Bayesian Methods and Model Selection for Latent Growth Curve Models with Missing Data

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

1 Introduction

There has been widespread interest in the analysis of change in social and behavioral sciences (e.g., Singer and Willett 2003). Growth modeling, in particular, is becoming increasingly important in these areas. Among the most popular growth models, *latent growth curve models* (LGCMs) are statistical models designed 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). With an increase in complexity of LGCMs, comes an increase in difficulties estimating such models. First, missing data are almost inevitable with longitudinal data (e.g., Jelicic et al. 2009). Second, using conventional likelihood procedures may be challenging when estimating model parameters in complex models with complicated data structures. And third, even with effective estimation methods, model selection in such complex situations becomes difficult.

1.1 Missing Data

As LCGMs involve data collection on the same participants through multiple waves of surveys, tests, or questionnaires, missing data are almost inevitable. This is because some students may miss a test because of absence or fatigue or research

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.), *New Developments in Quantitative Psychology*,
Springer Proceedings in Mathematics & Statistics 66, DOI 10.1007/978-1-4614-9348-8_18,

participants may drop out of a study (e.g., Schafer 1997). Missing data can be investigated from their mechanisms, that is, by examining why missing data occur. Little and Rubin (2002) distinguished two mechanisms for missing data, *ignorable* and *non-ignorable*. For ignorable missingness, estimates are usually asymptotically consistent when the missingness is ignored (Little and Rubin 2002). This is because parameters that govern the missing process either are distinct from the parameters that govern the model outcomes or depend on the observed parameters in the fitted model. The non-ignorable missingness is also referred to as *missing not at random (MNAR)*, in which the missing data probability depends on unobserved outcomes or on some unobserved latent variables in the model.

With the appearance of missing data comes the challenge in estimating growth model parameters. Although there is a large literature addressing the problems of missing data in applied and quantitative psychology (e.g., Yuan and Lu 2008; Roth 1994), particularly in longitudinal studies (e.g., Jelicic et al. 2009), the majority of the literature is on ignorable missingness. This is mainly because (1) analysis models or techniques for non-ignorable missing data are traditionally difficult to implement and not yet well suited for widespread use (e.g., Baraldi and Enders 2010); and (2) missingness mechanisms are not testable (Little and Rubin 2002). At the same time, however, the analysis of non-ignorable missingness is a crucial and a serious concern in applied research areas, in which participants may be dropping out for reasons closely related to the response being measured (e.g., Enders 2011). Not attending to the non-ignorable missingness may result in severely biased statistical estimates, standard errors, and associated confidence intervals (e.g., Schafer 1997), and thus poses substantial risk of leading researchers to incorrect conclusions. Accordingly, this paper focuses on non-ignorable missingness and investigates its influences on model estimation for different types of missingness.

In a recent study of latent growth models, Lu et al. (2011) investigated non-ignorable missingness. However, the missingness in that study was only allowed to depend on latent class membership. In practice, the non-ignorable missingness in latent growth models can depend on many other latent variables such as individual starting level and growth rate. Furthermore, Lu et al. (2011) did not discuss how to identify the missingness mechanisms.

1.2 Bayesian Approach

In this study, a full Bayesian approach is used for parameter estimation. Previously, maximum likelihood methods were adopted for most of the studies, and statistical inferences were carried out using conventional likelihood procedures (e.g., Yuan and Lu 2008). Recently, Bayesian methods have been proposed as an alternative approach (e.g., Muthén and Asparouhov 2012) to estimate complex models. The advantages of Bayesian methods include their intuitive interpretations of statistical results, their flexibility in incorporating prior information about how data behave in similar contexts and findings from experimental research, their capacity for

dealing with small sample sizes (such as occur with special populations), and their expandability in the analysis of complex statistical models with complicated data structure (e.g., Lee 2007).

In a Bayesian approach, when the joint distribution is complex or unknown but the conditional distribution of each variable is available for each set of variables, Gibbs sampling algorithm (Geman and Geman 1984) can be adopted. The Gibbs sampling generates Markov chains which can be shown to be ergodic (Geman and Geman 1984), and thus the sequence of samples after convergence can be viewed from the joint probability distribution of all parameters. It is also shown that each variable from the Markov chain converges to the marginal distribution of that variable (Robert and Casella 2004).

1.3 Model Selection Criteria

Model selection criteria can be used to compare models to identify the best-fit model. Akaike (1974) proposed the Akaike's information criterion (AIC). AIC offers a relative measure of the information lost. For Bayesian models, the Bayes factor is used for hypothesis testing. But the Bayes factor is usually difficult or even impossible to calculate, especially for models that involve many random effects, large numbers of unknowns parameters, or improper priors. To approximate the Bayes factor, Schwarz (1978) developed the Bayesian information criterion (BIC) or 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) which is based on the Rissanen Information Criteria (RIC, Rissanen 1978) for auto-regressions. The deviance information criterion (DIC) (Spiegelhalter et al. 2002) is a recently developed criterion designed for complex hierarchical models. It is based on the posterior distribution of the log-likelihood, following the original suggestion of Dempster (1974) for model choice in the Bayesian framework, and it is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. DIC is usually regarded as a Bayesian version or generalization of the AIC and BIC. For all these criteria, the model with a smaller value is better supported by data.

In a Bayesian context, currently there are no well-defined model selection criteria for latent growth models with missing data (e.g., Celeux et al. 2006). The problem is mainly due to random effects and missing data. For random effects models, the likelihood function can be an observed-data likelihood, a complete-data likelihood, or a conditional likelihood. Briefly speaking, an observed-data likelihood does not explicitly include latent variables, such as random-effects; a complete-data likelihood includes all auxiliary variables in the model; and a conditional likelihood is the joint likelihood function of the observed outcomes and the missingness indicator conditional on the random-effects, and thus the likelihood only includes

random-effects, with no fixed-effects involved (e.g., Celeux et al. 2006). Also, the missing data part can be either included in or excluded from the log-likelihood functions.

1.4 Goals and Structure

The goals of the paper are to propose latent growth models with non-ignorable missingness, to estimate the models via a Bayesian approach, and to evaluate the performance of model selection criteria.

The rest of the paper consists of six sections. Section 2 describes the proposed growth models. Three non-ignorable missingness selection models are presented and formulated. Section 3 presents a full Bayesian method to estimate the latent growth models through data augmentation and Gibbs sampling algorithms. Section 4 proposes model selection criteria in a Bayesian context for growth models with missing data. Section 5 conducts simulation studies. Estimates from models with different non-ignorable missingness and different sample sizes are summarized, analyzed, and compared. Conclusions based on the simulation studies are drawn. Section 6 discusses the implications and future directions of this study. In addition, the Appendices present some technical details.

2 Latent Growth Models

The LGCMs can be expressed by a regression equation with latent variables being regressors. Specifically, for a longitudinal study with N subjects and T measurement time points, let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ be a $T \times 1$ random vector, where y_{it} stands for the outcome or observation of individual i on occasion t ($i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$), and let η_i be a $q \times 1$ random vector containing q continuous latent variables. A LGCM for the outcome $\mathbf{y}_i$ related to the latent η_i can be written as

$$\mathbf{y}_i = \Lambda \eta_i + \mathbf{e}_i \tag{1}$$

$$\eta_i = \beta + \xi_i, \tag{2}$$

where Λ is a $T \times q$ matrix consisting of factor loadings, $\mathbf{e}_i$ is a $T \times 1$ vector of residuals or measurement errors that are assumed to follow a T-dimensional multivariate normal distribution, i.e., $\mathbf{e}_i \sim MN_T(\mathbf{0}, \Theta)$, and ξ_i is a $q \times 1$ vector that is assumed to follow a q-dimensional multivariate distribution, i.e., $\xi_i \sim MN_q(\mathbf{0}, \Psi)$. In LGCMs, β is a vector of *fixed effects* and ξ_i is a vector of *random effects* (e.g., Fitzmaurice et al. 2004). The vector β, η_i, and the matrix Λ determine the growth trajectory of the model.

2.1 Selection Models for Non-ignorable Missingness

To address the non-ignorable missingness, there are two general approaches, *pattern-mixture models* (Little and Rubin 1987) and *selection models* (Glynn et al. 1986). In both cases, the statistical analysis requires joint modeling of dependent variable and missing data processes. In this research, selection models are used, mainly because (1) substantively selection models seem more natural for considering the behavior of the response variable in the full target population of interests, rather than in the sub-populations defined by missing data patterns (e.g., Fitzmaurice et al. 2008), and (2) the selection models formulation leads directly to the joint distribution of both dependent variables and the missingness (e.g., Fitzmaurice et al. 2008):

$$p(\mathbf{y}_i, \mathbf{m}_i | \nu, \phi, \mathbf{x}_i) = p(\mathbf{y}_i | \nu, \mathbf{x}_i)\, p(\mathbf{m}_i | \mathbf{y}_i, \nu, \phi, \mathbf{x}_i)$$

where $\mathbf{x}_i$ is a vector of covariates for individual i, $\mathbf{y}_i$ is a vector of individual i's outcome scores, $\theta = (\nu, \phi)$ are all parameters in the model, in which ν are parameters for the growth model and ϕ are parameters for the missingness, and $\mathbf{m}_i$ is a vector $\mathbf{m}_i = (m_{i1}, m_{i2}, \ldots, m_{iT})'$ that indicates the missingness status for $\mathbf{y}_i$. Specifically, if y_i is missing at time point t, then $m_{it} = 1$. Otherwise, $m_{it} = 0$.

Let $\tau_{it} = p(m_{it} = 1)$ be the probability that y_{it} is missing, then m_{it} follows a Bernoulli distribution of τ_{it}, and the density function of m_{it} is

$$p(m_{it}) = \tau_{it}^{m_{it}} (1 - \tau_{it})^{1 - m_{it}}. \tag{3}$$

For different non-ignorable missingness patterns, the expressions of τ_{it} are different. In Lu et al. (2011), τ_{it} is a function of latent class membership and thus the missingness is *latent class dependent (LCD)*. However, the non-ignorable missingness mechanism could be much more complex in reality. For example, the missingness may be related to the latent intercept, the latent slope of growth, or the potential outcome variables. In these cases, the missing data probabilities depend on latent variables, and thus missingness is non-ignorable. We propose three basic non-ignorable missingness models in detail as follows.

(1) Latent Intercept-Dependent (LID) Missingness: This pattern assumes that the missingness depends on individual's latent intercept, or initial level, I_i, and some observed covariates $\mathbf{x}_i$. The rate of missingness τ_{Iit} is expressed as a probit link function of I_i and $\mathbf{x}_i$

$$\tau_{Iit} = \Phi(\gamma_{0t} + I_i \gamma_{It} + \mathbf{x}_i' \gamma_{xt}) = \Phi(\omega_{Ii}' \gamma_{It}), \tag{4}$$

where $\mathbf{x}_i$ is an r-dimensional vector, $\omega_{Ii} = (1, I_i, \mathbf{x}_i')'$ and $\gamma_{It} = (\gamma_{0t}, \gamma_{It}, \gamma_{xt}')'$. Note that if the vector $\gamma_{It} = 0$, then the missingness is ignorable. A path diagram of the LGCM with an LID missingness is illustrated in Fig. 1.

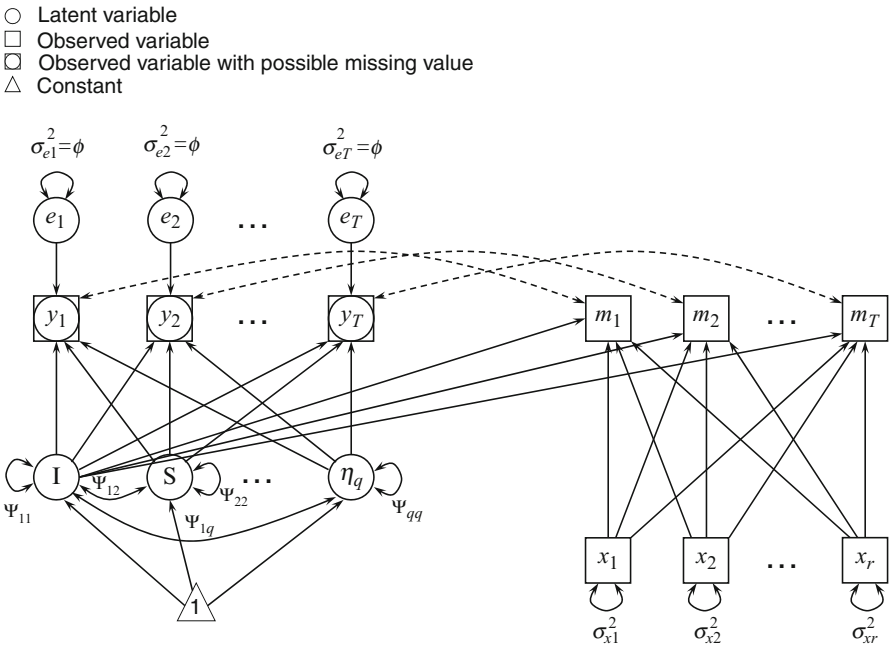

Fig. 1 Path diagram of a latent growth model with latent intercept-dependent missingness (LID), where the rate of missingness $p(m_t)$ depends on covariates x_rs and individual's latent intercept, or initial level, I

(2) Latent Slope-Dependent (LSD) Missingness: This pattern assumes the missingness depends on the latent slope of individuals, S_i. The missing data rate τ_{it} is expressed as a probit link function of S_i and covariates $\mathbf{x}_i$,

$$\tau_{Sit} = \Phi(\gamma_{0t} + S_i\gamma_{St} + \mathbf{x}_i'\gamma_{xt}) = \Phi(\omega_{Si}'\,\gamma_{St}), \tag{5}$$

with $\omega_{Si} = (1, S_i, \mathbf{x}_i')'$ and $\gamma_{St} = (\gamma_{0t}, \gamma_{St}, \gamma_{xt}')'$. Its path diagram is drawn in Fig. 2.

(3) Latent Outcome-Dependent (LOD) Missingness: This pattern assumes that the missing data rates depend on the potential outcomes that may be missing. With covariates $\mathbf{x}_i$, we express τ_{it} as a probit link function as follows.

$$\tau_{yit} = \Phi(\gamma_{0t} + y_{it}\gamma_{yt} + \mathbf{x}_i'\gamma_{xt}) = \Phi(\omega_{yit}'\,\gamma_{yt}), \tag{6}$$

with $\omega_{yit} = (1, y_{it}, \mathbf{x}_i')'$ and $\gamma_{yt} = (\gamma_{0t}, \gamma_{yt}, \gamma_{xt}')'$. The path diagram illustrating the LGCMs with LOD missingness is illustrated in Fig. 3.

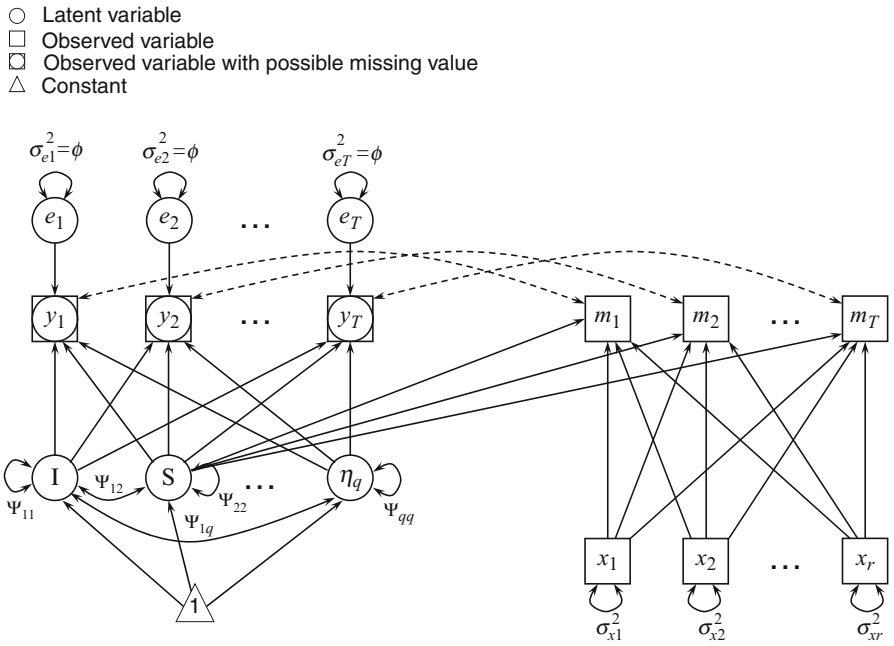

Fig. 2 Path diagram of a latent growth model with latent slope-dependent missing data where $p(m_t)$ depends on covariates x_rs and the latent slope S

3 Bayesian Estimation

In this research, a full Bayesian estimation approach is used to estimate growth models. The algorithm is described as follows. First, model-related latent variables are added via the data augmentation method (Tanner and Wong 1987). By including auxiliary variables, the likelihood function for each model is obtained. Second, proper priors are adopted. Third, with the likelihood function and the priors, based on the Bayes' Theorem, the posterior distribution of the unknown parameters is readily available. We obtain conditional posterior distributions instead of the joint posteriors because the integrations of marginal posterior distributions of the parameters are usually hard to obtain explicitly for high-dimensional data. Fourth, with conditional posterior distributions, Markov chains are generated for the unknown model parameters by implementing a Gibbs sampling algorithm (Geman and Geman 1984). Finally, the statistical inferences are conducted based on converged Markov chains.

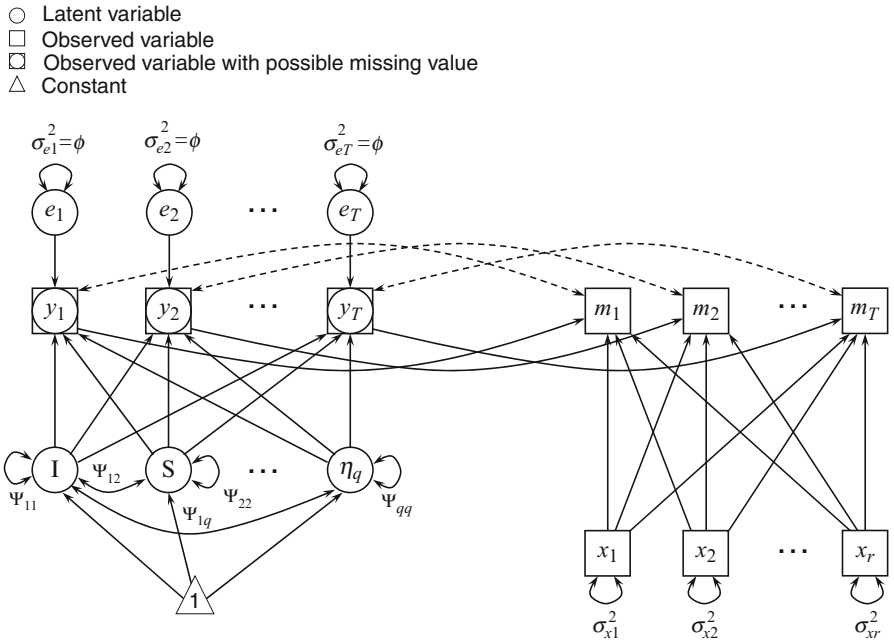

Fig. 3 Path diagram of a latent growth model with potential outcome-dependent missing data where $p(m_t)$ depends on covariates x_rs and the outcome y

3.1 Data Augmentation and Likelihood Functions

In order to construct the likelihood function explicitly, we use the data augmentation algorithm (Tanner and Wong 1987). The observed outcomes $\mathbf{y}_i^{obs}$ can be augmented with the missing values $\mathbf{y}_i^{mis}$ such that $\mathbf{y}_i = (\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis})'$ for individual i. Also, the missing data indicator variable $\mathbf{m}_i$ is added to models. Then the joint likelihood function of the selection model for the ith individual can be expressed as

$$L_i(\eta_i, \mathbf{y}_i, \mathbf{m}_i) = [p(\eta_i)\, p(\mathbf{y}_i|\eta_i)]\; p(\mathbf{m}_i|\mathbf{y}_i, \eta_i, \mathbf{x}_i).$$

For the whole sample, the likelihood function is specifically expressed as

$$\begin{aligned} L(\mathbf{y}, \eta, \mathbf{m}) \propto \prod_{i=1}^{N} \Bigg\{ & |\Psi|^{-1/2} \exp\left[-\frac{1}{2} (\eta_i - \beta)' \Psi^{-1} (\eta_i - \beta) \right] \\ & \times |\phi|^{-T/2} \exp\left[-\frac{1}{2\phi} (\mathbf{y}_i - \Lambda \eta_i)' (\mathbf{y}_i - \Lambda \eta_i) \right] \\ & \times \prod_{t=1}^{T} \left[\tau_{it}^{m_{it}} (1 - \tau_{it})^{1 - m_{it}} \right] \Bigg\}, \end{aligned} \tag{7}$$

where τ_{it} is defined by Eq. (4) for the LID missingness, (5) for the LSD missingness, and (6) for the LOD missingness.

3.2 Priors, Posteriors, and Gibbs Sampling

We assume that all posterior distributions exist in this study. Commonly used proper priors (e.g., Lee 2007) are adopted. Specifically, (1) an inverse Gamma distribution prior is used for $\phi \sim IG(v_0/2, s_0/2)$ where v_0 and s_0 are given hyper-parameters. The density function of an inverse Gamma distribution is $p(\phi) \propto \phi^{-(v_0/2)-1}\exp(-s_0/(2\phi))$. (2) An inverse Wishart distribution prior is used for Ψ. With hyper-parameters m_0 and $\mathbf{V}_0$, $\Psi \sim IW(m_0, \mathbf{V}_0)$, where m_0 is a scalar and $\mathbf{V}_0$ is a $q \times q$ matrix. Its density function is $p(\Psi) \propto |\Psi|^{-(m_0+q+1)/2}\exp[-\mathrm{tr}(\mathbf{V}_0\Psi^{-1})/2]$. (3) For β a multivariate normal prior is used, and $\beta \sim MN_q(\beta_0, \Sigma_0)$ where the hyper-parameter β_0 is a q-dimensional vector and Σ_0 is a $q \times q$ matrix. (4) The prior for γ_t $(t = 1, 2, \ldots, T)$ is chosen to be a multivariate normal distribution $\gamma_t \sim MN_{(2+r)}(\gamma_{t0}, \mathbf{D}_{t0})$, where γ_{t0} is a $(2+r)$-dimensional vector, $\mathbf{D}_{t0}$ is a $(2+r) \times (2+r)$ matrix, and both are pre-determined hyper-parameters.

After constructing the likelihood function and assigning the priors, the joint posterior distribution for unknown parameters is readily available. Considering the high-dimensional integration for marginal distributions of parameters, the conditional distribution for each parameter is obtained instead. The derived conditional posteriors are provided by the equations for parameters in the Appendix. In addition, the conditional posteriors for the latent variable η_i and the augmented missing data $\mathbf{y}_i^{mis}$ $(i = 1, 2, \ldots, N)$ are also provided by their corresponding equations in the Appendix.

After obtaining conditional posteriors, the Markov chain for each model parameter is generated by implementing a Gibbs sampling algorithm (Geman and Geman 1984). Specifically, suppose $\theta = (\theta_1, \theta_2, \ldots, \theta_M)$ is a vector of model parameters, latent variables, and missing values. We start with a set of initial values for θs. At the sth iteration, $\theta^{(s)}$ is generated. To obtain $\theta^{(s+1)}$, each $\theta^{(s+1)}$ is generated from its corresponding posterior distribution, derived in the Appendix, with renewed parameters.

3.3 Statistical Inference

After passing convergence tests, the generated Markov chains can be viewed as from the joint and marginal distributions of all parameters. The statistical inference can then be conducted based on the generated Markov chains.

For different loss functions of θ, the point estimates are different. For example, if a square loss function, $LF = (\theta - \hat{\theta})^2$, is used, then the posterior mean is the estimate of θ; but if an absolute loss function, $LF = |\theta - \hat{\theta}|$, is used, then its estimate

is the posterior median. There are other function forms, such as 0–1 loss function, but in this research we take the square loss function.

Let $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$ denote a vector of all the unknown parameters in the model. Then the converged Markov chains can be recorded as $\theta^{(s)}, s = 1,2,\ldots,S$, and each parameter estimate $\widehat{\theta}_j$ $(j = 1,2,\ldots,p)$ can be calculated as $\widehat{\theta}_j = \sum_{s=1}^{S} \theta_j^{(s)}/S$ with standard error (SE) $s.e.(\widehat{\theta}_j) = \sqrt{\sum_{s=1}^{S}(\theta_j^{(s)} - \widehat{\theta}_j)^2/(S-1)}$. To get the credible (confidence) intervals, both percentile intervals and the highest posterior density intervals (HPD, Box and Tiao 1973) of the Markov chains can be used. Percentile intervals are obtained by sorting $\theta_j^{(s)}$. HPD intervals may also be referred to as minimum length confidence intervals for a Bayesian posterior distribution, and for symmetric distributions HPD intervals obtain equal tail area probabilities.

4 Model Selection Criteria

Model selection criteria play an important role in comparing competing models. In this section, Bayesian model selection criteria are proposed for latent growth models with missing data.

The general mathematical forms of selection criteria are closely related to each other. Almost all of them try to find a balance between the accuracy and the complexity of a model. First, the accuracy of a model can be measured by deviance, which is defined as $D(\theta) = -2log(p(\mathbf{y}|\theta)) + C$ for some constant C. In a Bayesian context, the most popular way to calculate the deviance is to plug the expectation of θ. So we have $D(\hat{\theta}) = -2log(p(\mathbf{y}|E_{\theta|\mathbf{y}}[\theta])) + C$, which can be estimated by $D(\hat{\theta}) \approx -2log(p(\mathbf{y}|\hat{\theta})) + C$. For latent growth models with missing data, $D(\hat{\theta})$ can be calculated as

$$D(\hat{\theta}) = -2\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(1-m_{it})l_{it}(y|\hat{\theta}) + l_{it}(m|\hat{\theta})\right] \tag{8}$$

in which m_{it} is the missing data indicator for individual i at occasion t, $\hat{\theta}$ is the posterior mean of parameter estimates across S converged Markov iterations, and $l_{it}^{(s)}(y)$ and $l_{it}^{(s)}(m)$ are the conditional likelihood functions of y_{it} and m_{it}, respectively, for individual i at occasion t. When y_{it} is missing, $m_{it} = 1$, the likelihood of y_{it} is excluded. When y_{it} are normally distributed, the log-likelihood function is

$$l_{it}(y_N) = -\frac{1}{2}log(2\pi|\phi|) - \frac{(y_{it} - I_i - tS_i)^2}{2\phi}$$

Table 1 Model selection criteria

Criterion(Index) =	Deviance +	Penalty
Dhat.AIC	$D(\hat{\theta})$	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	$\overline{D(\theta)}$	$2(\overline{D(\theta)} - D(\hat{\theta}))$
rough DIC	$\overline{D(\theta)}$	var($D(\theta)$)/2

where I_i and S_i are obtained from the random effect model. For the missing data indicator m_{it}, the log-likelihood function is

$$l_{it}(m) = m_{it} log(\tau_{it}) + (1 - m_{it}) log(1 - \tau_{it}),$$

where τ_{it} varies for different missingness models.

The second part of a criterion is the complexity of a model, which is also called a penalty term. For AIC, the penalty is $2\,p$, where p is the number of model parameters. As the penalty of AIC is sometimes considered to be too lenient in that it selects saturated models in large samples (e.g., Janssen and De Boeck 1999), BIC uses $log(N)p$ as the penalty, where N is the sample size. CAIC is another improved version of AIC. Compared with BIC, CAIC adds an extra p in penalty, which makes CAIC favor smaller models slightly more than BIC. Also, ssBIC improves BIC. The penalty in ssBIC is $log((N+2)/24)\,p$. For DIC, the penalty takes the difference between $E_{\theta|y}[D]$ and $D(E_{\theta|y}[\theta])$, where $E_{\theta|y}[D] = E_{\theta|\mathbf{y}}[-2log(p(\mathbf{y}|\theta))] + C$ is a Monte Carlo estimation of the expectation deviance and can be estimated as the posterior mean across the converged Markov chain,

$$\overline{D(\theta)} = -\frac{2}{S}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{1=t}^{T}\left[(1 - m_{it}) l_{it}^{(s)}(y) + l_{it}^{(s)}(m)\right]. \tag{9}$$

In DIC, $pD = E_{\theta|y}[D] - D(E_{\theta|y}[\theta])$ is a measure of the effective model parameters or the complexity of the model, and it is approximated by $pD = \overline{D(\theta)} - D(\hat{\theta})$. In practice, rough DIC (RDIC, sometimes called DICV in some literature, e.g., Oldmeadow and Keith 2011) is an approximation of formal DIC (e.g., Sturtz et al. 2005). It takes $\overline{D(\theta)}$ as its deviance and $pV = Var(D(\theta))/2$ as its penalty.

In summary, the model selection criteria for latent growth models with missing data in this study are listed in Table 1.

5 Simulation Studies

In this section, simulation studies are conducted to evaluate the performance of the proposed latent growth models and the model selection criteria in a Bayesian context.

5.1 Simulation Design and Implementation

In the simulation we focus on linear LGCMs to simplify the presentation. Higher order LGCMs can be easily expanded by adding quadratic or higher order terms.

First, four waves of complete LGCM data $\mathbf{y}_i$ are generated based on Eqs. (1) and (2). The random effects consist of the intercept I_i and the slope S_i, with $Var(I_i) = 1$, $Var(S_i) = 4$, and $Cov(I_i, S_i) = 0$. The fix-effects are (I,S) $= (1, 3)$. The measurement errors are assumed to follow a normal distribution with mean 0 and standard deviation 1. In the simulation we also assume there is one covariate X generated from a normal distribution, $X \sim N(1, sd = 0.2)$. Missing data are created based on different pre-designed missingness rates. We assume the true missingness is LSD (also noted as the XS missingness in this study because the missingness depends on the latent individual slope S and covariate X). With LSD, the bigger the slope is, the more the missing data. For the sake of simplicity in the simulation, the missingness rate is set the same for every occasion. Specifically, we set the missingness probit coefficients as $\gamma_0 = (-1, -1, -1, -1)$, $\gamma_x = (-1.5, -1.5, -1.5, -1.5)$, and $\gamma_S = (0.5, 0.5, 0.5, 0.5)$. With the setting, missingness rates are generated based on Eq. (5). If a participant has a latent growth slope 3, with a covariate value 1, his or her missingness rate at each wave is $\tau \approx 16\%$; and if the slope is 5, with the same covariate value, the missing rate increases to $\tau \approx 50\%$; but when the slope is 1, the missingness rate decreases to $\tau \approx 2.3\%$.

Next, we fit data with LGCMs with different missingness. Specifically, the model design with different missingness is shown in Table 2, where the symbol "✓" shows the related factors on which the missing data rates depend. For example, when both "X" and "I" are checked, the missingness depends on the individual's latent intercept "I" and the observed covariate "X." Four types of missingness are studied: LID (also noted as XI in Table 2), LSD (XS), LOD (XY), and ignorable (X). The shaded model, LSD (XS), is the true model we used for generating the simulation data. Five levels of sample size (N = 1,000, N = 500, N = 300, N = 200 and N = 100) are investigated, and for each sample size, 100 converged replications are analyzed and summarized.

The simulation studies are implemented by the following algorithm. (1) Set the counter $R = 0$. (2) Generate complete longitudinal growth data according to predefined model parameters. (3) Create missing data according to missing data mechanisms and missing data rates. (4) Generate Markov chains for model parameters through the Gibbs sampling procedure. (5) Test the convergence of

Table 2 Model design in the simulation study

LGCM: N = 1000, 500, 300, 200 and 100				
	Missingness			
Model	X[2]	I[3]	S[4]	Y[5]
Ignorable (X)	√			
LID (XI)	√	√		
LSD (XS)[1]	√		√	
LOD (XY)	√			√

[1] The shaded model is the true model.
[2] Observed covariates.
[3] Individual latent intercept. If checked, the missingness is non-ignorable.
[4] Individual latent slope. If checked, the missingness is non-ignorable.
[5] Individual potential outcome y. If checked, the missingness is non-ignorable.

generated Markov chains. (6) If the Markov chains pass the convergence test, set $R = R + 1$ and calculate and save the parameter estimates. Otherwise, set $R = R$ and discard the current replication of simulation. (7) Repeat the above process till $R = 100$ to obtain 100 replications of valid simulation.

In step 4, priors carrying little prior information are adopted (Zhang et al. 2007). Specifically, for φ_1, we set $\mu_{\varphi_1} = \mathbf{0}_2$ and $\Sigma_{\varphi_1} = 10^3\mathbf{I}_2$. For ϕ, we set $v_{0k} = s_{0k} = 0.002$. For β, it is assumed that $\beta_{k0} = \mathbf{0}_2$ and $\Sigma_{k0} = 10^3\mathbf{I}_2$. For Ψ, we define $m_{k0} = 2$ and $\mathbf{V}_{k0} = \mathbf{I}_2$. Finally, for γ_t, we let $\gamma_{t0} = \mathbf{0}_3$ and $\mathbf{D}_{t0} = 10^3\mathbf{I}_3$, where $\mathbf{0}_d$ and $\mathbf{I}_d$ denote a d-dimensional zero vector and a d-dimensional identity matrix, respectively. In step 5, the iteration number of burn-in period is set. The Geweke convergence criterion indicated that less than 10,000 iterations were adequate for all conditions in the study. Therefore, a conservative burn-in of 20,000 iterations was used for all iterations. And then the Markov chains with a length of 20,000 iterations are saved for convergence testing and data analysis. After step 7, twelve summary statistics are reported based on 100 sets of converged simulation replications. For the purpose of presentation, let θ_j represent the jth parameter, also the true value in the simulation. The twelve statistics are defined below. (1) The average estimate (est.$_j$) across 100 converged simulation replications of each parameter is obtained as $\text{est.}_j = \bar{\hat{\theta}}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}/100$, where $\hat{\theta}_{ij}$ denotes the estimate of θ_j in the ith simulation replication. (2) The simple bias (BIAS.smp$_j$) of each parameter is calculated as $\text{BIAS.smp}_j = \bar{\hat{\theta}}_j - \theta_j$. (3) The relative bias (BIAS.rel$_j$) of each parameter is calculated using $\text{BIAS.rel}_j = (\bar{\hat{\theta}}_j - \theta_j)/\theta_j$ when $\theta_j \neq 0$ and $\text{BIAS.rel}_j = \bar{\hat{\theta}}_j - \theta_j$ when $\theta_j = 0$. (4) The empirical standard error (SE.emp$_j$) of each parameter is obtained as $\text{SE.emp}_j = \sqrt{\sum_{i=1}^{100}(\hat{\theta}_{ij} - \bar{\hat{\theta}}_j)^2/99}$. (5) The average standard error (SE.avg$_j$) is calculated by $\text{SE.avg}_j = \sum_{i=1}^{100} \hat{s}_{ij}/100$, where $\hat{s}_{ij}$ denotes the estimated standard error of $\hat{\theta}_{ij}$. (6) The average mean square error (MSE)

of each parameter is obtained 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, $\text{MSE}_{ij} = (\text{Bias}_{ij})^2 + (\hat{s}_{ij})^2$. (7) The average lower and (8) upper limits of the 95% percentile confidence interval (CI.low_j and CI.upper_j) are, respectively, defined as $\text{CI.low}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}^{l}/100$, and $\text{CI.upper}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}^{u}/100$ where $\hat{\theta}_{ij}^{l}$ and $\hat{\theta}_{ij}^{u}$ denote the 95% lower and upper limits of CI for the jth parameter, respectively. (9) The coverage probability of the 95% percentile confidence interval (CI.cover_j) of each parameter is obtained using $\text{CI.cover}_j = [\#(\hat{\theta}_{ij}^{l} \le \theta_j \le \hat{\theta}_{ij}^{u})]/100$. (10) The average lower, (11) upper limits, and (12) the coverage probability of the 95% highest posterior density credible interval (HPD, Box and Tiao 1973) of each parameter are similarly defined by HPD.low_j, HPD.upper_j, and HPD.cover_j, respectively.

5.2 Simulation Results

In this section, we show simulation results for the estimates obtained from the true model and mis-specified models, and the performance of model selection criteria.

First, we investigate the estimates obtained from the true model. Tables 3 and 4 show the summarized estimates for different sample sizes (N = 1,000, N = 500, N = 300, N = 200, and N = 100). From both tables, except for the small sample size N = 100, one can see that (1) all the estimate biases are very small; (2) the difference between the empirical SEs and the average SEs is very small, which indicates the SEs are estimated accurately; (3) both percentile interval and HPD interval coverage probabilities are very close to the theoretical percentage 95%, which means the type I error for each parameter is close to the specified 5% so that we can use the estimated confidence intervals to conduct statistical inference; and (4) this true model has 100% convergence rate.

In order to conveniently compare estimates for different sample sizes, we further summarize Tables 3 and 4 by calculating five summary statistics across all parameters, which are shown in Table 5. The first statistic is the average absolute relative biases (|Bias.rel|) across all parameters, which is defined as $|\text{Bias.rel}| = \sum_{j=1}^{p} |\text{Bias.rel}_j|/p$, where p is the total number of parameters in a model. Second, we obtain the average absolute differences between the empirical SEs and the average Bayesian SEs (|SE.diff|) across all parameters by using $|\text{SE.diff}| = \sum_{j=1}^{p} |\text{SE.emp}_j - \text{SE.avg}_j|/p$. Third, we calculate the average percentile coverage probabilities (CI.cover) across all parameters by using $\text{CI.cover} = \sum_{j=1}^{p} \text{CI.cover}_j/p$. Fourth, we calculate the average HPD coverage probabilities (HPD.cover) across all parameters by using $\text{HPD.cover} = \sum_{j=1}^{p} \text{HPD.cover}_j/p$. Fifth, the convergence rate for the study is calculated.

Table 5 shows that, except for the case for N = 100, the true mode can recover model parameters very well, by checking (1) the small average absolute relative biases of estimates, |Bias.rel|, (2) the small average absolute differences between the empirical SEs and the average SEs, |SE.diff|, and (3) the almost 95%

Table 3 Summarized estimates of the true models with LSD (XS) missingness

					BIAS		SE			95% percentile CI			95% HPD interval		
		Parameter	True	est.[a]	smp.[b]	rel.[c]	emp.[d]	avg.[e]	MSE[f]	Lower	Upper	Cover	Lower	Upper	Cover
N = 1,000, Summarized based on 100 converged replications with a convergence rate of 100/100 = 100%.															
Growth curve		I	1	0.998	−0.002	−0.002	0.05	0.053	0.005	0.894	1.101	0.99	0.894	1.101	0.98
		S	3	3.003	0.003	0.001	0.079	0.077	0.012	2.853	3.155	0.97	2.853	3.154	0.96
		var(I)	1	1.011	0.011	0.011	0.105	0.102	0.022	0.82	1.22	0.94	0.814	1.213	0.94
		var(S)	4	3.99	−0.01	−0.003	0.232	0.232	0.107	3.56	4.468	0.94	3.545	4.449	0.93
		cov(IS)	0	0.001	0.001	0.001	0.119	0.112	0.026	−0.221	0.217	0.94	−0.218	0.218	0.94
		var(e)	1	1	0	0	0.043	0.042	0.004	0.92	1.086	0.92	0.918	1.084	0.93
Missingness parameters	Wave 1	γ_{01}	−1	−1.025	−0.025	0.025	0.184	0.174	0.065	−1.375	−0.694	0.93	−1.365	−0.69	0.94
		γ_{x1}	−1.5	−1.541	−0.041	0.027	0.138	0.123	0.036	−1.795	−1.314	0.92	−1.783	−1.307	0.93
		γ_{S1}	0.5	0.515	0.015	0.03	0.066	0.062	0.008	0.4	0.641	0.9	0.397	0.636	0.92
	Wave 2	γ_{02}	−1	−1.038	−0.038	0.038	0.191	0.171	0.067	−1.385	−0.714	0.96	−1.376	−0.711	0.97
		γ_{x2}	−1.5	−1.551	−0.051	0.034	0.129	0.119	0.034	−1.798	−1.33	0.95	−1.786	−1.323	0.94
		γ_{S2}	0.5	0.521	0.021	0.042	0.066	0.06	0.008	0.41	0.643	0.95	0.408	0.639	0.94
	Wave 3	γ_{03}	−1	−1.067	−0.067	0.067	0.186	0.172	0.069	−1.417	−0.741	0.94	−1.407	−0.737	0.94
		γ_{x3}	−1.5	−1.557	−0.057	0.038	0.117	0.116	0.03	−1.796	−1.341	0.97	−1.785	−1.334	0.97
		γ_{S3}	0.5	0.529	0.029	0.058	0.063	0.058	0.008	0.42	0.648	0.89	0.418	0.643	0.91
	Wave 4	γ_{04}	−1	−1.034	−0.034	0.034	0.18	0.173	0.063	−1.384	−0.709	0.94	−1.374	−0.704	0.93
		γ_{x4}	−1.5	−1.539	−0.039	0.026	0.122	0.114	0.029	−1.773	−1.325	0.95	−1.763	−1.319	0.94
		γ_{S4}	0.5	0.514	0.014	0.027	0.058	0.057	0.007	0.407	0.63	0.95	0.405	0.625	0.95

(continued)

Table 3 (continued)

N = 500, Summarized based on 100 converged replications with a convergence rate of 100/100 = 100%.														
Growth curve		I	1	0.986	−0.014	−0.014	0.076	0.074	0.011	0.841	1.132	0.93	0.841	1.132	0.95
		S	3	3.001	0.001	0	0.097	0.109	0.021	2.789	3.216	0.97	2.788	3.213	0.97
		var(I)	1	0.976	−0.024	−0.024	0.146	0.144	0.042	0.712	1.274	0.97	0.7	1.26	0.97
		var(S)	4	4.001	0.001	0	0.388	0.329	0.258	3.403	4.691	0.9	3.373	4.652	0.9
		cov(IS)	0	−0.009	−0.009	−0.009	0.155	0.157	0.049	−0.324	0.294	0.96	−0.319	0.297	0.96
		var(e)	1	1.014	0.014	0.014	0.06	0.061	0.007	0.901	1.141	0.96	0.897	1.136	0.96
Missingness parameters	Wave 1	γ_{01}	−1	−1.082	−0.082	0.082	0.254	0.255	0.137	−1.609	−0.608	0.95	−1.587	−0.596	0.97
		γ_{x1}	−1.5	−1.606	−0.106	0.071	0.181	0.186	0.079	−2.002	−1.275	0.95	−1.975	−1.258	0.97
		γ_{S1}	0.5	0.54	0.04	0.081	0.083	0.092	0.017	0.375	0.735	0.95	0.368	0.722	0.94
	Wave 2	γ_{02}	−1	−1.096	−0.096	0.096	0.281	0.252	0.152	−1.61	−0.624	0.89	−1.591	−0.615	0.89
		γ_{x2}	−1.5	−1.615	−0.115	0.077	0.204	0.18	0.088	−1.996	−1.291	0.91	−1.971	−1.275	0.94
		γ_{S2}	0.5	0.546	0.046	0.092	0.104	0.088	0.021	0.385	0.73	0.87	0.379	0.719	0.88
	Wave 3	γ_{03}	−1	−1.068	−0.068	0.068	0.32	0.248	0.169	−1.572	−0.602	0.93	−1.555	−0.594	0.93
		γ_{x3}	−1.5	−1.613	−0.113	0.075	0.279	0.174	0.123	−1.978	−1.295	0.9	−1.958	−1.283	0.93
		γ_{S3}	0.5	0.536	0.036	0.072	0.116	0.084	0.022	0.381	0.71	0.92	0.378	0.702	0.91
	Wave 4	γ_{04}	−1	−1.123	−0.123	0.123	0.261	0.257	0.15	−1.652	−0.647	0.94	−1.628	−0.633	0.95
		γ_{x4}	−1.5	−1.579	−0.079	0.053	0.174	0.168	0.066	−1.933	−1.274	0.95	−1.913	−1.261	0.96
		γ_{S4}	0.5	0.543	0.043	0.086	0.089	0.085	0.017	0.388	0.719	0.92	0.382	0.71	0.92
N = 300, Summarized based on 100 converged replications with a convergence rate of 100/100 = 100%.															
Growth curve		I	1	1.001	0.001	0.001	0.104	0.097	0.02	0.81	1.192	0.89	0.811	1.192	0.89
		S	3	2.984	−0.016	−0.005	0.149	0.14	0.042	2.712	3.262	0.93	2.71	3.259	0.93
		var(I)	1	1.014	0.014	0.014	0.183	0.19	0.07	0.673	1.418	0.96	0.654	1.392	0.96
		var(S)	4	3.975	−0.025	−0.006	0.416	0.425	0.354	3.22	4.886	0.96	3.174	4.82	0.96
		cov(IS)	0	0.054	0.054	0.054	0.212	0.205	0.09	−0.359	0.449	0.94	−0.351	0.454	0.93
		var(e)	1	1.011	0.011	0.011	0.073	0.08	0.012	0.867	1.179	0.96	0.86	1.17	0.96

Missingness parameters	Wave 1	γ_{01}	−1	−1.094	−0.094	0.094	0.341	0.345	0.249	−1.822	−0.468	0.97	−1.778	−0.441	0.97
		γ_{x1}	−1.5	−1.65	−0.15	0.1	0.265	0.253	0.162	−2.209	−1.217	0.92	−2.155	−1.185	0.94
		γ_{s1}	0.5	0.548	0.048	0.097	0.121	0.124	0.033	0.331	0.82	0.97	0.318	0.794	0.97
	Wave 2	γ_{02}	−1	−1.106	−0.106	0.106	0.452	0.34	0.341	−1.819	−0.486	0.93	−1.782	−0.467	0.93
		γ_{x2}	−1.5	−1.692	−0.192	0.128	0.345	0.253	0.23	−2.243	−1.254	0.89	−2.196	−1.227	0.90
		γ_{s2}	0.5	0.566	0.066	0.132	0.158	0.121	0.046	0.354	0.827	0.93	0.343	0.807	0.92
	Wave 3	γ_{03}	−1	−1.139	−0.139	0.139	0.397	0.335	0.293	−1.845	−0.527	0.91	−1.801	−0.503	0.92
		γ_{x3}	−1.5	−1.648	−0.148	0.099	0.305	0.236	0.175	−2.152	−1.233	0.86	−2.115	−1.21	0.92
		γ_{s3}	0.5	0.566	0.066	0.132	0.141	0.115	0.038	0.361	0.811	0.9	0.352	0.794	0.91
	Wave 4	γ_{04}	−1	−1.217	−0.217	0.217	0.411	0.356	0.347	−1.976	−0.576	0.9	−1.932	−0.552	0.90
		γ_{x4}	−1.5	−1.681	−0.181	0.121	0.263	0.241	0.163	−2.203	−1.257	0.9	−2.161	−1.231	0.92
		γ_{s4}	0.5	0.583	0.083	0.165	0.138	0.118	0.041	0.372	0.839	0.88	0.363	0.82	0.91

[a] The parameter estimate, defined by $\text{est.}_j = \bar{\hat{\theta}}_j = \sum_{i=1}^{100} \hat{\theta}_{ij}/100$

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

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

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

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

[f] The mean square error, defined by $\text{MSE}_j = \sum_{i=1}^{100} \text{MSE}_{ij}/100$, where $\text{MSE}_{ij} = (\text{Bias}_{ij})^2 + (\hat{s}_{ij})^2$

Table 4 Summarized estimates of the true models with LSD (XS) missingness (cont'd)

					BIAS		SE			95% percentile CI			95% HPD interval		
		Parameter	True	est.	smp.	rel.	emp.	avg.	MSE	Lower	Upper	Cover	Lower	Upper	Cover
N = 200, Summarized based on 100 converged replications with a convergence rate of 100/106 = 94.34%.															
Growth curve		I	1	1.011	0.011	0.011	0.099	0.119	0.024	0.779	1.244	0.98	0.779	1.243	0.98
		S	3	2.975	−0.025	−0.008	0.177	0.171	0.061	2.643	3.314	0.93	2.642	3.312	0.94
		var(I)	1	1.011	0.011	0.011	0.228	0.233	0.107	0.601	1.516	0.94	0.572	1.476	0.92
		var(S)	4	4	0	0	0.474	0.522	0.498	3.095	5.135	0.97	3.029	5.041	0.96
		cov(IS)	0	0.065	0.065	0.065	0.257	0.252	0.134	−0.447	0.549	0.92	−0.436	0.557	0.92
		var(e)	1	1.027	0.027	0.027	0.098	0.099	0.02	0.851	1.238	0.95	0.84	1.224	0.95
Missingness parameters	Wave 1	γ_{01}	−1	−1.3	−0.3	0.3	0.671	0.5	0.901	−2.399	−0.449	0.93	−2.306	−0.402	0.94
		γ_{x1}	−1.5	−1.874	−0.374	0.249	0.745	0.424	1.113	−2.868	−1.227	0.88	−2.735	−1.169	0.91
		γ_{S1}	0.5	0.647	0.147	0.293	0.323	0.197	0.202	0.334	1.1	0.91	0.311	1.045	0.92
	Wave 2	γ_{02}	−1	−1.278	−0.278	0.278	0.69	0.468	0.838	−2.303	−0.463	0.87	−2.227	−0.426	0.89
		γ_{x2}	−1.5	−1.779	−0.279	0.186	0.456	0.349	0.451	−2.578	−1.209	0.91	−2.487	−1.163	0.9
		γ_{S2}	0.5	0.627	0.127	0.254	0.244	0.171	0.117	0.343	1.014	0.9	0.324	0.976	0.91
	Wave 3	γ_{03}	−1	−1.191	−0.191	0.191	0.505	0.436	0.5	−2.133	−0.419	0.91	−2.05	−0.377	0.93
		γ_{x3}	−1.5	−1.721	−0.221	0.147	0.502	0.314	0.426	−2.428	−1.193	0.9	−2.348	−1.15	0.94
		γ_{S3}	0.5	0.586	0.086	0.172	0.183	0.152	0.068	0.326	0.926	0.91	0.309	0.889	0.95
	Wave 4	γ_{04}	−1	−1.27	−0.27	0.27	0.594	0.467	0.67	−2.304	−0.457	0.86	−2.209	−0.404	0.90
		γ_{x4}	−1.5	−1.808	−0.308	0.205	0.397	0.336	0.382	−2.56	−1.24	0.82	−2.48	−1.195	0.89
		γ_{S4}	0.5	0.618	0.118	0.236	0.204	0.16	0.085	0.345	0.98	0.88	0.325	0.942	0.89

N = 100, Summarized based on 100 converged replications with a convergence rate of 100/142 = 70.42%.															
Growth curve		I	1	1.031	0.031	0.031	0.167	0.168	0.057	0.701	1.359	0.96	0.701	1.359	0.97
		S	3	2.983	−0.017	−0.006	0.236	0.242	0.115	2.514	3.467	0.95	2.51	3.46	0.94
		var(I)	1	0.933	−0.067	−0.067	0.305	0.323	0.206	0.408	1.665	0.93	0.355	1.574	0.91
		var(S)	4	3.965	−0.035	−0.009	0.829	0.747	1.261	2.743	5.656	0.91	2.623	5.458	0.91
		cov(IS)	0	0.069	0.069	0.069	0.333	0.357	0.246	−0.666	0.748	0.93	−0.646	0.762	0.95
		var(e)	1	1.078	0.078	0.078	0.157	0.151	0.054	0.82	1.409	0.93	0.801	1.38	0.94
Missingness parameters	Wave 1	γ_{01}	−1	−3.257	−2.257	2.257	5.794	1.333	42.792	−6.264	−1.131	0.84	−5.922	−1.018	0.86
		γ_{x1}	−1.5	−4.314	−2.814	1.876	7.492	1.277	69.337	−7.171	−2.396	0.8	−6.739	−2.251	0.85
		γ_{s1}	0.5	1.626	1.126	2.252	2.881	0.55	10.353	0.788	2.857	0.8	0.746	2.698	0.84
	Wave 2	γ_{02}	−1	−3.011	−2.011	2.011	5.719	1.322	41.711	−6.062	−1.027	0.85	−5.696	−0.893	0.88
		γ_{x2}	−1.5	−3.772	−2.272	1.515	6.947	1.283	61.237	−6.811	−1.927	0.82	−6.385	−1.774	0.85
		γ_{s2}	0.5	1.436	0.936	1.871	2.57	0.549	8.564	0.653	2.71	0.81	0.586	2.527	0.86
	Wave 3	γ_{03}	−1	−2.877	−1.877	1.877	5.93	1.2	42.401	−5.493	−0.898	0.89	−5.233	−0.806	0.91
		γ_{x3}	−1.5	−3.86	−2.36	1.573	6.955	1.153	58.835	−6.508	−2.086	0.83	−6.125	−1.932	0.85
		γ_{s3}	0.5	1.388	0.888	1.776	2.567	0.467	7.977	0.641	2.428	0.85	0.596	2.289	0.89
	Wave 4	γ_{04}	−1	−2.831	−1.831	1.831	5.646	1.297	39.835	−5.902	−0.891	0.89	−5.522	−0.753	0.90
		γ_{x4}	−1.5	−3.386	−1.886	1.257	5.379	1.127	37.532	−6.048	−1.745	0.81	−5.622	−1.586	0.88
		γ_{s4}	0.5	1.222	0.722	1.444	1.944	0.457	4.854	0.552	2.312	0.84	0.491	2.152	0.88

Note: Abbreviations are as given in Table 3

Table 5 Summary and comparison of simulation results of the true model

		|Bias.rel|[a]	|SE.diff|[b]	MSE[c]	CI.cover[d]	HPD.cover[e]	CVG.rate[f] (%)
N	1,000	0.025	0.007	0.033	0.942	0.942	100
	500	0.052	0.021	0.079	0.932	0.939	100
	300	0.089	0.031	0.150	0.922	0.930	100
	200	0.160	0.090	0.366	0.909	0.924	94.34
	100	1.202	2.664	23.743	0.869	0.893	70.42

[a]The average absolute relative bias across all 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 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 parameters, defined by $MSE = \sum_{j=1}^{p}[(\text{Bias}_j)^2 + (\hat{s}_j)^2]/p$. The smaller, the better

[d]The average percentile coverage probability across all 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 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

average percentile coverage probabilities, CI.cover, and the average HPD coverage probabilities, HPD.cover. With the increase of the sample size, both the point estimates and standard errors get more accurate.

Second, we compare the estimates obtained from the true model and different mis-specified models. In this study the true model is the LGCM with LSD (XS) missingness, and there are three mis-specified models, the LGCM with LID (XI) missingness, the LGCM with LOD (XY) missingness, and the LGCM with ignorable missingness (see Table 2 for simulation design). The estimates from the mis-specified models, such as LID (XI) missingness, LOD (XY) missingness, and ignorable missingness, are also summarized, but not included in this paper due to limit space.

To compare estimates from different models, we further summarize and visualize some statistics. Figure 4a compares the point estimates of intercept and slope for all models when N = 1,000. The true value of slope is 3 but the estimate is 2.711 when the missingness is ignored. Actually, for the model with ignorable missingness, the slope estimates are all less than 2.711 for all sample sizes in our study. Figure 4b focuses on the coverage of slope. When the missingness is ignored, it is as low as 4% for N = 1,000, and 21% for N = 500 (the coverage for N = 1,000 is lower because the SE for N = 1,000 is smaller than the SE for N = 500). As a result, conclusions based on the model with ignorable missingness will be severely misleading. Figure 4b also shows that the slope estimate from the model with the mis-specified missingness, LID (XI), has low coverage, with 76% for N = 1,000 and 87% for N = 500. So the conclusions based on this model may still be incorrect. Figure 4c compares the true model and the model with another type of mis-specified missingness, LOD (XY) for N = 1,000. For the wrong model, the coverage is 51%

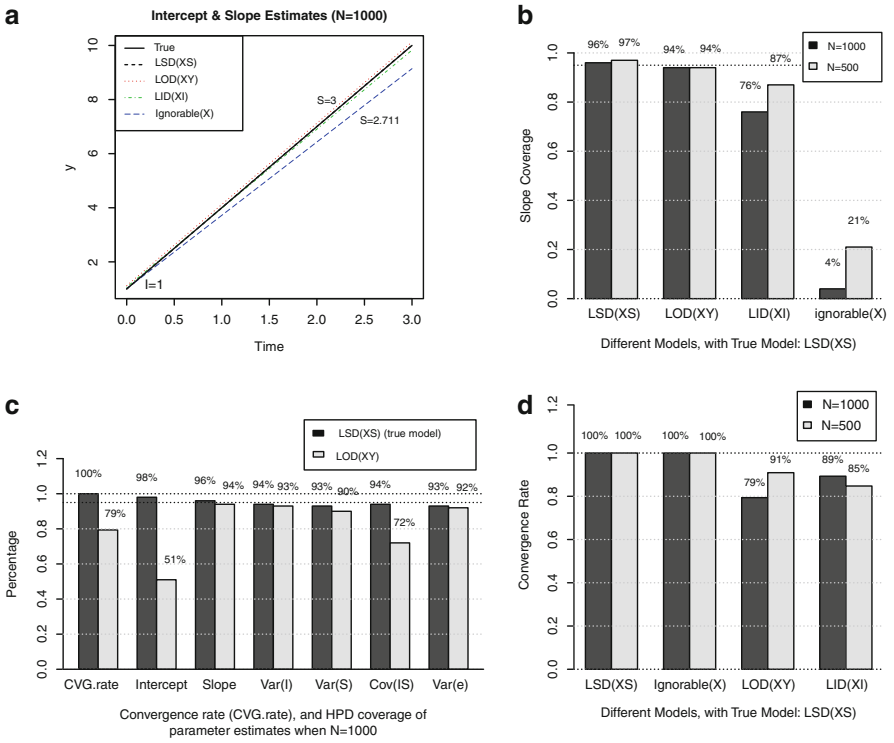

Fig. 4 Comparison of four models/missingness mechanisms. (**a**) Intercept and slope estimates for all models (True Int=1, True Slope=3), (**b**) Slope coverage for all models (Theoretical coverage=95%), (**c**) Parameter coverage for LSD(XS) and LOD (XY) (Theoretical value=95%), and (**d**) Convergence rates for all models (The closer to 100%, the better)

for intercept, and 72% for Cov(I,S). Finally, Fig. 4d compares the convergence rates for all models. One can see that the convergence rates of LOD (XY) and LID (XI) models are much lower than those of the true model LSD (XS) and the model with ignorable missingness. When the missingness is ignored, the number of parameters is smaller than that of non-ignorable models, and then convergence rate gets higher.

In summary, the estimates from mis-specified models may result in severely misleading conclusions, especially when the missingness is ignored. Also, the convergence rate of a mis-specified model is usually lower than that of the true model.

Third, regarding model selection, Table 6 lists the selection proportions across all replications. It shows that almost all the criteria, except for the rough DIC, can correctly identify the true model with high certainty.

Table 6 Model selection proportion

Criteron[1]	Non-ignorable missingness			ignorable missingness	Non-ignorable missingness			ignorable missingness
	LSD (XS) [2]	LOD (XY)	LID (XI)		LSD (XS)	LOD (XY)	LID (XI)	
	N = 1000				N = 500			
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
Rough DIC	0.013	0.000	0.987	0.000	0.038	0.000	0.962	0.000
	N = 300				N = 200			
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
Rough DIC	0.1125	0.000	0.8875	0.000	0.2	0.03125	0.76875	0.000
	N = 100							
Dhat.AIC	0.7125	0.28125	0.00625	0.000				
Dhat.BIC	0.7125	0.28125	0.00625	0.000				
Dhat.CAIC	0.7125	0.28125	0.00625	0.000				
Dhat.ssBIC	70625	0.28125	0.00625	0.000				
DIC	0.70625	0.175	0.11875	0.000				
Rough DIC	0.1125	0.04375	0.84375	0.000				

[1] The definition of each criterion is given in Table 1.

[2] The shaded model is the true model.

[3] The shaded cell has the largest proportion. For each criterion, the sum of all proportions might be larger than 1 because models may have the same lowest index value.

5.3 Simulation Conclusions

Based on the simulation studies, we conclude as follows. (1) The proposed Bayesian method can accurately recover model parameters (both point estimates and standard errors). (2) The small difference between the empirical SE and the average SE indicates that the Bayesian method used in the study can estimate the standard errors accurately. (3) With the increase of the sample size, estimates get closer to their true values and standard errors become more accurate. (4) Ignoring the non-ignorable missingness can lead to severely incorrect conclusions. (5) Mis-specified missingness may also result in misleading conclusions. (6) Almost all the criteria, except for the rough DIC, can correctly identify the true model with high certainty. (7) The non-convergent model might be a sign of a wrong model.

6 Real Data Analysis

In this section, we illustrate the application of the Bayesian latent growth curve model with missing data through the analysis of mathematical ability growth data from the NLSY97 survey (Bureau of Labor Statistics, U.S. Department of Labor 1997). The data set available to us consisted of $N = 362$ youths who were administered the Peabody Individual Achievement Test (PIAT) Mathematics Assessment yearly from 1997, when they were 12 years old and in Grade 7, to 2000, when they were 15 years old and in Grade 10. Figure 5 plots the data, which shows the four measures of mathematical ability increased over time with a roughly linear trend.

Table 7 presents the summary statistics. The missing data rates range from 5.801% to 12.707%. Information on mothers' education (in years) was also included in the sample. In this analysis, we are interested to see how mathematical ability grew over the 4-year period, and if mothers' education influenced the missing data pattern.

First, for comparison purposes, we fit four models with different types of missingness, LSD, LID, LOD, and ignorable. For each model, the burn-in period for Gibbs sampling was generated long enough to make sure Markov chains for parameters converged. To test convergence, the history plot and Geweke test statistic

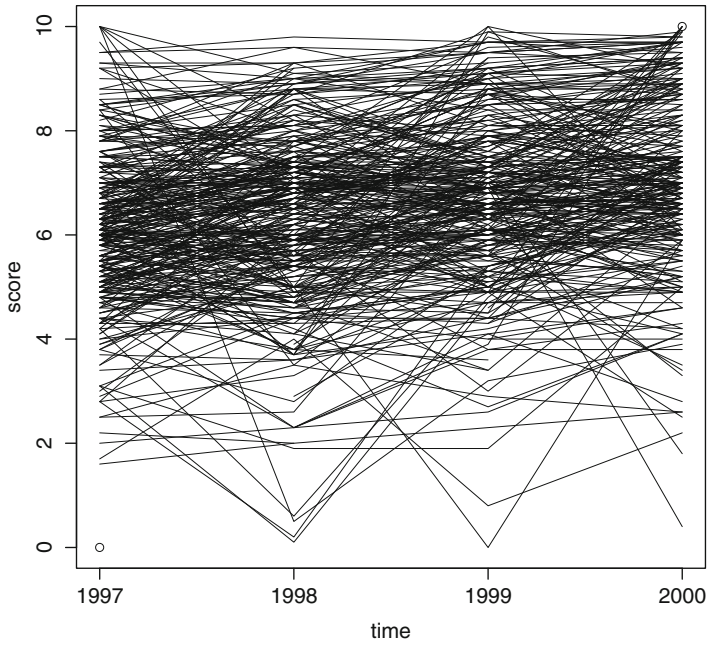

Fig. 5 Plot for the PIAT math data

Table 7 Summary statistics for the PIAT math data

	1997	1998	1999	2000
Mean	6.110	6.309	6.722	6.959
Standard deviation	1.560	1.698	1.679	1.770
Missing data (count)	22	21	39	46
Missing data rate (%)	6.077	5.801	10.773	12.707

for each unknown model parameter were examined. Except for the LID model, all the other three models converged. Table 8 shows the Geweke test statistics for all the model parameters are smaller than 1.96, which indicates the convergence of Markov chains (Geweke 1992). The next 90,000 iterations are then saved for data analysis. The results of the three models are provided in Table 8. In the table, the ratio of Monte Carlo error (MC error) and standard deviation (SD) for each parameter is around or smaller than 0.05, which indicates parameter estimates are accurate (Spiegelhalter et al. 2003). MC error is an important statistic providing a measure of the variability of each parameter estimate in the MCMC chain. The lower the MC error, the more precise the parameter estimate. Overall, we conclude that the results from the real data analysis can be used for further inference. A quick look at the results from the three models shows that the growth parameters do not differ much, even for the model with ignorable missingness. This is due to the low missing data rates for our data set. However, for missingness parameters, different missingness models have different results which, in turn, leads to different interpretations of the data.

Second, the model selection criteria were used to identify the best-fit model. Table 9 shows all the available indices. As one can see, the LSD model is favored by all the criteria. The results from the best-fit model, LSD, reveal that (1) none of γ_{xt}s, the coefficients for the covariate, are significant at the α level of 0.05, which implies that the missingness is not related to mothers' education level; however, (2) the missingness is significantly negatively correlated with the latent slope in 1999 and 2000, which implies that in these 2 years the youth with a low mathematical growth is more likely to miss a test.

7 Discussion

Latent growth curve models are becoming increasingly complex and with this comes an increase in concerns about estimating these models. In this study, we examined several growth models designed to address problems common to almost all longitudinal research, namely, that of missing data. Three new non-ignorable missingness mechanisms were considered: latent intercept missingness, latent slope missingness, and outcome-dependent missingness. A fully Bayesian approach was implemented using data augmentation and Gibbs sampling to estimate these models in the presence of the three types of non-ignorable missingness. Simulation results

Table 8 Estimates from different models in real data analysis

			Mean	S.D.[1]	MCs.e./S.D.[2]	Lower[2.5%]	Upper[97.5%]	Geweke t[3]
			Model with LSD (XS) missingness [4]					
Growth Curve	Parameters	Intercept	6.060 [5]	0.083	0.001	5.895	6.223	−0.718
		Slope	0.288	0.030	$7.3E-4$	0.230	0.348	0.170
		Var(I)	1.697	0.171	0.002	1.387	2.057	0.928
		Var(S)	0.078	0.020	$7.4E-4$	0.046	0.121	−1.280
		Cov(I,S)	−0.039	0.038	$8.7E-4$	−0.120	0.031	−0.199
		Var(e)	1.011	0.054	0.001	0.909	1.121	1.734
Missingness Parameters	1997	γ_{01}	−2.574	0.625	0.033	−3.847	−1.450	−1.448
		γ_{x1}	0.081	0.046	0.002	−0.004	0.175	1.512
		γ_{S1}	−0.089	0.840	0.030	−1.797	1.550	0.205
	1998	γ_{02}	−1.656	0.516	0.025	−2.681	−0.636	−0.162
		γ_{x2}	0.022	0.039	0.002	−0.054	0.103	0.313
		γ_{S2}	−0.926	0.796	0.025	−2.613	0.526	−0.989
	1999	γ_{03}	−1.710	0.695	0.039	−3.164	−0.407	1.387
		γ_{x3}	0.083	0.054	0.003	−0.020	0.195	−1.146
		γ_{S3}	−4.332	2.878	0.073	−12.457	−1.170	−1.201
	2000	γ_{04}	−0.875	0.482	0.025	−1.823	0.021	0.484
		γ_{x4}	0.009	0.037	0.002	−0.061	0.085	−0.230
		γ_{S4}	−1.838	0.920	0.032	−3.967	−0.319	−1.258
			Model with LOD (XY) missingness					
Growth Curve	Parameters	Intercept	6.002	0.084	$8.8E-4$	5.838	6.167	0.315
		Slope	0.333	0.032	$6.0E-4$	0.271	0.396	0.205
		Var(I)	1.738	0.187	0.002	1.396	2.128	0.035
		Var(S)	0.103	0.022	$3.2E-4$	0.064	0.150	0.437
		Cov(I,S)	−0.057	0.050	$6.6E-4$	−0.161	0.036	0.243
		Var(e)	0.972	0.053	$4.7E-4$	0.873	1.080	−1.124
Missingness Parameters	1997	γ_{01}	−0.986	0.760	0.041	−2.491	0.539	0.204
		γ_{x1}	0.102	0.052	0.003	0.005	0.211	−0.466
		γ_{Y1}	−0.345	0.117	0.006	−0.591	−0.133	0.201
	1998	γ_{02}	−1.794	0.681	0.036	-3.213	−0.543	−0.162
		γ_{x2}	0.026	0.040	0.002	−0.053	0.104	−0.593
		γ_{Y2}	−0.019	0.082	0.004	−0.178	0.145	0.902
	1999	γ_{03}	−1.258	0.586	0.031	−2.344	−0.034	−1.477
		γ_{x3}	0.050	0.038	0.001	−0.024	0.124	1.272
		γ_{Y3}	−0.092	0.069	0.003	−0.230	0.045	0.804
	2000	γ_{04}	−1.638	0.606	0.033	-2.740	−0.371	−0.142
		γ_{x4}	4.4E-4	0.032	0.002	−0.060	0.064	−0.464
		γ_{Y4}	0.067	0.070	0.004	−0.076	0.187	0.529
			Model with ignorable (X) missingness					
Growth Curve	Parameters	Intercept	6.051	0.082	$4.7E-4$	5.890	6.210	1.303
		Slope	0.311	0.030	$2.5E-4$	0.252	0.369	−1.376
		Var(I)	1.683	0.183	0.001	1.349	2.064	1.431
		Var(S)	0.100	0.021	$3.0E-4$	0.062	0.145	1.052
		Cov(I,S)	−0.043	0.049	$5.7E-4$	−0.144	0.049	−1.945
		Var(e)	0.966	0.052	$3.8E-4$	0.869	1.072	−1.446

(continued)

Table 8 (continued)

Missingness Parameters	1997	γ_{01}	−2.554	0.553	0.029	−3.640	−1.433	−0.254
		γ_{x1}	0.080	0.043	0.002	−0.007	0.163	0.305
	1998	γ_{02}	−1.906	0.541	0.028	−2.963	−0.869	−0.282
		γ_{x2}	0.026	0.043	0.002	−0.058	0.109	0.294
	1999	γ_{03}	−1.784	0.420	0.021	−2.598	−0.978	−0.394
		γ_{x3}	0.044	0.033	0.002	−0.021	0.109	0.389
	2000	γ_{04}	−1.189	0.381	0.019	−1.914	−0.463	−0.287
		γ_{x4}	0.004	0.031	0.002	−0.056	0.061	0.281

Note:
Standard deviation.
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).
Geweke test t value. An absolute value less than 1.96 indicates
The shaded model is selected to be the best-fit model by all criteria in this study.
The shaded parameter estimate is significant from zero.

Table 9 Model selection in real data analysis

	non-ignorable missingness			ignorable
Criterion	LOD (XY)	LSD (XS)	LID (XI)	missingness
Dhat.AIC	4125.000	4083.000	N/A	4151.000
Dhat.BIC	4195.050	4153.050	N/A	4205.483
Dhat.CAIC	4213.050	4171.050	N/A	4219.483
Dhat.ssBIC	4137.944	4095.944	N/A	4161.067
DIC	4959.000	4953.000	N/A	4979.000
rough DIC	5730.878	5714.752	N/A	5731.980

Note:
The definition of each criterion is given in Table 1.
The shaded cell has the smallest value.

showed that the Bayesian method was able to accurately recover parameters in all models considered.

Next, Bayesian model selection criteria were studied to identify the best-fit model in the context of the correct missing mechanisms. Almost all the criteria were able to correctly identify the true model with high certainty.

We also illustrated the application of the Bayesian latent growth curve model with missing data through the analysis of mathematical ability growth data from the NLSY97 survey. In this example, the focus was on seeing how mathematical ability grew over the 4-year period, and whether mothers' education influenced the missing data pattern. Using the model selection criteria introduced in this study, we were able to identify the best-fit of the models considered. The results obtained from the best-fit model showed that mathematical ability grew significantly, and the missing data mainly depended on student's latent rate of growth. Further, mothers' education did not significantly influence the missing data pattern.

The models proposed in this paper can be further developed in various ways. First, the missingness in the simulation was assumed to be independent across time points. If this assumption is violated, likelihood functions will be different. For

example, if the missingness depends on the previous session, then autocorrelations might be involved, and the likelihood will be much more complicated. Furthermore, the missingness in practice can be a combination of different types of missingness, quite probably leading to development of increasingly more complex models. Second, additional model selection criteria could be considered, for example, Bayes factors and predictive posterior probabilities. Also, designing new criteria is an interesting topic for future work. It might be useful, for example, to consider observed-data or complete-data likelihood functions for random effects models for $p(\mathbf{y}|\theta)$. Third, the data considered in the study were assumed to be normally distributed. However, in reality data are seldom normally distributed, particularly in behavioral and educational sciences (e.g., Micceri 1989). When data have heavy tails, or are contaminated with outliers, robust models (e.g., Huber 1996) should be adopted to help reduce the sensitivity to small deviations from the assumption of normality. Fourth, latent population heterogeneity (e.g., McLachlan and Peel 2000) may exist in the collected longitudinal data. Growth mixture models (GMMs) can be considered to provide a flexible set of models for analyzing longitudinal data with latent or mixture distributions (e.g., Bartholomew and Knott 1999).

Appendix

The Derived Posteriors for LGCMs with Non-ignorable Missingness:

(1) Let $\eta = (\eta_1, \eta_2, \ldots, \eta_N)$, and the conditional posterior distribution for ϕ can be easily derived as an Inverse Gamma distribution,

$$\phi|\eta, \mathbf{y} \sim IG\,(a_1/2, b_1/2)\,,$$

where $a_1 = v_0 + NT$, and $b_1 = s_0 + \sum_{i=1}^{N}(\mathbf{y}_i - \Lambda\eta_i)'(\mathbf{y}_i - \Lambda\eta_i)$.

(2) Notice that $tr(\mathbf{AB}) = tr(\mathbf{BA})$, so the conditional posterior distribution for Ψ is derived as an Inverse Wishart distribution,

$$\Psi|\beta, \eta \sim IW\,(m_1, \mathbf{V}_1)\,,$$

where $m_1 = m_0 + N$, and $\mathbf{V}_1 = \mathbf{V}_0 + \sum_{i=1}^{N}(\eta_i - \beta)(\eta_i - \beta)'$.

(3) By expanding the terms inside the exponential part and combining similar terms, the conditional posterior distribution for β is derived as a multivariate normal distribution,

$$\beta|\Psi, \eta \sim MN(\beta_1, \Sigma_1),$$

where $\beta_1 = \left(N\Psi^{-1} + \Sigma_0^{-1}\right)^{-1}\left(\Psi^{-1}\sum_{i=1}^{N}\eta_i + \Sigma_0^{-1}\beta_0\right)$, and $\Sigma_1 = \left(N\Psi^{-1} + \Sigma_0^{-1}\right)^{-1}$.

(4) The conditional posterior for γ_t, $(t = 1, 2, \ldots, T)$, is a distribution of

$$p(\gamma_t|\omega,\mathbf{x},\mathbf{m}) \propto \exp\Bigg[-\frac{1}{2}(\gamma_t-\gamma_{t0})'\mathbf{D}_{t0}^{-1}(\gamma_t-\gamma_{t0}) + \sum_{i=1}^{N}\left\{m_{it}\log\Phi(\omega_i'\gamma_t)+(1-m_{it})\log[1-\Phi(\omega_i'\gamma_t)]\right\}\Bigg].$$

where $\Phi(\omega_i'\gamma_t)$ is defined by Eqs. (4), (5), or (6).

(5) By expanding the terms inside the exponential part and combining similar terms, the conditional posterior distribution for η_i, $i = 1,2,\ldots,N$, is derived as a Multivariate Normal distribution,

$$\eta_i|\phi,\Psi,\beta,\mathbf{y}_i \sim MN(\mu_{\eta i},\Sigma_{\eta i}),$$

where $\mu_{\eta i} = \left(\frac{1}{\phi}\Lambda'\Lambda+\Psi^{-1}\right)^{-1}\left(\frac{1}{\phi}\Lambda'\mathbf{y}_i+\Psi^{-1}\beta\right)$, and $\Sigma_{\eta i} = \left(\frac{1}{\phi}\Lambda'\Lambda+\Psi^{-1}\right)^{-1}$.

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

$$\mathbf{y}_i^{mis}|\eta_i,\phi \sim MN\left[\Lambda\eta_i,\mathbf{I}_T\phi\right],$$

where $\mathbf{I}_T$ is a $T\times T$ identity matrix. The dimension and location of $\mathbf{y}_i^{mis}$ 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.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5–37.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis: Kendall's library of statistics* (2nd ed., Vol. 7). New York, NY: Edward Arnold.

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

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Hoboken, NJ: Wiley.

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. Available from http://www.bls.gov/nls/nlsy97.htm

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

Dempster, A. (1974). The direct use of likelihood for significance testing. In *Proceedings of Conference on Foundational Questions in Statistical Inference* (pp. 335–352). University of Aarhus: Aarhus.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*, 1–16.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (Eds.). (2008). *Longitudinal data analysis*. Boca Raton, FL: Chapman & Hall.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

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 J. M. Bernado, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 169–193). Oxford, UK: Clarendon.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Drawing inferences from self-selected samples. In H. Wainer (Ed.), (pp. 115–142). New York: Springer.

Huber, P. (1996). *Robust statisticalprocedures* (2nd ed.). Philadelphia: SIAM.

Janssen, R., & De Boeck, P. (1999). Confirmatory analyses of componential test structure using multidimensional item response theory. *Multivariate Behavioral Research, 34*, 245–268.

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.

Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester, UK: Wiley.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley-Interscience.

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.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

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

Muthén, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335.

Oldmeadow, C., & Keith, J. M. (2011). Model selection in Bayesian segmentation of multiple DNA alignments. *Bioinformatics, 27*, 604–610.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*, 465–471.

Robert, C. P., & Casella, G. (2004). *Monte Carlostatistical methods*. New York, NY: Springer.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537–560.

Schafer, J. L. (1997). *Analysisof incomplete multivariate data*. Boca Raton, FL: Chapman & Hall/CRC.

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. *Psychometrics, 52*, 333–343.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583–639.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS manual Version 1.4. (Cambridge CB2 2SR, UK: MRC Biostatistics Unit, Institute of Public Health, Robinson Way. http://www.mrc-bsu.cam.ac.uk/bugs)

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*, 1–16.

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., & Lu, Z. (2008). SEM with missing data and unknown population using two-stage ML: Theory and its application. *Multivariate Behavioral Research, 43*, 621–652.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development, 31*(4), 374–383.