notre Dame, Notre Dame, IN 46556, USA
e-mail: zhangzhiyong@nd.edu

R.E. Millsap et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics 89, DOI 10.1007/978-3-319-07503-7_21

the Bayes factor, Schwarz (1978) developed the Bayesian information criterion (BIC, sometimes called the Schwarz criterion). To obtain more precise criteria, Bozdogan (1987) proposed the consistent Akaike information criterion (CAIC), and Sclove (1987) proposed the sample-size adjusted Bayesian information criterion (ssBIC). The deviance information criterion (DIC, Spiegelhalter et al. 2002) is a recently developed criterion designed for hierarchical models. It is based on the posterior distribution of the log-likelihood and is useful in Bayesian model selection problems where the posterior distributions have been obtained by Markov chain Monte Carlo (MCMC) simulation. DIC is usually regarded as a generalization of AIC and BIC. It is defined analogously to AIC or BIC with a penalty term of the number equal to effective model parameters in Bayesian models. In practice, rough DIC (RDIC or DICV in some literature, e.g., Oldmeadow and Keith 2011) is an approximation of DIC. The mathematical forms of AIC, BIC, CAIC, ssBIC, and DIC are closely related to each other. They all try to find a balance between accuracy and complexity of the fitting model. The accuracy of a model can be shown by a deviance $D(\theta) = -2log(f(y|\theta)) + C$ for some constant C where θ is a vector of model parameters. For all the criteria above, the model with a smaller criterion value is better supported by data.

Bayesian approach is becoming increasingly important in estimating models as it provides many advantages in dealing with complex statistical models with complicated data structure (e.g., Dunson 2000). However, there is no well-defined model selection criterion in a Bayesian context (e.g., Celeux et al. 2006). There are at least three problems. First, in a Bayesian context there are two versions of deviance because the Bayesian procedure generates Monte Carlo Markov chains for each parameter. One version is the posterior estimate which can be expressed as $D(\hat{\theta}) = -2log(p(y|E_{\theta|y}[\theta])) + C$, which is analogous to a frequentist estimate. It can be estimated by adopting a point parameter estimate of θ. Another version is the Monte Carlo estimate of the expected deviance, which can be calculated as $\overline{D(\theta)} = E_{\theta|y}[-2log(p(y|\theta))] + C$, which is based on Bayesian iterations. It can be estimated as the posterior mean across a converged Markov chain. Conceptually, $\overline{D(\theta)}$ is the average of all deviances, and $D(\hat{\theta})$ is the deviance of the average of all estimates. The second problem is related to the complexity of the raw data. The data often come from heterogeneous populations which almost unavoidable contain outliers and attrition. The estimates from mis-specified models may result in severely misleading conclusions. The third problem relates to the likelihood function. When latent variables are considered in statistical models, the likelihood function can be an observed-data likelihood function, a complete-data likelihood function, or a conditional likelihood function (Celeux et al. 2006). Furthermore, if data come from heterogeneous populations, the class membership indicator may have different versions, for example, a posterior mode or a posterior mean. Also, with missing data, the likelihood functions have different ways to construct.

To address these problems, new criteria are expected. As latent growth modeling is becoming increasingly popular in applied research, such as in statistical, psychological, behavioral, and educational areas, in this study we consider to use latent growth models to test the performance of proposed model selection criteria.

Specifically, the goal of this paper is to examine the performance of the Bayesian model selection criteria with more general growth models, such as non-normally distributed growth models, robust growth mixture models, and robust extended growth mixture models. Lu et al. (2013b) proposed a series of Bayesian criteria, based on the traditional model selection criteria. However, in Lu et al. (2013b) the performances of these criteria were investigated when data are non-mixture, normally distributed, and with simple non-ignorable missingness. And only latent growth models were used. In this study, data are more complex. We conduct five simulation studies. The results will help inform the selection of growth models by researchers seeking to provide people with accurate estimates of growth across a variety of possible contexts.

21.2 Robust Growth Models with Non-ignorable Missingness

Our investigation of the performance of the Bayesian selection criteria involves fitting growth models to complex data. In this section, different types of growth models are briefly introduced. Given the fact that the data used in growth models are almost inevitably contain attrition (e.g., Little and Rubin 2002; Yuan and Lu 2008; Lu et al. 2011) and outliers (e.g., Maronna et al. 2006), different types of growth models are developed, which include traditional latent growth curve models with missing data (Lu et al. 2013b), robust growth curve models (Zhang et al. 2013) with missing data (Lu et al. 2013a), growth mixture models (e.g., Bartholomew and Knott 1999) with missing data (Lu and Zhang 2014), extended growth mixture models (EGMMs, Muthén and Shedden 1999) with missing data (Lu and Zhang 2014), and robust growth mixture models with missing data (Lu and Zhang 2014).

In the following, we discuss three types of models: traditional growth models (including growth curve models, growth mixture models, and extended growth mixture models), robust growth models (including three types of robust models), and models that account for missingness (we mainly focus on non-ignorable missingness). By combining different elements of these models, it becomes possible to consider a series of growth models with a variety of missing data mechanisms and contaminated data.

21.2.1 Traditional Growth Models

The density for a latent growth curve model is

$$\begin{cases} y_i = \Lambda \eta_i + \mathbf{e}_i , \\ \eta_i = \boldsymbol{\beta} + \boldsymbol{\xi}_i , \end{cases} \tag{21.1}$$

where y_i is a $T \times 1$ vector of outcomes for participant $i\,(i = 1,\ldots,N)$, η_i is a $q \times 1$ vector of latent effects, Λ is a $T \times q$ matrix of factor loadings for η_i, $\mathbf{e}_i$ is a $T \times 1$ vector of residual or measurement errors, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed-effects, and $\boldsymbol{\xi}_i$ captures the variation of η_i. We have to note that $\mathbf{e}_i$ and $\boldsymbol{\xi}_i$ are usually assumed normally distributed but not necessary. When data have outliers and are heavy-tailed, this assumption might cause estimate biases. To reduce the effects of outliers, we adopt robust models in this study.

The density function of a growth mixture model is

$$f(y_i) = \sum_{k=1}^{K} \pi_k f_k(y_i), \tag{21.2}$$

where π_k is the invariant class probability (or weight) for class k satisfying $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$ (e.g., McLachlan and Peel 2000), and $f_k(y_i)(k = 1,\ldots,K)$ is the density of a latent growth model for class k.

For extended growth mixture models (EGMMs, Muthén and Shedden 1999), π_k is not invariant across individuals. It is allowed to vary individually depending on covariates, so it is expressed as $\pi_{ik}(\mathbf{x}_i)$. If a probit link function is used, then

$$\begin{cases} \pi_{i1}(\mathbf{x}_i) = \Phi(X_i'\,\boldsymbol{\varphi}_1), \\ \pi_{ik}(\mathbf{x}_i) = \Phi(X_i'\,\boldsymbol{\varphi}_k) - \Phi(X_i'\,\boldsymbol{\varphi}_{k-1}),\ (k = 2,3,\ldots,K-1) \\ \pi_{iK}(\mathbf{x}_i) = 1 - \Phi(X_i'\,\boldsymbol{\varphi}_{K-1}), \end{cases} \tag{21.3}$$

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 $\Phi(X_i'\boldsymbol{\varphi}_K) \equiv 1$.

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), that is, $\mathbf{z}_i \sim MultiNomial(\pi_{i1}, \pi_{i2}, \ldots, \pi_{iK})$.

21.2.2 Robust Growth Models

When data have outliers and are heavy-tailed, robust methods are used to reduce the effects of outliers. As t-distributions are more robust than normal distributions, the following are robust growth models (Lu et al. 2013a; Zhang et al. 2013).

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

$$\begin{cases} \mathbf{e}_i \sim Mt_T(\mathbf{0}, \boldsymbol{\Theta}, \nu), \\ \boldsymbol{\xi}_i \sim MN_q(\mathbf{0}, \boldsymbol{\Psi}), \end{cases} \tag{21.4}$$

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 covariance matrix $\boldsymbol{\Psi}$.

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

$$\begin{cases} \mathbf{e}_i \sim MN_T(\mathbf{0},\boldsymbol{\Theta}), \\ \boldsymbol{\xi}_i \sim Mt_q(\mathbf{0},\boldsymbol{\Psi},u). \end{cases} \tag{21.5}$$

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

$$\begin{cases} \mathbf{e}_i \sim Mt_T(\mathbf{0},\boldsymbol{\Theta},\nu), \\ \boldsymbol{\xi}_i \sim Mt_q(\mathbf{0},\boldsymbol{\Psi},u). \end{cases} \tag{21.6}$$

21.2.3 Non-ignorable Missingness

To build models with non-ignorable missingness, selection models (Glynn et al. 1986; Little 1993, 1995) are used. For individual i, let $\mathbf{m}_i = (m_{i1}, m_{i2}, \dots, m_{iT})'$ be a missing data indicator for 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 $m_{it} \sim \text{Bernoulli}(\tau_{it})$, so its density function is $f(m_{it}) = \tau_{it}^{m_{it}}(1-\tau_{it})^{(1-m_{it})}$. The missingness probability τ_{it} can have different forms. Lu and Zhang (2014) proposed the following non-ignorable missingness mechanisms for mixture models.

(1) Latent-Class-Intercept-Dependent (LCID) missingness in which τ_{it} is a function of latent class, covariates, and latent individual initial levels. For example, students are more likely to miss a test if their starting levels of that course are low. We model it as follows.

$$\tau_{it} = \boldsymbol{\Phi}(\mathbf{z}_i'\boldsymbol{\gamma}_{zt} + I_i\gamma_{It} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt}), \tag{21.7}$$

where I_i is the latent initial levels for individual i, γ_{It} is the coefficient for I_i, $\boldsymbol{\gamma}_{zt}$ is the coefficient for class membership, and $\boldsymbol{\gamma}_{xt}$ are coefficients for covariates. For non-mixture homogenous growth models, LCID can be simplified to Latent-Intercept-Dependent (LID) without the class membership indicator $\mathbf{z}_i$ and expressed as $\tau_{it} = \boldsymbol{\Phi}(\gamma_{0t} + I_i\gamma_{It} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt})$, where γ_{0t} is the intercept.

(2) Latent-Class-Slope-Dependent (LCSD) missingness in which τ_{it} is a function of latent class, covariates, and latent individual slopes of growth. For example, students are more likely to miss a test if they have slow growth of the course. In this case, τ_{it} can be modelled as

$$\tau_{it} = \boldsymbol{\Phi}(\mathbf{z}_i'\boldsymbol{\gamma}_{zt} + S_i\gamma_{St} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt}), \tag{21.8}$$

where S_i is the latent slope for individual i, and γ_{St} is the coefficient for S_i. Similarly, for non-mixture homogenous growth models, LCSD is simplified to Latent-Slope-Dependent (LSD) case as $\tau_{it} = \Phi(\gamma_{0t} + S_i\gamma_{St} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt})$.

(3) Latent-Class-Outcome-Dependent (LCOD) missingness in which τ_{it} is a function of latent class, covariates, and potential outcomes that may be missing. For example, a student who feels he/she is not doing well on the test may be more likely to give up taking the rest of the test. We express τ_{it} as

$$\tau_{it} = \Phi(\mathbf{z}_i'\boldsymbol{\gamma}_{zt} + y_{it}\gamma_{yt} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt}), \tag{21.9}$$

where y_{it} is the potential outcomes for individual i at time t, and γ_{yt} is the coefficient for y_{it}. And LCOD can be simplified to Latent-Outcome-Dependent (LOD) for non-mixture homogeneous growth models with a probability of missingness $\tau_{it} = \Phi(\gamma_{0t} + y_{it}\gamma_{yt} + \mathbf{x}_i'\boldsymbol{\gamma}_{xt})$.

In a more general framework, LCID and LCSD can be further encompassed into Latent-Class-Random Effect-Dependent missingness as intercept and slope are different random effects according to different situations under consideration. And for non-mixture structure, LID and LSD are encompassed into Latent-Random Effect-Dependent missingness.

21.3 Bayesian Selection Criteria

Based on Lu et al. (2013a), model selection criteria are proposed in the framework of Bayesian growth models with missing data. The definitions of selection criteria are listed in Table 21.1. The model selection criteria in the table are based on two versions of deviance in the Bayesian context, $E_{D|y}[D(\theta)]$ and $D(E_{\theta|y}[\theta])$. As we have discussed in the introduction section, $E_{\theta|y}[D]$ is the expected value of all the deviances, and $D(E_{\theta|y}[\theta])$ is the deviance score based on the expected parameters. For different models, the detailed mathematical form of these two deviances is different. In this paper, we focus on both homogeneous and heterogenous latent growth models with non-ignorable missing data.

(1) We first look at the homogeneous growth curve models with non-ignorable missing data. One version of deviance, $E_{D|y}[D(\theta)]$, is approximated by

$$\begin{aligned} E_{D|y}[D(\theta)] \approx \overline{D(\theta)} &= -\frac{2}{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{1=t}^{T} l_{it}^{(s)}(\theta|y,m) \\ &= -\frac{2}{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{1=t}^{T}\left[(1-m_{it}^{(s)})l_{it}^{(s)}(y) + l_{it}^{(s)}(m)\right], \end{aligned} \tag{21.10}$$

Table 21.1 Model selection criteria

Criterion(Index) =	Deviance +	Penalty
Dbar.AIC[a]	$\overline{D(\theta)}$[b]	2 p
Dbar.BIC[c]	$\overline{D(\theta)}$	log(N) p
Dbar.CAIC	$\overline{D(\theta)}$	(log(N)+1) p
Dbar.ssBIC	$\overline{D(\theta)}$	log((N+2)/24) p
RDIC	$\overline{D(\theta)}$	var(Dbar)/2
Dhat.AIC	$D(\hat{\theta})$[d]	2 p
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[e]	$D(\hat{\theta})$	2 pD

[a] p is the number of parameters, which are on the same level as the likelihood value is.
[b] $\overline{D(\theta)}$ is shown as in Eq. (21.10) for growth curve models and as in Eq. (21.13) for growth mixture models. It is one type of the approximations of the deviance score.
[c] N is the sample size.
[d] $D(\hat{\theta})$ is shown as in Eq. (21.12) for growth curve models and as in Eq. (21.14) for growth mixture models
[e] $pD = \overline{D(\theta)} - D(\hat{\theta})$

where S is the number of iterations for converged Markov chains, $l_{it}^{(s)}(\theta|y,m) = log(L_{it}^{(s)}(\theta|y,m))$ is a conditional joint loglikelihood function (see, Celeux et al. 2006) of y and m, m_{it} is the missing data indicator for individual i at time t with a likelihood function $l_{ikt}(m) = m_{it}log(\tau_{it}) + (1-m_{it})log(1-\tau_{it})$, where τ_{it} is the missing data rate for individual i at time t and is defined differently for different missingness models as in the previous section. When y_{it} is missing, the corresponding likelihood is excluded. So combining y and m, the conditional likelihood function of a selection model with non-ignorable missing data can be expressed as

$$L_{it}(\theta|y,m) = [f(y_{it}|\eta_i)(1-\tau_{it})]^{(1-m_{it})}\ \tau_{it}^{m_{it}}, \tag{21.11}$$

And the other version of deviance, $D(E_{\theta|y}[\theta])$, is approximated by

$$D(E_{\theta|y}[\theta]) \approx D(\hat{\theta}) = -2\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(1-m_{it})l_{it}(y|\hat{\theta}) + l_{it}(m|\hat{\theta})\right], \tag{21.12}$$

where $\hat{\theta}$ is the posterior mean of parameter estimates across S iterations.

(2) For growth mixture models with missing data, $E_{\theta|y}[D]$ is expressed as

$$E_{D|y}[D(\theta)] \approx \overline{D(\theta)} = -\frac{2}{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}^{(s)} \sum_{1=t}^{T}\left[(1-m_{it})l_{ikt}^{(s)}(y) + l_{ikt}^{(s)}(m)\right], \quad (21.13)$$

where $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iK})$ is the class membership indicator which follows a multinomial distribution, $\mathbf{z}_i \sim MultiNomial(\pi_{i1}, \pi_{i2}, \ldots, \pi_{iK})$, and $z_{ik}^{(s)}$ is the class membership estimated at iteration s. And

$$D(E_{\theta|y}[\theta]) \approx D(\hat{\theta}) = -2\sum_{i=1}^{N}\sum_{k=1}^{K}\hat{z}_{ik}\sum_{t=1}^{T}\left[(1-m_{it})l_{ikt}(y|\hat{\theta}) + l_{ikt}(m|\hat{\theta})\right], \quad (21.14)$$

where $\hat{z}_{ik}$ is the posterior mode of class membership, $\hat{\theta}$ is the posterior mean of parameter estimates across all S iterations. In both the $\overline{D(\theta)}$ and $D(\hat{\theta})$ definitions of deviance, $l_{ikt}(y)$ and $l_{ikt}(m)$ are the conditional loglikelihood functions for y_{it} and m_{it}, respectively, for individual i in class k at time t.

If people calculate deviance scores using $D(\hat{\theta})$, then $\overline{D(\theta)}$ is the sum of an approximation of the deviance score ($D(\hat{\theta})$) and some penalties. The difference between $\overline{D(\theta)}$ and $D(\hat{\theta})$ can be quantified by a statistic called pD (Spiegelhalter et al. 2002),

$$pD = \overline{D(\theta)} - D(\hat{\theta}). \quad (21.15)$$

Based on the Jensen's inequality (Casella and George 1992), when $D(\theta)$ is convex, then $\overline{D(\theta)} \geq D(\hat{\theta})$ and as a result pD is positive. When $D(\theta)$ is concave, then $\overline{D(\theta)} \leq D(\hat{\theta})$ and pD is negative.

21.4 Simulation Studies

In this section, five simulation studies are conducted to evaluate the performance of the Bayesian criteria. For each study, four waves of complete data were generated first and then missing data were created on each occasion according to pre-designed missing data rates. After data are generated, full Bayesian methods are used by adopting uninformative priors, obtaining conditional posterior distributions through application of a data augmentation algorithm, generating Markov chains through a Gibbs sampling procedure, conducting convergence testing, and making statistical inference for model parameters. For all simulations, the software OpenBUGS is used for the implementation of Gibbs sampling, and R codes are written for data-generation, convergence testing, and parameter estimation.

The five studies are designed such that the data complexity increases from study 1 to study 5. Studies 1–2 focus on non-mixture growth data and thus, latent growth curve models with missing data are used. Studies 3–5 focus on mixture growth data and thus, growth mixture models with missing data are used. Simulation factors include measurement error distributions, random effect distributions, missingness patterns, sample size, and class separation (Anderson and Bahadur 1962). Under each condition, 100 converged replications are used to calculate the model selection proportion. Table 21.2 lists the design details.

Study 1 investigated the performance of the Bayesian criteria when data were non-mixture homogenous, normally distributed with non-ignorable missingness. The true model was NN-XS, which was the model with normally distributed measurement errors ($\mathbf{e}_i$) at level 1 and random effects ($\boldsymbol{\xi}_i$) at level 2, with missingness depending on covariate x and latent slope S. Specifically, $\mathbf{e}_i \sim MN(\mathbf{0}, \mathrm{I})$, $\eta_i \sim MN_q(\boldsymbol{\beta}, \boldsymbol{\Psi})$ where $\boldsymbol{\beta} = (\text{Intercept}, \text{Slope}) = (1, 3)$ and $\boldsymbol{\Psi}$ was a 2 by 2 symmetric matrix with $Var(I) = 1$, $Cov(I, S) = 0$, and $Var(S) = 4$. For missingness, the bigger the latent slope was, the higher the missing data rate would be. The missingness probit coefficients were set as $\gamma_0 = (-1, -1, -1, -1)$, $\gamma_x = (-1.5, -1.5, -1.5, -1.5)$, and $\gamma_S = (0.5, 0.5, 0.5, 0.5)$. For example, if a participant had a latent growth slope 3, with a covariate value 1, then his or her missing probability at each wave was $\tau \approx 16\%$; if the slope was 5, with the same covariate value, the missing probability increased to $\tau = 50\%$; but if the slope was 1, then the missing probability decreased to $\tau = 2.3\%$. The covariate x was also generated from a normal distribution, $x \sim N(1, sd = 0.2)$. In study 1, totally there were 16 conditions with 4 missingness mechanisms (XS non-ignorable, XY non-ignorable, XI non-ignorable, and ignorable) combined with 4 levels of sample size (1,000, 500, 300, and 200). Table 21.3 lists the model selection proportions across 100 replications for each of these criteria across all conditions in study 1. The largest proportion across four missingness models is indicated in the shaded cell for each criterion. When sample size is relatively large, 1,000 or 500, all of the model selection criteria, except for the rough DIC (RDIC), correctly identify the true model with 100 %. When sample size becomes smaller, 300 or 200, except for the RDIC, all of the model selection criteria choose the true model with certainty above 93 %. Comparing the criteria defined based on Dbar with those defined based on Dhat, one can see that the former performs a little bit better.

Study 2 investigated the performance of these criteria when data were non-mixture homogeneous with outliers and non-ignorable missingness. The main difference between study 2 and 1 was that the data for study 2 contain outliers such that they are not normally distributed. So robust growth curve models were used. The true model was TN-XS, which means measurement errors ($\mathbf{e}_i$) at level 1 followed a t-distribution. Specifically, $\mathbf{e}_i$ were generated from a t distribution with 5 degrees of freedom and a scale matrix I, i.e., $\mathbf{e}_i \sim Mt(\mathbf{0}, \mathrm{I}, 5)$. Other settings were kept the same as those in study 1. In this study, totally 32 conditions were considered with 4 data distributions (NN, TN, NT, and TT), 4 missingness patterns (XS non-ignorable, XY non-ignorable, XI non-ignorable, and ignorable), and 2 levels of sample size (1,000 and 500). Table 21.4 lists the model selection proportions. The largest proportion across 16 missingness models is indicated in the shaded cell for each criterion. Except for the RDIC, all of the model selection criteria correctly identify the true model. TT-XS is a competing model, which also gains high selection probabilities. This is because the normal distribution is almost identical

Table 21.2 Simulation study design

Study	Model	Data distribution				Missingness depends on					Sample size			Class separation[c]	
		$\mathbf{e}_i$[a]		η_i[b]											
		N[d]	t[e]	N	t	C[f]	X[g]	I[h]	S[i]	Y[j]	Different			M	S
Study 1	Normal LGCMs: use relative small sample sizes due to single-class data														
	NN-ignorable	✓		✓			✓								
	NN-XI	✓		✓			✓	✓							
	NN-XS[k]	✓		✓			✓		✓						
	NN-XY	✓		✓			✓			✓					
Study 2	Robust LGCMs: use relative small sample sizes due to single-class data														
	TN-ignorable		✓	✓			✓								
	TN-XI		✓	✓			✓	✓							
	TN-XS		✓	✓			✓		✓						
	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	✓		✓			✓			✓					

(continued)

Table 21.2 (continued)

Study	Model	Data distribution				Missingness depends on					Sample size	Class separation[c]	
		$\mathbf{e}_i^{a}$		η_i^{b}									
		N[d]	t[e]	N	t	C[f]	X[g]	I[h]	S[i]	Y[j]	Different	M	S
Study 3	Robust GMMs (RGMMs): use relative large sample sizes due to multiple classes data, and use small class separation due to fixed class probabilities												
	TN-ignorable		✓	✓			✓					✓	
	TN-XI		✓	✓			✓	✓				✓	
	TN-XS		✓	✓			✓		✓			✓	
	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	✓		✓			✓			✓		✓	
Study 4	Robust Extended GMMs (REGMMs): select 5 competing models based on the performance in Study 3 use relative large sample sizes due to multiple-class data and varied class probabilities												
	TN-CXS		✓	✓		✓	✓		✓			✓	✓
	TN-CX		✓	✓		✓	✓					✓	✓

	TT-CXS		✓			✓		✓	✓		✓							✓	✓
	NN-CXS	✓			✓			✓	✓		✓							✓	✓
	NN-CX	✓			✓			✓	✓									✓	✓
Study 5	Single-Class LGCMs vs. Multiple-Class RGMMs																		
	1 Class LGCMs																		
	TN-XS		✓		✓				✓		✓								
	TT-XS		✓			✓			✓		✓								
	NN-XS	✓			✓				✓		✓								
	2 Classes RGMMs																		
	TN-XS		✓		✓				✓		✓							✓	
	TT-XS		✓			✓			✓		✓							✓	
	NN-XS	✓			✓				✓		✓							✓	
	3 Classes RGMMs																		
	TN-XS		✓		✓				✓		✓								
	TT-XS		✓			✓			✓		✓								
	NN-XS	✓			✓				✓		✓								
	4 Classes RGMMs																		
	TN-XS		✓		✓				✓		✓								
	TT-XS		✓			✓			✓		✓								
	NN-XS	✓			✓				✓		✓								

[a] Measurement errors
[b] Random effects
[c] Class separation (Anderson and Bahadur 1962) when generating data (S: small = 1.7, M: medium = 2.7)
[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] The shaded model is the true model

Table 21.3 Model selection proportion in study 1

	N = 1,000				N = 500			
	Non-ignorable			Ignorable	Non-ignorable			Ignorable
Criterion[a]	NN-XS[b]	NN-XY[c]	NN-XI[d]	NN[e]	NN-XS	NN-XY	NN-XI	NN
Dbar.AIC	1[f]	0.000	0.000	0.000	1	0.000	0.000	0.000
Dbar.BIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
Dbar.CAIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
Dbar.ssBIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
RDIC	0.013	0.000	0.987	0.000	0.038	0.000	0.962	0.000
Dhat.AIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
Dhat.BIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
Dhat.CAIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
Dhat.ssBIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
DIC	1	0.000	0.000	0.000	1	0.000	0.000	0.000
	N = 300				N = 200			
Dbar.AIC	0.98125	0.01875	0.000	0.000	0.975	0.025	0.000	0.000
Dbar.BIC	0.98125	0.01875	0.000	0.000	0.975	0.025	0.000	0.000
Dbar.CAIC	0.98125	0.01875	0.000	0.000	0.975	0.025	0.000	0.000
Dbar.ssBIC	0.98125	0.01875	0.000	0.000	0.975	0.025	0.000	0.000
Rough DIC	0.1125	0.000	0.8875	0.000	0.2	0.03125	0.76875	0.000
Dhat.AIC	0.95	0.05	0.000	0.000	0.9375	0.06875	0.000	0.000
Dhat.BIC	0.95	0.05	0.000	0.000	0.9375	0.06875	0.000	0.000
Dhat.CAIC	0.95	0.05	0.000	0.000	0.9375	0.06875	0.000	0.000
Dhat.ssBIC	0.95	0.05	0.000	0.000	0.9375	0.06875	0.000	0.000
DIC	1	0.000	0.000	0.000	0.98125	0.0125	0.00625	0.000

[a] The definition of each criterion is given in Table 21.1

[b] The shaded model is the true model. The model is normal-distribution-based with latent-slope-dependent missingness

[c] The model is normal-distribution-based with potential-outcome-dependent missingness

[d] The model is normal-distribution-based with latent-intercept-dependent missingness

[e] The model is normal-distribution-based with ignorable missingness

[f] The shaded cell has the largest proportion

to a t-distribution with large degrees of freedom. The degrees of freedom of t is also estimated by the model. Also, the Dbar-based criteria perform a little bit better than the Dhat-based criteria. Among them, Dbar-based BIC and CAIC perform best.

Study 3 was designed for mixture data with outliers and non-ignorable missing data. As data were mixture, growth mixture models were used. In this study, the true model was 2-class mixture TN-XS RGMM. Only intercepts of these 2 classes were different, with 5 for class 1 and 1 for class 2. Other settings for each class were the same as in study 2. Both classes have t_5 distributed measurement errors. Based on Anderson and Bahadur (1962), the class separation is around 2.7. In this study, we assumed they are traditional mixture models, i.e., class probabilities were fixed. We were fixed as (50 %, 50 %) in this study. Similar as in study 2, there were 32 conditions considered with 4 data distributions (NN, TN, NT, and TT), 4 missingness patterns (XS non-ignorable, XY non-ignorable, XI non-ignorable, and ignorable), and 2 levels of sample size (1,000 and 1,500). As mixture data require more data to obtain estimates, we increased the sample size. Table 21.5 shows the results for study 3. The shaded cell indicates the largest proportion across 16 missingness models for each criterion. Again, almost all of the model selection criteria correctly identify the true model. And the Dbar-based criteria perform a little bit better than the Dhat-based criteria. Specifically, Dbar-based BIC and CAIC perform best among these criteria, and then Dbar-based ssBIC also performs well.

Study 4 extended study 3 such that the class probabilities were not fixed. Instead, they depended on values of covariates. Also, the non-ignorable missingness in this study was allowed to depend on the corresponding observations' latent class membership. The true model in this study was 2-class mixture TN-CXS robust extended growth mixture models (REGMM). The differences between this study and study 3 were (1) the class proportions in this study were predicted by the value of covariate x; (2) the missing data rates were predicted by the latent class membership; (3) both medium size, 2.7, and small size, 1.7, class separations were used. Specifically, for small class separation, the intercept for class 1 was 3.5 and the intercept for class 2 was 1. To simplify the simulation, based on the findings in study 3, 5 competing mixture models (TN-CXS, TT-CXS, TN-CX, NN-CXS, and NN-CX) were chosen to fit the data. Totally, we considered 20 conditions with 5 mixture models, 2 levels of sample size (1,500 and 1,000), and 2 levels of class separation (2.7 and 1.7). Table 21.6 shows the model selection proportions in study 4. Again, almost all of the model selection criteria correctly identify the true model. Specifically, Dbar-based BIC and CAIC perform best among these criteria.

Study 5 focused on the number of classes. In this study, different growth curve models with different numbers of classes were fitted and compared. In total, 9 conditions were considered, including 3 models (TN-XS, TT-XS, NN-XS) and 3 numbers of classes (1, 2, and 3). The true model was the 2-class mixture TN-XS model. The simulation results for study 5 were presented in Table 21.7. Among these criteria, Dhat-based criteria perform better than Dhbar-based criteria. Specifically, Dhat-based BIC and CAIC perform best, and ssBIC and AIC also provide high certainty.

Table 21.4 Model selection proportion in study 2

Criterion		N = 1,000				N = 500			
		Non-ignorable				Non-ignorable			
		XS[a]	XY	XI	Ignorable	XS	XY	XI	Ignorable
Dbar.AIC	TN[b]	0.519	0.000	0.000	0.000	0.597	0.013	0.000	0.000
	TT[c]	0.469	0.000	0.000	0.012	0.377	0.000	0.000	0.000
	NT[d]	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
	NN[e]	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
Dbar.BIC	TN	0.781	0.000	0.000	0.000	0.855	0.013	0.000	0.000
	TT	0.200	0.000	0.000	0.019	0.113	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000
Dbar.CAIC	TN	0.819	0.000	0.000	0.000	0.888	0.012	0.000	0.000
	TT	0.162	0.000	0.000	0.019	0.075	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000
Dbar.ssBIC	TN	0.625	0.000	0.000	0.000	0.631	0.012	0.000	0.000
	TT	0.362	0.000	0.000	0.012	0.338	0.000	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000
RDIC	TN	0.000	0.000	0.106	0.000	0.000	0.000	0.094	0.000
	TT	0.000	0.000	0.100	0.000	0.000	0.000	0.113	0.000
	NT	0.000	0.000	0.394	0.000	0.000	0.000	0.390	0.000
	NN	0.000	0.000	0.400	0.000	0.000	0.000	0.403	0.000
Dhat.AIC	TN	0.544	0.000	0.000	0.000	0.547	0.025	0.000	0.000
	TT	0.506	0.006	0.000	0.000	0.447	0.019	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.BIC	TN	0.675	0.006	0.000	0.000	0.717	0.025	0.000	0.000
	TT	0.319	0.000	0.000	0.000	0.245	0.013	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.CAIC	TN	0.700	0.006	0.000	0.000	0.788	0.025	0.000	0.000
	TT	0.294	0.006	0.000	0.000	0.169	0.012	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dhat.ssBIC	TN	0.575	0.006	0.000	0.000	0.588	0.025	0.000	0.000
	TT	0.419	0.006	0.000	0.000	0.369	0.012	0.000	0.000
	NT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	NN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DIC	TN	0.325	0.000	0.000	0.000	0.415	0.006	0.000	0.000
	TT	0.462	0.000	0.000	0.194	0.409	0.000	0.000	0.000
	NT	0.012	0.000	0.000	0.000	0.088	0.000	0.000	0.000
	NN	0.006	0.000	0.000	0.000	0.082	0.000	0.000	0.000

[a]Other abbreviations are as given in Table 21.3
[b]Growth model with t-distributed measurement errors and normally distributed random effects
[c]Growth model with t-distributed measurement errors and t-distributed random effects
[d]Growth model with normally distributed measurement errors and t-distributed random effects
[e]Growth model with normally distributed measurement errors and random effects

Table 21.5 Model selection proportion in study 3

Criterion		N = 1,500				N = 1,000			
		Non-ignorable			Ignorable	Non-ignorable			Ignorable
		XS	XY	XI		XS	XY	XI	
Dbar.AIC	TN	0.621	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

Abbreviations are as given in Table 21.3

Table 21.6 Model selection proportion in study 4

Criterion	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 = 1,500							Class separation = 2.7, N = 1,000				
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 = 1,500							Class separation = 1.7, N = 1,000				
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

Abbreviations are as given in Table 21.3

Table 21.7 Model selection proportion in study 5

	2 CLASSES			1 CLASS			3 CLASSES		
Criterion	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

Abbreviations are as given in Table 21.3

21.5 Application

In this section, a real data set on mathematical growth is analyzed to demonstrate the application of the criteria. 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, U.S. Department of Labor 1997), which were collected from $N = 1{,}510$ adolescents yearly from 1997 to 2001 when each adolescent was administered the Peabody Individual Achievement Test (PIAT) Mathematics Assessment to measure their mathematical ability. There are some outliers at all five grades. Lu et al. (2011) conducted a power transformation to normalize the sample and assumed the data are normally distributed without outliers. In this study, however, we use the original non-transformed data with outliers, so robust methods are used. Also, different non-ignorable missingness mechanisms are considered. 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.

The analysis is conducted following the steps in Table 21.8. 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, to compare models with different non-ignorable missingness and numbers of classes, 10 models 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 ensure estimates are based on the Markov chains that have converged. For testing convergence, the history plot is examined and the Geweke's z statistic (Geweke 1992) is checked for each parameter. The Geweke's z statistics for all the parameters are smaller than

Table 21.8 Steps and fitting models in real data analysis

Step 1:	Fit a tentative 2 classes model, and check the estimated df of t									
		$\mathbf{e}_i$		η_i		missingness				
	Model	N	T	N	T	C	X	I	S	Y
	TT-ignorable		✓		✓					
Step 2:	Try models with different missingness and number of classes									
	2 Classes RGMMs									
	TN-X		✓	✓			✓			
	TN-XI		✓	✓			✓	✓		
	TN-XS		✓	✓			✓		✓	
	TN-XY		✓	✓			✓			✓
	2 Classes REGMMs									
	TN-CX		✓	✓		✓	✓			
	TN-CXI		✓	✓		✓	✓	✓		
	TN-CXS		✓	✓		✓	✓		✓	
	TN-CXY		✓	✓		✓	✓			✓
	3 Classes GMMs									
	NN-X	✓		✓			✓			
	4 Classes GMMs									
	NN-X	✓		✓			✓			
Step 3:	Compare selection criteria									
Step 4:	Interpret results obtained from the selected model									

Abbreviations are as given in Table 21.2

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). In step 3, model selection criterion is used to compare the ten models. The indices are listed in Table 21.9. And in step 4, the results obtained from the final selected model are interpreted.

As suggested by Dhat.CAIC, Dhat.ssBIC, Dhat.BIC, and Dhat.AIC, without further substantive information, the TN-CXY model would appear to be a good candidate for best-fitting model. Table 21.10 provides the results of the TN-CXY REGMM model. It can be seen 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 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)

Table 21.9 Model selection in real data analysis

Criterion[a]	2 CLASSES TN-CXS	TN-CXY	TN-CXI	TN-CX	TN-XS	TN-XY	TN-XI	TN-X	3 CLASSES NN-X	4 CLASSES 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	17,583.52	17,663.52	17,693.52	17,666.92	17,556.92	17,646.92	17,666.92	17,650.32	17,536.92	17,328.15
Dbar.CAIC	17,619.52	17,699.52	17,729.52	17,697.92	17,587.92	17,677.92	17,697.92	17,676.32	17,567.92	17,366.15
Dbar.ssBIC	17,469.15	17,549.15	17,579.15	17,568.44	17,458.44	17,548.44	17,568.44	17,567.72	17,438.44	17,207.44
RDIC	22,759.24	22,704.5	22,378.14	22,601.28	22,562.65	22,755.44	22,973.52	22,520.18	22,843.52	23,333.2
Dhat.AIC	15,192	14,942	17,482	19,822	21,922	23,622	25,722	27,352	15,872	15,716
Dhat.BIC	15,383.52	15,133.52	17,673.52	19,986.92	22,086.92	23,786.92	25,886.92	27,490.32	16,036.92	15,918.15
Dhat.CAIC	15,419.52	15,169.52	17,709.52	20,017.92	22,117.92	23,817.92	25,917.92	27,516.32	16,067.92	15,956.15
Dhat.ssBIC	15,269.15	15,019.15	17,559.15	19,888.44	21,988.44	23,688.44	25,788.44	27,407.72	15,938.44	15,797.44
DIC	19,520	19,930	17,450	15,120	12,800	11,280	9,220	7,620	18,810	18,460

The shaded cell has the smallest value
[a]The definition of each criterion is given in Table 21.1

Table 21.10 Estimates of TN-CXY REGMM in real data analysis

		Parameter	Mean	S.D.	MC.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	2.900	0.050
	Class 2	Intercept	6.196	0.047	0.020	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	0.020
		Var(S)	0.034	0.004	0.022	0.027	0.042	0.010
		Cov(I,S)	0.010	0.014	0.021	−0.018	0.037	−0.023
		Var(e)	0.372	0.020	0.033	0.336	0.412	−0.061
		df_y	3.200	0.195	0.040	2.850	3.600	−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	0.310	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	−0.190	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	0.330
	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	0.050	0.065	−0.266	−0.084	0.378
	Grade 11	γ^*_{05}	−1.298	0.421	0.065	−2.130	−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	0.040	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,c]The lower 2.5 percentile and upper 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. (21.3)
[g]The probit coefficient of the class membership 1 at Grade 7, defined in Eq. (21.9)
[h]The probit coefficient of the class membership 2 at Grade 7, defined in Eq. (21.9)
[i]The probit coefficient of the covariate at Grade 7, defined in Eq. (21.9)
[j]The probit coefficient of the potential output Y at Grade 7, defined in Eq. (21.9)

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.

21.6 Conclusions and Future Research

Based on the results from the five simulation studies, one can conclude that (1) almost all of the model selection criteria, except for the rough DIC (RDIC), can correctly choose the true model with high certainty; (2) if the number of classes is correctly identified, then the Dbar-based criteria perform better than the Dhat-based criteria; if candidate models have different numbers of classes, then the Dhat-based criteria might be used to select the best-fit model; (3) across five studies, CAIC and BIC provide higher probabilities than those ssBIC, AIC, or DIC does. The results will help inform the selection of growth models by researchers seeking to provide people with accurate estimates of growth across a variety of possible contexts. The real data analysis demonstrated the application of the criteria to typical longitudinal growth studies such as educational, psychological, and social research. Future research of this study includes proposing more effective model selection criteria, such as Bayes factors, and testing their performance with more practice statistical models, such as survival models.

Acknowledgments The authors thank the reviewer Dr. Daniel Bolt for his very helpful comments and suggestions, which greatly improved the quality of this article.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 1919(6):716–723

Anderson TW, Bahadur RR (1962) Classification into two multivariate normal distributions with different covariance matrices. Ann Math Stat 33:420–431

Bartholomew DJ, Knott M (1999) Latent variable models and factor analysis: Kendall's library of statistics, vol 7, 2nd edn. Edward Arnold, New York

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, U.S. 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. Retrieved from http://www.bls.gov/nls/nlsy97.htm

Casella G, George EI (1992) Explaining the Gibbs sampler. Am Stat 46(3):167–174

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

Dunson DB (2000) Bayesian latent variable models for clustered mixed outcomes. J R Stat Soc B 62:355–366

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernado JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 4. Clarendon Press, Oxford, pp 169–193

Glynn RJ, Laird NM, Rubin DB (1986) In: Wainer H (ed) Drawing inferences from self-selected samples. Springer, New York, pp 115–142

Little RJA (1993) Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc 88:125–134

Little RJA (1995) Modelling the drop-out mechanism in repeated-measures studies. J Am Stat Assoc 90:1112–1121

Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. Wiley-Interscience, New York

Lu Z, Zhang Z (2014) Robust growth mixture models with non-ignorable missingness data: models, estimation, selection, and application. Comput Stat Data Anal 71:220–240

Lu Z, Zhang Z, Lubke G (2011) Bayesian inference for growth mixture models with latent-class-dependent missing data. Multivariate Behav Res 46:567–597

Lu Z, Zhang Z, Cohen A (2013a) Bayesian inference for latent growth curve models with non-ignorable missing data. Struct Equ Modeling (manuscript submitted for publication)

Lu ZL, Zhang Z, Cohen A (2013b) In: Millsap RE, van der Ark LA, Bolt DM, Woods CM (eds) New developments in quantitative psychology, vol 66. Springer, New York, pp 275–304

Maronna RA, Martin RD, Yohai VJ (2006) Robust statistics: theory and methods. Wiley, New York

McLachlan GJ, Peel D (2000) Finite mixture models. Wiley, New York

Muthén B, Shedden K (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55(2):463–469

Oldmeadow C, Keith JM (2011) Model selection in Bayesian segmentation of multiple DNA alignments. Bioinformatics 27:604–610

Schwarz GE (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464

Sclove LS (1987) Application of mode-selection criteria to some problems in multivariate analysis. Psychometrics 52:333–343

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 64(4):583–639

Spiegelhalter DJ, 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)

Yuan K-H, Lu Z (2008) SEM with missing data and unknown population using two-stage ML: theory and its application. Multivariate Behav Res 43:621–652

Zhang Z, Lai K, Lu Z, Tong X (2013) Bayesian inference and application of robust growth curve models using student's t distribution. Struct Equ Modeling 20(1):47–78