A Nonparametric Bayesian Approach for Longitudinal Nonnormal and Missing Data in Growth Curve Models

Dingjing Shi
Chongqing, China

Master of Arts, Psychology
Master of Science, Education
Bachelor of Arts, English

A Dissertation Presented to the Graduate Faculty of
the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
13 July 2020

Committee Members:

Xin Tong, PhD (Chair)
Steven Boker, PhD
Hudson Golino, PhD
Jingjing Li, PhD (McIntire School of Commerce)

A Nonparametric Bayesian Approach for Longitudinal Nonnormal and Missing Data in Growth Curve Models

Dingjing Shi

Abstract

Growth curve models (GCMs) are widely used to analyze longitudinal data. This study highlights two aspects of novelties to GCMs. The first novelty solves a practical issue. The study proposes a nonparametric Bayesian selection GCM to simultaneously handle missing and nonnormal data in longitudinal studies. Bayesian methods are used for the estimation. Multiple imputation and selection model approaches are used to handle the ignorable and non-ignorable missing data, respectively. The second novelty advances methodological theory. The proposed modeling framework is an infinite mixture modeling. The study develops new Bayesian model selection criteria to evaluate infinite mixture models in a Bayesian context and compares the performance of new criteria with existing criteria in Bayesian infinite mixture modeling. A Monte Carlo simulation study is conducted to assess the performance of the proposed modeling approach and the corresponding model selection criteria. Simulation results show that the nonparametric Bayesian GCMs perform better than the traditional normal-distribution-based GCM in analyzing the nonnormal and ignorable missing data. The BNP selection GCM in general outperforms the other two GCMs when data are nonnormal and non-ignorable missing. The new Bayesian model selection criteria have the potential to select the Bayesian infinite mixture models. A real data example using the NLSY97 survey data is provided to illustrate the application of the proposed method and model selection criteria.

Acknowledgements

First of all, I would like to express my appreciation to my advisor, Dr. Xin Tong, for opening my eyes to the world of quantitative psychology. I want to thank Dr. Tong for her patience, mentorship, and for being rigorous with me on my doctoral study. With her guidance and mentorship, she has taught me to be persistent and independent. I want to thank my professor and dissertation committee member, Dr. Steve Boker. I am grateful for his insights, encouragement, and support at each critical stage of my doctoral study. He continues to inspire me as I step into the next research journey. I am thankful to my professor and dissertation committee member, Dr. Hudson Golino, for all his valuable advice and his friendship. I have learned a lot from him as a scholar, professor, and thinker. I would like to thank Dr. Jingjing Li, for serving on my dissertation committee and for giving me excellent advice and input on my dissertation and future work.

I also want to express my gratitude to many professors I have met throughout my graduate school journey. They have helped shape me into who I am today. I am particularly thankful to Dr. Joanne Peng from Indiana University for being my first graduate school mentor that took me to a world completely new. I am thankful to Dr. Raymond Smith for inspiring me with his wisdom and knowledge. I want to thank all the schoolmates, classmates and fellow students I had at Indiana University and the University of Virginia, and thank all my dear friends for their unconditional help, their encouragement, and their company.

I would like to sincerely thank my parents for their love and for giving me the courage to reach for the stars and chase the dreams. I want to express my sincere thanks to my husband. Without his love and support, I could be nowhere near where I am in this journey. Lastly, I want to thank my little Peppa and George, for their love and comfort.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Growth curve modeling is a widely-used approach to analyze longitudinal data and can directly investigate intraindividual change over time and interindividual differences in change (McArdle and Nesselroade, 2014; Meredith and Tisak, 1990). Although the normal-distribution-based growth curve model (GCM) is commonly used, it faces the challenge of analyzing nonnormal outcomes which are often observed in social and behavioral research (Cain et al., 2017; Micceri, 1989). Analyzing longitudinal nonnormal data as if they were normal may lead to biased and inefficient parameter estimates and eventually come to misleading conclusions when conducting statistical inference (Shin et al., 2009; Zu and Yuan, 2010). Strategies including data transformation (Azuero et al., 2010) or exclusion of outliers (Osborne and Overbay, 2004) are applied by substantive researchers but can lead to problems such as uninterpretable estimation results (Osborne, 2005) or underestimated standard errors (SEs) (Yuan et al., 2002). Various robust procedures have been developed to address nonnormality by downweighing outliers that are far from the center of the majority of the data (Hampel et al., 1986; Huber, 1981; Yuan et al., 2002). Researchers also bring in Bayesian framework to model normal and nonnormal data, particularly through the specification of error distributions that are parametric in nature (Zhang, 2016).

The Bayesian nonparametric approach (BNP) does not restrict distributional assumptions to specific parametric forms and can better reflect true underlying distributions of latent variables or measurement errors. The BNP typically uses Dirichlet process (DP) priors to model distributions over spaces of distributions and has been applied to complex model structures to analyze nonnormal data. For instance, Kleinman and Ibrahim (1998b) introduced the

BNP in random effects model to place a DP mixture prior to random effects to make inferences to the nonnormal and correlated data. The idea has been further applied to generalized linear mixed effects models (Kleinman and Ibrahim, 1998a). Kottas and Gelfand (2001) proposed a BNP approach to median regression modeling for the robust analysis of censored data. Song et al. (2010) introduced the BNP to structural equation models to model nonnormal residuals in the measurement equation and thus flexibly model multivariate nonnormal data. Kelava and Brandt (2014) used BNP to study nonlinear interaction and quadratic effects in multilevel structural equation mixture models. Recently, Tong and Zhang (2019) proposed a nonparametric Bayesian growth curve model (BNP GCM) and used BNP with a DP mixture prior to analyze nonnormal longitudinal data. In particular, the traditional parametric normality assumption of measurement errors was replaced by an unknown distribution, constructed using a DP mixture prior.

The DP used in BNP GCM is an infinite-dimensional generalization of the Dirichlet distribution and thus the BNP GCM is an infinite mixture model. Despite its flexibility, two important aspects of BNP growth curve modeling have not yet been studied, which highlight the novelties of the current study. The first novelty solves a practical issue. Although BNP GCM has been demonstrated to be effective in estimating longitudinal nonnormal data, it can be more useful with an additional capability to handle missing data. The study proposes a BNP selection GCM to handle missing and nonnormal data simultaneously in longitudinal studies. The second novelty advances methodological theory. Although Bayesian model evaluation is an important topic, model evaluation in Bayesian mixture models is an understudied area. With regard to the nonparametric Bayesian growth curve model, which is a type of the infinite mixture model, even less is known about its model evaluations in the Bayesian literature. This study develops new Bayesian model selection criteria to evaluate infinite mixture models and compares the performance of the new criteria with existing criteria in Bayesian infinite mixture modeling. The new fit criteria may also have the potential to evaluate other models including finite mixture models in future studies.

Rubin (1976) distinguished three missing data mechanisms depending on how missing data are generated: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), respectively. Data are MCAR if the missingness does not depend on any factors that influence the outcome variable, meaning that the missingness is completely unrelated to any missing or observed data. Data are MAR when the

missingness on the outcome variable is only related to observed variables that influence the outcome variable. MCAR and MAR data are also called ignorable missing data because the missingness is independent of other variables or related to only observed variables so the missingness can be ignored or explained. The missingness is MNAR or non-ignorable when the missingness depends on the missing data itself or other unobserved factors and can not be ignored. Ignorable and non-ignorable missingness should be analyzed with different missing data analytical methods in order to obtain reliable parameter estimates (Allison, 2000; Little and Rubin, 2019; Schafer, 1997).

Traditional ad hoc missing data handling methods such as listwise deletion, pairwise deletion or mean substitution are viewed as "quick yet inappropriate" methods to handle missing data (Little and Rubin, 2019) and are generally not recommended by methodologists. These methods lead to biased estimates for MAR data and loss of power in general. Full Information Maximum Likelihood (FIML) estimation and Multiple Imputation (MI) were developed as the "state-of-the-art" methods (Schafer and Graham, 2002) to handle missing data. Numerous research studies have shown that under the ignorable missingness (i.e., MAR and MCAR), FIML and MI produce reliable parameter estimates and are recommended to use (e.g., Allison, 2003; Collins et al., 2001).

However, neither MI nor FIML properly handles the non-ignorable missing data and may result in inconsistent parameter estimates (e.g., Enders, 2001b; Schafer, 1997). Under the non-ignorable missingness, additional models must be included to account for the reasons why data are missing. One strategy is to include an existing modeling framework. For example, Enders (2011b) discussed using joint models (e.g., selection models, pattern mixture models) in growth curve analysis to study the MNAR data. In the current study, a selection model is added to the nonparametric Bayesian growth curve modeling. The proposed method turns the unexplained missingness in the longitudinal outcome variable into a missing data problem whose missingness mechanism can be explained (i.e., ignorable missingness), the mechanism of which can then be easily handled by methods such as MI.

Model evaluation has always been an essential topic in Bayesian methods. Although there is an abundant literature to evaluate Bayesian models in general (Gelman et al., 1996; Kass and Raftery, 1995; Spiegelhalter et al., 2002), literature and guidelines on how to select mixture models in the Bayesian context are much smaller in quantity. Directly applying the commonly used criterion such as the Deviance Information Criterion to mixture models is

problematic (Celeux et al., 2000; DeIorio and Robert, 2002). Lu et al. (2015) proposed model selection criteria in Bayesian growth mixture models and found they can correctly identify true finite mixture models. The nonparametric Bayesian growth curve modeling is an infinite mixture model. In Bayesian model evaluation literature, there is a lack of understanding on how to select infinite mixture models in a Bayesian context. This study proposes model selection criteria based on measures of information criterion to evaluate and select a best infinite mixture model in the Bayesian context. The study further compares the newly-proposed model selection criteria with a set of existing Bayesian model selection criteria. The existing model selection criteria were developed by Lu et al. (2015) in the context of Bayesian growth mixture models, or a type of finite mixture models with missing data and outliers. This dissertation develops new model selection criteria and investigates the performance of existing Bayesian criteria that were developed for finite mixture models. A total of thirteen Bayesian criteria are studied and compared in nonparametric Bayesian growth curve modeling, or in the context of Bayesian longitudinal infinite mixture modeling.

In sum, the dissertation proposes a BNP selection GCM to simultaneously handle missing and nonnormal data and develops Bayesian model selection criteria to evaluate the performance of the proposed models. Chapter 2 reviews normal-distribution-based linear growth curve models and discusses common approaches to address nonnormality in longitudinal data. Nonparametric Bayesian approaches to handling nonnormal data and particularly their applications to GCMs are discussed. Chapter 3 reviews the literature in missing data analysis. Chapter 4 proposes nonparametric Bayesian selection growth curve modeling to simultaneously handle nonnormal and missing data. Chapter 5 proposes Bayesian model evaluation criteria to select best fitting Bayesian mixture models. In particular, four new model selection criteria are proposed and a total of thirteen model selection criteria are investigated for nonparametric Bayesian growth curve modeling. The performance of the proposed measures are evaluated by comparing three longitudinal models in the GCM family, namely the normal-distribution-based GCM, BNP GCM and BNP selection GCM. Chapter 6 conducts a simulation study to systematically evaluate the performance of the proposed approach and the model selection criteria. In Chapter 7, the Peabody Individual Achievement Test mathematical data from the National Longitudinal Survey of Youth 1997 Cohort is used as a real data example to illustrate the application of the proposed models as well as the application of the model selection criteria.

Chapter 8 ends the study with discussions and future directions.

Chapter 2

GCM with Nonnormal Data

This chapter reviews normal-distribution-based linear growth curve models (GCMs) and discusses common approaches to addressing nonnormality in longitudinal data. It then reviews nonparametric Bayesian approaches to handle nonnormal data and their applications to complex models and GCMs.

2.0.1 Growth curve models

Longitudinal data are common in psychological, educational, developmental and health studies. Growth curve modeling is a popular technique to analyze longitudinal data. Growth curve modeling simultaneously studies intraindividual change over time and interindividual differences in intraindividual changes (McArdle and Nesselroade, 2014; Meredith and Tisak, 1990). In other words, growth curve modeling investigates how within-person performances change over time and how different the changes are between individuals. GCMs have gained popularity as they can be fitted under the structural equation modeling (SEM) framework.

A typical form of linear GCMs can be written as

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\Lambda} \boldsymbol{b}_i + \boldsymbol{e}_i, \\ \boldsymbol{b}_i &= \boldsymbol{\beta} \boldsymbol{X}_i + \boldsymbol{u}_i \end{aligned} \tag{2.1}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ is a vector of observations for individual i, $i = 1, ..., N$ with N denoting sample size and T denoting number of measurement occasions. The factor loading matrix $\boldsymbol{\Lambda}$ (e.g., $\boldsymbol{\Lambda} = ((1, 1, \ldots, 1)', (0, 1, \ldots, T-1)')$ for equal time intervals) represents the linear growth trajectories. The

$\boldsymbol{e}_i = (e_{it}, ..., e_{iT})'$ is a vector of intraindividual measurement errors for the ith individual and is usually assumed to be normally distributed as $\boldsymbol{e}_i \sim MN(\mathbf{0}, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ indicates level-1 residual variance. The vector of latent variables $\boldsymbol{b}_i = (b_{Li}, b_{Si})'$ are functions of fixed effects $\boldsymbol{\beta} = ((\beta_{00}, \beta_{01}, ..., \beta_{0m})', (\beta_{10}, \beta_{11}, ..., \beta_{1m})')'$, and random residuals, with a vector of possible covariates $\boldsymbol{X}_i = (X_{1i}, X_{2i}, \ldots, X_{mi})'$, where m is the number of covariates in the model. The vector of random effects $\boldsymbol{u}_i = (u_{Li}, u_{Si})'$ are residuals that remain unexplained by the level-2 predictors and are assumed to be normally distributed as $\boldsymbol{u}_i \sim MN(\mathbf{0}, \boldsymbol{\Psi})$ in traditional normal-distribution-based GCMs. Figure 2.1 is a path diagram of the linear GCM.

GCMs can be estimated using Bayesian methods. In Bayesian inference, parameter estimates are obtained and all statistical inferences are made from the posterior distributions of model parameters. In a Bayesian approach to estimate GCMs, we combine prior distributions of model parameters and information of the data to obtain the joint posterior distributions of parameters. Zhang et al. (2013) demonstrated the joint probability and likelihood function of the linear GCM. In the linear GCMs, the joint probability distribution of $\boldsymbol{y}_i$, $\boldsymbol{b}_i$ is

$$\begin{aligned} p(\boldsymbol{y}_i, \boldsymbol{b}_i|\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\beta}) &= p(\boldsymbol{b}_i|\boldsymbol{\Psi}, \boldsymbol{\beta}) p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\Phi}, \boldsymbol{\Lambda}) \\ &= (2\pi)^{-m/2} |\boldsymbol{\Psi}|^{-1/2} exp[-\frac{1}{2}(\boldsymbol{b}_i - \boldsymbol{\beta})' \boldsymbol{\Psi}^{-1} (\boldsymbol{b}_i - \boldsymbol{\beta})] \\ &\times (2\pi)^{-T/2} |\boldsymbol{\Phi}|^{-1/2} exp[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Lambda}\boldsymbol{b}_i)' \boldsymbol{\Phi}^{-1} (\boldsymbol{y}_i - \boldsymbol{\Lambda}\boldsymbol{b}_i)]. \end{aligned}$$

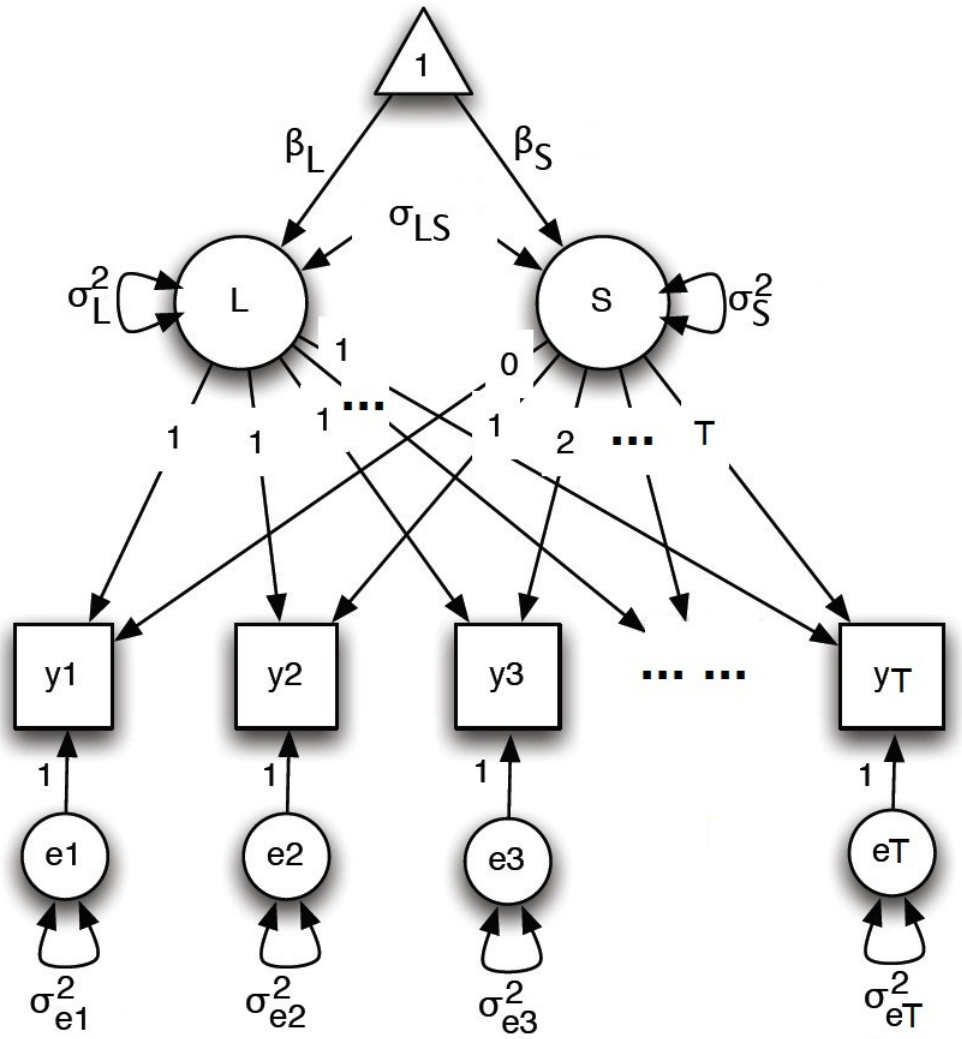
Thus, the likelihood function for the linear GCM is

$$\begin{aligned} L &= \prod_{i=1}^{N} p(\boldsymbol{y}_i, \boldsymbol{b}_i|\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \boldsymbol{\Lambda}) \\ &\propto |\boldsymbol{\Psi}|^{-N/2} exp[\sum_{i=1}^{N} (\boldsymbol{b}_i - \boldsymbol{\beta})' \boldsymbol{\Psi}^{-1} (\boldsymbol{b}_i - \boldsymbol{\beta})] \times |\boldsymbol{\Phi}|^{-N/2} exp[-\sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{\Lambda}\boldsymbol{b}_i)' \boldsymbol{\Phi}^{-1} (\boldsymbol{y}_i - \boldsymbol{\Lambda}\boldsymbol{b}_i)], \end{aligned}$$

where the unknown parameters in the linear GCM include $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Let $p(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ denote the joint prior distribution of these parameters. Thus, the joint posterior distribution is

$$p(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}|\boldsymbol{y}_i) \propto \int p(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \times L d\boldsymbol{b}.$$

Figure 2.1: PATH DIAGRAM OF THE LINEAR GROWTH CURVE MODEL



The conditional posterior distributions for the linear GCM derived by Zhang et al. (2013) can be used as a transition kernel to the joint distribution. The Markov chain Monte Carlo (MCMC) algorithm (e.g., Gibbs sampler) in Bayesian methods generates a sequence of samples from the joint probability distribution of two or more random variables (Casella and George, 1992). In specific, Gibbs sampler alternatively samples one parameter conditional on the current values of other parameters sampled from conditional posterior distributions. The Markov chains converge to stationary distributions after a sufficient number of iterations (Geman and Geman, 1984). Gibbs sampling is especially useful when the joint probability distribution is too complex or unknown at all, but the conditional distribution for each parameter can be made available.

The Gibbs sampling algorithm is used to generate Markov chains to construct estimates of parameters from the linear GCM. Details of the Gibbs sampling algorithm for the linear GCMs are given below.

1. Start with initial values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\Psi}^{(0)}$, $\boldsymbol{\Phi}^{(0)}$, $\boldsymbol{b}_i^{(0)}$.
2. Assume at the jth iteration, we have $\boldsymbol{\beta}^{(j)}$, $\boldsymbol{\Psi}^{(j)}$, $\boldsymbol{\Phi}^{(j)}$, $\boldsymbol{b}_i^{(j)}$.
3. At the $(j+1)th$ iteration,
   - 3.1. Sample $\boldsymbol{\beta}^{(j+1)}$ from $p(\boldsymbol{\beta}|\boldsymbol{\Psi}^{(j)}, \boldsymbol{b}_i^{(j)})$;
   - 3.2. Sample $\boldsymbol{\Psi}^{(j+1)}$ from $p(\boldsymbol{\Psi}|\boldsymbol{\beta}^{(j+1)}, \boldsymbol{b}_i^{(j)})$;
   - 3.3. Sample $\boldsymbol{\Phi}^{(j+1)}$ from $p(\boldsymbol{\Phi}|\boldsymbol{b}_i^{(j)}, \boldsymbol{y}_i)$;
   - 3.4. Sample $\boldsymbol{b}_i^{(j+1)}$, $i = 1, \ldots, N$ from $p(\boldsymbol{b}_i|\boldsymbol{\Phi}^{(j+1)}, \boldsymbol{\Psi}^{(j+1)}, \boldsymbol{\beta}^{(j+1)}, \boldsymbol{y}_i)$.
4. Repeat Step 3.

2.0.2 Nonnormality

Traditional GCMs discussed above are based on the normality assumption of intraindividual measurement errors and random effects. However, longitudinal data in social, psychological and behavioral sciences often violate the normality assumption. Normal-distribution-based longitudinal models and designs in particular face the statistical challenge of analyzing nonnormal outcomes. Analyzing longitudinal nonnormal data as if they were normal may lead to biased and inefficient parameter estimates (Shin et al., 2009; Zu and Yuan, 2010). In a typical two-level normal-distribution-based GCM with complete data, failure to model the nonnormality at Level One will not bias parameter estimation of fixed effects, but will affect standard error (SE) estimates for both fixed and random effects, which will damage the validity

of hypothesis testing and may eventually distort statistical inference (Brandt and Klein, 2015).

There are several approaches to addressing nonnormality in longitudinal data. One common approach is the data transformation so that residuals will be unskewed. In a Breast Cancer Educational Intervention study, Azuero et al. (2010) used the natural logarithm as a link function to model the right-skewed longitudinal outcome of monthly health-related out-of-pocket expense data. Bernier et al. (2011) used an arc sine transformation to the skewed longitudinal data that are related to the health and life quality from the Statistics Canada's National Population Health Survey and then continued with normal-based growth curve analysis to study the health status changes over time.

Another approach to addressing nonnormality is to use robust procedures. The rationale of many robust procedures is to weight each observation according to its distance from the center of the majority of the data, so that outliers are downweighted (Hampel et al., 1986; Huber, 1981). Many other robust procedures have used Bayesian framework to model normal and nonnormal data through the specification of error distributions (Zhang, 2016). Robust procedures based on Student's t distributions have been developed and advanced to model heavy-tailed data or data containing outliers (Lange et al., 1989; Yuan and Zhang, 2012). Zhang et al. (2013); Tong and Zhang (2019) proposed robust GCMs and modeled measurement errors using Student's t distributions. Wang et al. (2008) developed a tobit GCM to model limited response outcome data with ceiling effects. Ferron et al. (2002) conducted simulation methods to examine the effects of misspecifying error structures in two-level longitudinal growth model.

2.0.3 Bayesian nonparametric approach for nonnormal data

Researchers typically model normal and nonnormal data using different parametric forms. In practice, shapes of nonnormal data can vary. Specifying a parametric distribution hardly reflects true underlying distributions of latent variables or measurement errors. Constraining inference to a specific parametric form can result in model misspecifications, which may eventually distort inferences and mislead statistical conclusions. A nonparametric Bayesian approach (BNP) relaxes the parametric assumptions of distribu-

tions and provides distributional flexibility and robustness against distributional misspecifications. By using a nonparametric approach, data distributions are not restricted to a specific parametric form; and by using the Bayesian method, full probability models for the data generating process are provided.

The BNP uses Dirichlet process (DP) to model distributions over spaces of distributions. DP was proposed as the first prior to define for spaces of distribution functions (Ferguson, 1973). The DP is a stochastic process whose sample realization is a set of probability distributions. The DP generates a random probability measure. For every measurable partition of the random probability measure, it follows a Dirichlet distribution. The space of all probability distributions for a random probability measure is infinite dimensional (Ghahramani, 2013) and DP is an infinite-dimensional generalization of the Dirichlet distribution.

Ferguson (1973) and Ferguson et al. (1974) showed that random distributions drawn from the DP are discrete. To describe random distributions that are continuous, Antoniak (1974) and Escobar (1994) developed a DP mixture approach to expand the space of probability distributions to continuous distributions. The DP mixture is implemented in a way that a mixture of certain distributions is constructed, which will be a continuous distribution, and the DP prior is assigned to the mixture component of the newly constructed continuous distribution. Posterior distributions are typically analytically intractable, so the inference instead usually involves computational simulation (e.g., Gelman et al., 2013). In the early nineties, the Markov chain Monte Carlo (MCMC) algorithm was introduced to draw samples from the DP mixture (e.g., Escobar, 1994; Zeger and Karim, 1991), the posterior distribution of which is highly complex over function spaces. Neal (2000) further developed MCMC algorithms based on collapsed Gibbs sampling and Metropolis-Hastings, which led to analytically tractable posterior distributions and made the application of the BNP methods possible.

Complex models have used the BNP approach to relax distributional assumptions and analyze nonnormal data. Kleinman and Ibrahim (1998b) introduced the BNP in random effects model and placed a DP mixture prior to random effects to make inferences to the nonnormal and correlated data. The idea has been further applied to generalized linear mixed effects models (Kleinman and Ibrahim, 1998a). Kottas and Gelfand (2001) proposed a BNP approach to median regression modeling for the robust analysis of censored data. Song et al. (2010) introduced the BNP using DP to structural equa-

tion models to model nonnormal data in SEM. Kelava and Brandt (2014) used BNP to study nonlinear interaction and quadratic effects in multilevel structural equation mixture model.

2.0.4 Bayesian nonparametric GCM

2.0.4.1 Dirichlet process (DP) priors

The BNP method using DP or DP mixture prior has been applied to GCMs for nonnormal data analysis. Tong and Zhang (2019) used the BNP method to propose a robust GCM (hereafter named BNP GCM) and found that the BNP GCM flexibly models nonnormal residuals and analyzes longitudinal nonnormal data. In the BNP GCM, the distribution of measurement errors is modeled by a random distribution function with a nonparametric prior. In particular, the traditional parametric normality assumption of intraindividual measurement errors, e.g., $\boldsymbol{e}_i \sim MN(0, \boldsymbol{\Phi})$, is replaced by an unknown distribution G_e. Because intraindividual measurement errors $\boldsymbol{e}_i$ follow a continuous distribution, the prior of the unknown distribution G_e is chosen to be the Dirichlet process mixture (DPM) (Antoniak, 1974) and can be expressed as

$$\begin{aligned} \boldsymbol{e}_i|\boldsymbol{\Phi}_i &\sim MN(\mathbf{0}, \boldsymbol{\Phi}_i), \\ \boldsymbol{\Phi}_i|G &\sim G, \\ G &\sim DP(\alpha, G_0). \end{aligned} \tag{2.2}$$

Equation 2.2 means that the intraindividual measurement errors $\boldsymbol{e}_i \sim G_e$ and $G_e \sim DPM$ are constructed by a mixture of multivariate normal distributions, where a DP prior is assigned to the mixing measure of the mixture of multivariate normal distributions for $\boldsymbol{e}_i$.

To illustrate how a DP prior is assigned to the weight proportion of the mixture of multivariate normal distribution, consider a random probability measure G distributed as DP and denoted as $G \sim DP(\alpha, G_0)$. For a measurable partition of Θ, $P_1, ...P_k$,$G(P_1), ..., G(P_k)$ follows a Dirichlet distribution,

$$(G(P_1), ..., G(P_k)) \sim Dirichlet(\alpha G_0(P_1), ..., \alpha G_0(P_k)). \tag{2.3}$$

Equation 2.3 defines DP and shows that for a random distribution drawn from the DP, if we add up the probability mass in a region $P \in \Theta$, there will be on average $G_0(P)$ mass in that region (Ferguson et al., 1974). For

a random probability measure $G \sim DP(\alpha, G_0)$, there are two parameters, α and G_0 in DP. G_0 is a base distribution, which is the mean distribution in the distribution space of the DP, and is expressed as $E(G) = G_0$. α is a concentration parameter, which can be understood as proportional to an inverse variance of the DP as $Var(G) = \frac{G_0(1-G_0)}{\alpha+1}$. The larger the α, the smaller the variance, and the more concentrated the DP is to have its mass around the mean. The concentration parameter α governs the discrete realization of the random probability measure G to its base distribution G_0.

2.0.4.2 Stick Breaking Construction

With a probability of one, which is the probability of a weight measure, G can be written as

$$G = \Sigma_{k=1}^{\infty} \pi_k \delta_{\theta_k},$$

where δ_{θ_k} shows the location of the parameters θ in the cluster and π_k corresponds to the weights or the mixing proportions of the locations and $\Sigma_{j=1}^{\infty} \pi_k = 1$. To illustrate how the mixing proportions are obtained, let q_i be draws from a sequence of identically and independently distributed Beta random variables, $q_i \sim Beta(1, \alpha)$. Define

$$\pi_k = q_k \Pi_{j=1}^{k-1}(1 - q_j).$$

The process of obtaining the mixing proportions is given below.

1. Draw q_1 from $Beta(1, \alpha)$, then $\pi_1 = q_1$;
2. Draw q_2 from $Beta(1, \alpha)$, then $\pi_2 = q_2(1 - q_1)$;
3. Draw q_3 from $Beta(1, \alpha)$, then $\pi_3 = q_3(1 - q_2)(1 - q_1)$;

$\vdots$

Repeat this process till all mixing proportions π_k are obtained. The idea of this process is similar to breaking off from a stick that has a total probability of one, and then iteratively breaking off from the remaining sticks till all mixing proportions are obtained. Sethuraman (1994) found that draws from this stick-breaking process are DP distributed and developed this process as the "stick-breaking construction". The distribution $G(\cdot)$ is a discrete

distribution as

$$G(\cdot) = \begin{cases} \delta_{\theta_1}, & p = \pi_1 \\ \delta_{\theta_2}, & p = \pi_2 \\ \vdots & \vdots \\ \delta_{\theta_k}, & p = \pi_k \\ \vdots & \vdots \end{cases}.$$

To define a continuous distribution, the DP can be used as the basis of a mixture model and the unknown distribution G_e can be constructed as a mixture of multivariate normal distributions.

2.0.4.3 Truncated Stick Breaking Construction

Theoretically, there can be an infinite number of clusters as $k = 1, ..., \infty$. In practice, a finite number of mixture components should be sufficient to describe distributions. Therefore we can obtain this mixture of multivariate normal distributions using the truncated stick-breaking construction (e.g., Ishwaran and James, 2001), by truncating the data to a maximum of C $(1 \leq C \leq N)$ possible mixture components, or in other words, by breaking the sticks C times. For $q_1, ..., q_C \sim Beta(1, \alpha)$, define $\pi_k = q_k \prod_{j=1}^{k-1}(1 - q_j)$, $k = 1, ..., C$. Then the mixing proportion p_k is $p_k = \frac{\pi_k}{\Sigma_{j=1}^{C} p'_j}$ and that $\Sigma_{k=1}^{C} p_k = 1$. The mixture of multivariate normal distribution G_e, which has weight proportions governed by the DP prior, can be expressed as

$$G_e = \begin{cases} MN(\boldsymbol{\mu}_e^{(1)}, \boldsymbol{\Phi}^{(1)}), & p = p_1 \\ MN(\boldsymbol{\mu}_e^{(2)}, \boldsymbol{\Phi}^{(2)}), & p = p_2 \\ \vdots & \vdots \\ MN(\boldsymbol{\mu}_e^{(C)}, \boldsymbol{\Phi}^{(C)}), & p = p_C \end{cases} \tag{2.4}$$

where $p_1, \ldots, p_C$ are mixing proportions and C is a finite number of clusters prespecified in the truncated sticking breaking procedure. The actual number of clusters can be directly estimated from the model and is equal or less than C.

Therefore, BNP GCM is represented as Equations 2.1 and 2.2. The intraindividual measurement error $\boldsymbol{e}_i \sim G_e$ and $G_e \sim DPM$ in BNP GCM, as illustrated in Equations 2.2 and2.4, is a mixture of multivariate normal

distributions and can be obtained by the truncated stick-breaking construction (Sethuraman, 1994). In this way, the BNP GCM relaxes the parametric assumptions of intraindividual measurement errors and flexibly models the nonnormality in the longitudinal data.

Ferguson (1973) found that the Dirichlet process is a conjugate prior. For a DP prior $G \sim DP(\alpha, G_0)$, the posterior of G is also a DP and expressed as $G \sim DP(\tilde{\alpha}, \tilde{G_0})$. The two parameters $\tilde{\alpha} = \alpha + N$, and $\tilde{G_0} = \frac{\alpha}{\alpha+N} G_0 + \frac{N}{\alpha+N} G_N$, where G_0 is the base distribution of the DP and G_N is the empirical distribution function of the data. In the MCMC procedure, we sample G from $DP(\tilde{\alpha}, \tilde{G_0})$, then sample $\boldsymbol{e}$ from the generated G, and replicate the procedure multiple times to finally obtain the empirical distribution of the measurement errors $\boldsymbol{e}$. Tong and Zhang (2019) found that the BNP GCM performs as well as, or better than, the traditional normal-based GCM when data are nonnormal.

The BNP GCM (Equations 2.1 and 2.2) is flexible to model the longitudinal nonnormal data. The following two chapters (Chapters 3 and 4) extend the nonparametric Bayesian growth curve modeling to handle longitudinal missing data and further develop a selection model to the BNP GCM approach to simultaneously deal with nonnormal and missing data in longitudinal research.

Chapter 3

Review of Missing Data

Chapter 3 reviews the literature in missing data. Missing data in longitudinal studies is a ubiquitous problem. Because growth data are often collected over time, they are likely missing at one or more occasions.

Although dealing with missing data is usually not the focus of substantive research, failure to do so may result in serious problems. First, studies have found that ignoring missing data at the analysis stage could introduce potential bias in parameter estimation (Ghosh and Pahwa, 2008; Rubin, 2004; Schafer, 1997). A respondent's profile could be very different with and without missing data. Thus, samples may no longer be representative of the population, and conclusions drawn solely from those who responded may be biased. Moreover, missing data can lead to inefficiency and a loss of information, which eventually reduces statistical power (Davey et al., 2009; Little, 1988). In addition, missing data make common statistical methods inappropriate or difficult to apply (Peng et al., 2006; Rubin, 2004).

Chapter 3 comprises two parts. The first part reviews missing data mechanisms, which lay foundations of missing data theory, and serve as vital assumptions of missing data treatment. The second part discusses missing data treatment methods, including traditional ad-hoc missing data treatment, modern missing data treatment and joint models.

3.0.1 Missing data mechanisms

3.0.1.1 Underlying theory: Modeling the probabilistic phenomenon

The key to missing data procedures is that missingness is treated as a probabilistic phenomenon (Rubin, 1976). Users define a set of variables with a joint probability distribution to reflect missingness. In specific, for each variable in a data set (referred to as the substantive variable hereafter), we define a corresponding indicator variable R to denote whether a score on the associated substantive variable is observed or missing (i.e., $R = 1$ if missing and 0 otherwise). We model the relationship between the indicator variable and the associated substantive variable and thus model the relations between the missingness and the observed and/or missing variables.

This probabilistic relationship between the indicator variable and the corresponding substantive variable is summarized as the missing data mechanism. Different relations define different missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR).

3.0.1.2 MAR and MCAR (Ignorable missingness)

Data are MAR if the probability that the data are missing depends on and only on observed variables, but not on missing variables (Rubin, 1976). MCAR mechanism is a weaker form of MAR and can be seen as a special case of MAR (Allison, 2000; Demirtas and Schafer, 2003; Peng et al., 2006). Data are MCAR if the probability of missingness is neither related to missing data nor related to the observed data. The missingness is purely a "chance mechanism" (Molenberghs et al., 2014, p.7).

MCAR is a more restrictive condition than MAR, and is a "benign condition" (Black et al., 2012) in terms of estimating parameters. Some traditional ad hoc methods such as listwise deletion provide unbiased estimates for MCAR data, but will suffer from a loss of power. In all, missingness from MAR and MCAR data is independent of all other variables or related to only observed variables and thus the missingness can be ignored or explained. MAR and MCAR mechanisms are also said to be the ignorable missingness (Little and Rubin, 2019; Schafer, 1997).

3.0.1.3 MNAR (Non-ignorable missingness)

The MNAR mechanism occurs when the MAR mechanism is violated (Demirtas and Schafer, 2003; Schafer and Graham, 2002). Data are MNAR when the probability of missingness depends on unobserved or missing data. The probability of missingness in MNAR depends on the variable that is missing itself or unobserved, and thus missingness can not be ignored or explained. The MNAR mechanism is also the non-ignorable missingness (Rubin, 1976; Little and Rubin, 2019; Schafer, 1997; Diggle and Kenward, 1994).

In sum, missing data mechanisms describe the nature of relationships between models that contain missing values and substantive variables that are observed or unobserved in the model. Different statistical methods are required to handle missing data with different mechanisms (e.g., Rubin, 1976; Little and Rubin, 2019).

3.0.2 Missing data analytical techniques

3.0.2.1 Traditional ad hoc method

Traditional ad hoc approaches to handling missing data include listwise deletion, pairwise deletion, and single imputation. Listwise deletion discards subjects that have missing values on one or more variables from the data set, and makes statistical analysis exclusively on units that have complete information on all the variables. Using listwise deletion can result in a loss of information. Simulation studies have shown that as long as the MCAR assumption holds, listwise deletion produces unbiased parameter estimates, but statistical power and precision of parameter estimation will suffer (Brown, 1994; Enders, 2001a; Roth and Switzer III, 1995; Wothke, 2000). Different from listwise deletion, pairwise deletion uses observed data information available from each unit to compute descriptive and inferential statistics of the sample. Pairwise deletion takes advantage of even partial information from the units and makes statistical inferences. Using pairwise deletion can be problematic in multivariate data settings as many multivariate techniques are based on the raw data matrix (Rubin, 1976).

Single imputation methods are also used for missing data analysis. Mean imputation substitutes the mean of the observed values for all missing values. Hot-deck imputation is another single imputation method that "fills in" each missing values from a random draw of observed values of the same variable in the data set. The biggest deficiency of hot-deck imputation is it

lacks theoretical ground in terms of statistical properties under MCAR or MAR assumptions Roth and Switzer III (1995). The amount of bias with hot-deck imputation is affected by the missing data rate (Switzer III et al., 1998) so the hot-deck imputation is not much used in practice. The regression imputation method imputes missing values with values predicted from a regression equation based on variables with observed values in the same data set. Similar to the mean imputation and the hot-deck imputation, regression imputation preserves units with missing information. Regression imputation is better than other traditional ad hoc methods in that it utilizes information available in the data set and makes reasonable inference. However, problems in regression imputation remains as Donner (1982); Taljaard et al. (2008) argued that a) regression models need to be specified and are subjective to different sample variables, b) the imputed values reduce the uncertainty; and c) applying regress in imputation to multivariate data with more than one variables having missing values can be difficult.

Traditional ad hoc methods are easy to implement, so have had many users for the last decade in mainstream education and psychology journals (Peugh and Enders, 2004) and are accessible in most statistical software. However, studies have shown that they typically produce biased results (Brown, 1994; Peng et al., 2006; Wothke, 2000) if data are not MCAR. Little and Rubin (2019) viewed them as "quick yet inappropriate" methods to missing data. They are usually not recommended by methodologists (Little and Rubin, 2019; Newman, 2014).

3.0.2.2 State-of-the-art method

In the past decades, two widely-accepted approaches to handling missing data, full information maximum likelihood (FIML) and multiple imputation (MI) were developed. The two approaches are viewed as the "state-of-the-art" methods (Schafer and Graham, 2002). Following the pioneering work of Little and Rubin (1986), numerous research has shown that FIML and MI produce accurate and efficient parameter estimates and are recommended to use when the missing data are ignorable (e.g., Allison, 2000; Collins et al., 2001; Enders, 2001b; Schafer, 1997).

FIML estimation is a model-based method that uses likelihood function to draw statistical inferences. It estimates parameters directly from the model and does not require a complete data set. Given ignorable missing data (i.e., MAR or MCAR), FIML maximizes the observed data likelihood to obtain

the maximum log-likelihood estimates of the model parameters. Because the sample log-likelihood takes advantage of all available information to improve estimation, FIML is superior to traditional missing data techniques. There are two key assumptions in FIML estimation to handling missing data and providing unbiased and efficient parameter estimates: the MAR assumption and the multivariate normality assumption. When the two assumptions are met, in addition to unbiased (Enders and Bandalos, 2001) and efficient (Wothke, 2000) parameter estimates, FIML produces valid model fit information (Enders, 2001b). Enders (2001a) studied the performance of FIML under the violation of the MAR and multivariate normality assumptions, and found that 1) as long as data are under MCAR and MAR mechanisms, FIML produces unbiased estimates regardless of missing rates, sample size or the distribution shape, and that 2) FIML produces negatively biased SEs and inflated rejection rates with nonnormal indicator variables in SEM. Therefore, Enders (2001a) recommended the rescaled statistics or the bootstrap methods to correct for the bias arisen from nonnormality.

MI is a second "state-of-the-art" method and takes on an approach different from FIML to handle MAR and MCAR data. MI fills in the missing values for individuals to have complete data records before the analysis starts, and combines the filled-in values into one single result. A multiple imputation analysis is comprised of three steps: an imputation step (filling in the missing values), an analysis step (common statistical analysis), and a pooling step (combining everything into one single result). The goal of the imputation step is to fill in multiple sets of estimated missing values, using the available information from the observed data. A variety of algorithms have been developed to suit for different models in the imputation step (e.g., Lavori et al., 1995; Buuren and Groothuis-Oudshoorn, 2010), one of which is the MCMC based method (Schafer, 1997).

3.0.2.3 Joint models

As previously discussed, the missing data mechanism is said to be ignorable if the missingness does not depend on the missing or unobserved data, and the parameters from the data model and the missingness model are distinct. Because missingness does not depend on the missing data, the ignorable missingness only uses observed values for estimation or imputation and does not need to include information about the values that are missing.

However, this is not the case for non-ignorable missingness. MNAR means

that the missing values in the outcome variable are related to the missingness. Under non-ignorable missingness, additional models have to be included to account for why data are missing. In other words, the missing data mechanism needs to be modeled in addition to the substantive model of interest with observed values. For such missingness, traditional ad hoc as well as the state-of-the-art methods will produce biased estimation results and invalid inferences (Arbuckle et al., 1996; Enders, 2001a,b; Gold and Bentler, 2000; Muthén et al., 1987; Olinsky et al., 2003; Wothke, 2000).

Instead, joint models are needed. The purpose of the joint models is to model the joint distribution of the outcome data and the missingness indicator. Selection models (Diggle and Kenward, 1994; Heckman, 1976) and pattern mixture models (Demirtas et al., 2008; Molenberghs et al., 2014) are two aspects of modeling the same joint distribution and are two widely-accepted joint models.

The selection model was first proposed by Heckman (1976) as a method to correct for the bias associated with analysis from the outcome data that are under MNAR mechanism. Diggle and Kenward (1994) further developed the selection model as a type of joint models to simultaneously model two types of variables, the outcome variable of substantive interest Y and the missingness indicator variable R. Because the joint model can jointly take on different forms, the selection model is flexible in its form. For example, in analyzing longitudinal MNAR data, the substantive model could be a growth curve model. Because the missingness indicator R is binary, either a logit or a probit model can be used to model the missingness indicator. One form of the selection model can be expressed as

$$\frac{log(P(R=1|y_{i,t}, y_{i,(t-1)}))}{log(P(R=0|y_{i,t}, y_{i,(t-1)}))} = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 y_{i,t} + \boldsymbol{\beta}_2 y_{i,(t-1)}, \tag{3.1}$$

where i denotes the ith participant, t is the time that the missingness or dropout occurs and $t-1$ denotes the time preceding the missingness or dropout.

Rather than studying the marginal distribution of the outcome variable, pattern mixture model stratifies samples into subgroups with a common missing data pattern and estimates model parameters separately within each pattern. The logic of pattern mixture models is similar to estimating missing data patterns with stratified cases, or like a multiple group missing data approach (Muthén et al., 1987), which was an old missing data treatment strategy before the principled FIML techniques came into shape. In other

words, pattern mixture models estimate common substantive models (e.g., growth curve model) using observed cases with a stratified subsample, within which each subsample has the same number of available cases. Take a longitudinal four-wave study as an example, given that data are in a monotone missing pattern and the first wave is assumed to have complete cases, which constitutes the first pattern; the second wave has dropout cases after the baseline wave and forms a second pattern, etc. In the end, there are four patterns composed from the four waves of data. Suppose a GCM is used for the analysis, a growth curve analysis is conducted within each pattern and forms a total of four sets of parameter estimates from four growth curve analyses. Depending on the form of individual patterns, the parameter estimates could be unique and become the final parameter estimate, or are different, and the final parameter estimate is the weighted average of the several sets of parameter estimates (Enders, 2011b).

This study will focus on selection models to analyze MNAR data. The following chapter proposes a selection model added to the BNP GCM to simultaneously handle longitudinal nonnormal and missing data. Using the selection model approach, the proposed method can handle either the ignorable or non-ignorable missing data.

Chapter 4

Nonparametric Bayesian GCM with Missing Data

Different missingness mechanisms frequently arise in longitudinal data (Enders, 2011a). Nonparametric Bayesian growth curve modeling can be more powerful with the ability to handle missing data with different mechanisms. The MCMC algorithm in Bayesian modeling naturally embeds the MI technique to handle the ignorable missing data. Non-ignorable missing data are more challenging to handle as even modern methods such as MI or FIML provide inconsistent parameter estimates (Demirtas and Schafer, 2003; Diggle and Kenward, 1994).

This chapter introduces a selection model added to the BNP GCM. The new model is expected to handle either the ignorable or non-ignorable missing data in longitudinal studies and provide consistent parameter estimates, in addition to its ability to handle nonnormal data. Chapter 4 first discusses MCMC-based multiple imputation of the ignorable missing data in BNP GCM. The second part of Chapter 4 develops a selection model added to the BNP GCM to handle the non-ignorable missing data.

4.0.0.1 Ignorable Missingness Treatment

This subsection discusses MCMC-based imputation of the ignorable missing data in nonparametric Bayesian GCM. The MI technique is embedded in the MCMC algorithm in Bayesian estimation and is used to handle the ignorable missing data. MI is most directly motivated from the Bayesian perspective (Little and Rubin, 2019).

MCMC methods draw pseudo-random numbers from a probability distribution, the idea of which is equivalent to MI in that MI fills in copies of data with different estimates of missing values. Using the Gibbs sampling algorithm, the nonparametric Bayesian GCM contains two steps to impute missing data: the imputation step (I-step) and the posterior step (P-step) (Schafer, 1997).

The imputation I-step can be represented as

$$Y_{mis(t+1)} \sim P(Y_{mis}|Y_{obs}, \boldsymbol{\theta}(t)), \tag{4.1}$$

where Y_{mis} and Y_{obs} denote the missing and observed values in the variables, respectively; t is the current iteration step, and $\boldsymbol{\theta}$ represent all the estimated parameters in the nonparametric Bayesian GCM. Equation 4.1 shows that the missing variables at the next stage $t+1$ are sampled from the conditional distribution of the missing variables, conditional on the observed variables and the estimated model parameters at the current iteration stage t. In the imputation I-step, the model parameters are treated as known.

Following the preceding I-step, the posterior P-step can be expressed as

$$\boldsymbol{\theta}_{t+1} \sim P(\boldsymbol{\theta}|Y_{obs}, Y_{mis(t+1)}). \tag{4.2}$$

Equation 4.2 means that model parameters $\boldsymbol{\theta}$ in the nonparametric Bayesian GCM are estimated at the next iteration stage $(t+1)$ by sampling from the posterior predictive distribution of $\boldsymbol{\theta}$, given the observed variables and the imputed missing values in the imputation I-step. By repeating the two steps, a Markov chain is created as $(Y_{mis_1}, \boldsymbol{\theta}_1)$, $(Y_{mis_2}, \boldsymbol{\theta}_2)$, ..., which converges in distribution to $P(Y_{mis}, \boldsymbol{\theta}|Y_{obs})$ after a sufficient number of iterations (Geman and Geman, 1984; Schafer, 1997). Gibbs sampling in the Bayesian framework combines MI with DA to obtain the final single parameter estimates.

4.0.0.2 Non-ignorable Missingness Treatment

Under ignorable missingness, both FIML and MI yield unbiased parameter estimates and are widely recommended to use (Allison, 2000; Newman, 2014; Rubin, 2004; Schafer and Olsen, 1998; Peugh and Enders, 2004; Pigott, 2001). In spite of this, the results produced by FIML and MI are biased when data are MNAR (Arbuckle et al., 1996; Muthén et al., 1987, 2011). In order to produce unbiased results under MNAR mechanism, joint models are needed. The following part proposes a selection model as an add-on to the

nonparametric Bayesian growth curve modeling so that both the ignorable and non-ignorable missing data can be handled in a comprehensive manner.

As introduced in Chapter 3, the selection model is a type of joint model to analyze the MNAR data. The idea of selection models is to jointly model two parts in one model, the marginal distribution of the outcome data, and the conditional distribution of the missingness indicator given the data. Equation 3.1 in Chapter 3 is one form of selection models (Diggle and Kenward, 1994). Best et al. (1996) proposed a simplified selection model where $\boldsymbol{\alpha} \equiv \mathbf{0}$, so the model assumes that the missingness in $\boldsymbol{y}$ is only related to the missingness itself and not related to any auxiliary variables.

Following the work by Best et al. (1996), a simplified selection model is added to the BNP GCM to analyze the MNAR data. A nonparametric Bayesian selection GCM is developed and named as nonparametric Bayesian selection growth curve model (BNP selection GCM). In addition to handle the longitudinal nonnormal data, the BNP selection GCM simultaneously deals with the ignorable or non-ignorable missing data.

Specifically, in order to address the non-ignorable missing data, a binary missingness indicator variable is proposed to be created. Because the missingness indicator is binary, we then used a probit model to build an added-on selection model structure to the existing modeling framework. In this way, the missingness in the data can be explained and the non-ignorable missing data problem is turned into an ignorable missing data problem.

A BNP selection GCM is constructed and presented as Equations 2.1, 2.2 and 4.3.

$$\begin{aligned} R_{it} &\sim Bernoulli(q_{it}), \\ q_{it} &= \mathbf{\Phi}(\tau_{0t} + \tau_{1t} b_{Li} + \tau_{2t} b_{Si}), \end{aligned} \tag{4.3}$$

where R_{it} is a missingness indicator for the ith observation at time t (1 if missing and 0 otherwise), and q_{it} is the probability for the ith observation to be missing at time t. In the newly developed model, a probit link function is added to the BNP GCM to model the missingness indicator, so that the probability of missingness is explained by the latent intercept and slope b_{Li} and b_{Si}. In a linear GCM, latent intercept and slope are often the key to understanding growth patterns.

By taking a simplified selection model approach, unnecessary auxiliary variables could be avoided (Best et al., 1996; Shi and Tong, 2020; Zhang et al., 2012). In longitudinal studies or clinical trials, it is not uncommon

that data being missing is only or majorly related to the missingness itself but nothing else (e.g., Enders, 2011b; Karahalios et al., 2012). For example, in psychological assessments or achievement tests, participants may miss the session because they had a low starting score and were demotivated later in the process, or they developed slower in the ability than peers and did not want to continue later sessions. In addition, Zhang et al. (2012) concluded through simulation studies that even when missingness originates from some auxiliary or unobserved variables other than the missingness itself, the simplified selection model is still able to well recover parameter estimates.

This chapter proposes a BNP selection GCM (Equations 2.1, 2.2 and 4.3) to handle missing and nonnormal data simultaneously in longitudinal studies. Note that as discussed previously, because DP is an infinite-dimensional generalization of the Dirichlet distribution and the nonparametric Bayesian GCM uses DP mixture as the prior for intraindividual measurement errors, the BNP selection GCM is essentially a type of longitudinal infinite mixture model. The next chapter develops Bayesian model evaluation criteria in the framework of longitudinal infinite mixture models.

Chapter 5

Model Selection in Bayesian Modeling

In previous chapters, a BNP selection GCM was proposed to simultaneously handle missing and nonnormal data in longitudinal studies. A natural question may rise: does the proposed model fit data better than the traditional normal-distribution-based GCM or the BNP GCM, and which model has the best fit to the data? The performance of the new model will be systematically compared with the other two GCMs in the simulation study. In this chapter, we first discuss model fit.

Although there are abundant studies on Bayesian model evaluations in general, such as Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), posterior predictive p value (ppp, Gelman et al., 1996; Meng et al., 1994) and Bayes factor (BF; Kass, 1993; Kass and Raftery, 1995; Raftery, 1999), literature and guidelines on how to select mixture models in the Bayesian context are in a smaller quantity. The proposed BNP selection GCM is a type of infinite mixture model. However, model evaluations on Bayesian infinite mixture models are rarely studied. This chapter develops Bayesian model selection criteria to select best fitting models in longitudinal infinite mixture models.It also discusses Bayesian model selection criteria previously developed for finite mixture models. The following chapter further investigates the performance of those Bayesian model selection criteria in the infinite mixture context in the simulation study.

5.0.1 DIC

As an information criterion (IC) based measure, DIC is one of the major model selection criteria in the Bayesian literature and has been widely used to select Bayesian models (e.g., Hsu et al., 2015; Shriner and Yi, 2009; Shi and Tong, 2017). DIC was introduced by Spiegelhalter et al. (2002) to compare the relative fit in Bayesian models. It was developed from the IC family for assessing and comparing hierarchical models in Bayesian methods. It can be seen as a generalization of the Akaike Information Criterion in a Bayesian context (e.g., Gelman et al., 2013).

Like all other measures in the IC family, DIC consists of two parts, a measurement of model fit and a measurement of model complexity. A trade-off in these two quantities is usually necessary when comparing models, as increasing model complexity is usually associated with a better fit. While DIC has been used extensively for Bayesian model comparisons (e.g., Hsu et al., 2015; Shriner and Yi, 2009; Shi and Tong, 2017), applying DIC directly to mixture models is problematic. When it comes to compare and select mixture models in a Bayesian context, it has been unclear which form of DIC to use (Celeux et al., 2000, 2006; Lu et al., 2015; Plummer, 2008). Methodological discrepancies exist in defining DIC for mixture models (Celeux et al., 2000; DeIorio and Robert, 2002) or mixture models with missing data (Celeux et al., 2006). Below first introduces DIC and then discusses methodological challenges of DIC in mixture models in more detail.

DIC tries to balance between the fit and parsimony of models. In the model fit part, the computation of DIC is based on the posterior distribution of log-likelihood from a MCMC analysis. The measurement of fit in DIC is represented by a deviance $D(\theta)$ and is defined as

$$D(\boldsymbol{\theta}) = -2log(f(\boldsymbol{y}|\boldsymbol{\theta})) + 2log(g(\boldsymbol{y})),$$

where $f(\cdot)$ represents the likelihood function of the model, $\boldsymbol{y}$ is the data and $\boldsymbol{\theta}$ is a vector of parameters. $g(\boldsymbol{y})$ depends only on the data and can be seen as a constant. As all other measures in the IC family, DIC has an additional component to penalize the model complexity, p_D, as the effective number of parameters in a model Celeux et al. (2006). p_D is defined as

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}}),$$

where $\overline{D(\theta)}$ is the posterior mean deviance and can be represented as

$$\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta}}[-2log(f(\boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{y})] + C,$$

where C is a constant. $\tilde{\boldsymbol{\theta}}$ is an estimate of the parameters $\boldsymbol{\theta}$ depending on the data $\boldsymbol{y}$. (Spiegelhalter et al., 2002) defined $D(\tilde{\boldsymbol{\theta}})$ as the deviance evaluated at the posterior mean of the model parameters. Taking the two parts together, DIC is defined as

$$\begin{aligned} DIC =& \overline{D(\boldsymbol{\theta})} + p_D \\ =& D(\tilde{\boldsymbol{\theta}}) + 2p_D \\ =& 2\overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}}). \end{aligned} \tag{5.1}$$

5.0.2 Criteria used for Finite Mixture Models

It is a methodological challenge to have a clear guideline on model selection criteria in a Bayesian context, particularly in mixture models (Celeux et al., 2006). Lu et al. (2015) developed model selection criteria in Bayesian growth mixture models and found that they can correctly identify true finite mixture models. Lu et al. (2015) studied two traditional versions of deviance statistics, $\overline{D(\boldsymbol{\theta})}$ and $D(\tilde{\boldsymbol{\theta}}_{mean})$, with penalty terms corresponding to AIC, BIC, $CAIC$ and $SABIC$, and found that both versions of deviance statistics perform generally well in finite growth mixture models with missing data in that they recovered the true number of latent classes.

This study uses the criteria developed by Lu et al. (2015) (hereafter named as FMM criteria) to study their performances in nonparametric Bayesian growth curve modeling, which is a longitudinal infinite mixture model. Our hypothesis is that the FMM criteria can perform equally well in longitudinal finite or infinite mixture models. Lu et al. (2015) uses the two traditional versions, $\overline{D(\boldsymbol{\theta})}$ and $D(\tilde{\boldsymbol{\theta}}_{mean})$, as the deviance statistics. The choice of penalty terms derives from the IC family. Particularly, in addition to using the traditional effective dimension, p_D, penalty components associated with AIC, BIC, $CAIC$ and $SABIC$ are studied. Akaike (1974, 1998) proposed Akaike Information Criterion (AIC), whose measure is based on minimizing an expected information loss to select the most parsimonious and correct model. AIC is expressed as

$$AIC = -2LL + 2 \times k,$$

where $-2LL$ is the -2 times log-likelihood of the estimated model and k is the number of estimated parameters in the model. The first part is used to fit the model and the second part serves as a "penalty" term to prevent overfitting of the model. The Bayesian Information Criterion (BIC, Schwarz

et al., 1978) is similar to AIC, but with a different penalty for the number of parameters. While AIC consistently adds a penalty of two times for every parameter being estimated, the penalty in BIC increases with sample size. BIC is expressed as

$$BIC = -2LL + log(N) \times k,$$

where N represents sample size and other terms are the same as those in AIC. BIC was developed to also approximate the Bayes factor, as the computation of the Bayes factor is difficult particularly in hierarchical models. Other modification indices in the IC family include the consistent AIC (CAIC, Schwarz et al., 1978) and sample size adjusted BIC (SABIC, Sclove, 1987)), which are widely used and reported in major SEM software (e.g., Muthén and Muthén, 2004; Neale et al., 2016). They are similar in mathematical representations but differ in penalty components. They are all based on the log likelihood and a penalty to the number of model parameters. CAIC uses $(log(N) + 1) \times k$ as the penalty for models with large number of parameters while $SABIC$ applies $(log((N + 2)/24)) \times k$ to penalize overfitting.

5.0.3 New (DIC-variant) Criteria

This section proposes four model evaluation criteria for longitudinal infinite mixture models. The new criteria derive from the DIC measurement (hereafter named as DIC-variant criteria) and can be seen as a part of the IC family.

Spiegelhalter et al. (2002) constructed the form of DIC (Equation 5.1) and focused mostly on selecting generalized linear models. This current form of DIC is less than ideal to select mixture models. For some hierarchical models, the conditional likelihood used in DIC can be invalid (Millar, 2009). In addition, DIC uses posterior mean to compare models but posterior mean is not meaningful in mixture models as their parameters are not from a unimodal distribution (Lunn et al., 2012).

Challenges of DIC in mixture models come from both the deviance part and the penalty part. On one hand, the deviance part has at least two challenges. The first challenge relates to the likelihood function. When missing data or latent variables are in the models, DICs can be associated with observed likelihood, complete likelihood or conditional likelihood (Celeux et al., 2006). Different likelihood functions lead to different representations of DICs

and computational complexity in DICs can thus vary. Second, as mentioned above, there are two possible versions of deviance statistics (i.e., $\overline{D(\boldsymbol{\theta})}$ and $D(\tilde{\boldsymbol{\theta}})$) in the Bayesian context to compute parameters in the Monte Carlo markov chains (Lu et al., 2015). In specific, $\overline{D(\boldsymbol{\theta})}$ represents the posterior mean deviance for each MCMC iteration, whereas $D(\tilde{\boldsymbol{\theta}})$ is the deviance of the average of the estimate $\tilde{\boldsymbol{\theta}}$. Note that Celeux et al. (2006) and Richardson (2002) explicitly pointed out that $\tilde{\boldsymbol{\theta}}$ does not necessarily need to be the posterior mean and thus alternative choices of posterior estimates are possible. The choice of the estimate $\tilde{\boldsymbol{\theta}}$ will affect the computation of deviance, which is central to the DIC criterion. For example, in the context of mixture models, posterior distributions of parameters in mixture models inevitably involve multimodal distributions. Merely choosing the posterior mean as the estimate of the posterior distributions may fail to fully represent the true mixture distributions, and thus affect the deviance statistics and eventually harm the performance of model selection measures.

On the other hand, the penalty part, which is the effective number of parameters p_D, is unreliable and depends on how the estimate $\tilde{\boldsymbol{\theta}}$ is chosen. As pointed out directly by Spiegelhalter et al. (2002), the fact that the effective dimension p_D depends on the choice of the estimate $\tilde{\boldsymbol{\theta}}$ makes it difficult to have an intrinsic definition to the dimension of p_D in DIC. Lunn et al. (2012) further argued that p_D is hard to calculate when model parameters contain categorical variables, where calculating posterior means is not meaningful.

Consider the above reasons, it is necessary to propose and evaluate new Bayesian model selection criteria in the mixture model context. This chapter specifically develops four model selection measures based on DIC and other criteria from the IC family. Particularly, variations of the two challenging and essential components in DIC - deviance statistics and penalty - are proposed. Each of the four proposed measure will be calculated for the three aforementioned growth curve modeling and results of each measure will be compared in the simulation study in the next chapter.

Due to the nature and shapes of posterior distributions in mixture models, which unavoidably have multi-modal distributions, posterior means do not fully capture information about the multi-modal distributions in the posterior. Using posterior means as the point estimate is what most traditional DIC does and is a major reason why DIC fails in accurately selecting mixture models. On the contrary, considering the posterior median as the point estimate of parameters for the mixtures models is reasonable. With different latent classes, mixture models are likely to experience with bi-modal or

Table 5.1: DIC BASED MODEL SELECTION CRITERIA

	Criterion (Index)	Deviance	Penalty
DIC-variant Criteria	Dmedian.AIC	D_{mdn}	$2 \times k$
	Dmedian.BIC	D_{mdn}	$log(N) \times k$
	Dmedian.SABIC	D_{mdn}	$log((N+2)/24) \times k$
	Dmedian.CAIC	D_{mdn}	$(log(N)+1) \times k$
DIC	Dbar.DIC	$\overline{D(\theta)}$	pD
FMM Criteria	Dbar.AIC	$\overline{D(\theta)}$	$2 \times k$
	Dbar.BIC	$\overline{D(\theta)}$	$log(N) \times k$
	Dbar.SABIC	$\overline{D(\theta)}$	$log((N+2)/24) \times k$
	Dbar.CAIC	$\overline{D(\theta)}$	$(log(N)+1) \times k$
	Dhat.AIC	$D(\hat{\theta})$	$2 \times k$
	Dhat.BIC	$D(\hat{\theta})$	$log(N) \times k$
	Dhat.SABIC	$D(\hat{\theta})$	$log((N+2)/24) \times k$
	Dhat.CAIC	$D(\hat{\theta})$	$(log(N)+1) \times k$

multi-modal posterior distributions. Unlike the mean statistic, which is sensitive to each single point, the median statistics are less sensitive to one single point and better describes a point that appears more than once. Therefore, posterior median may summarize the multi-modal distributions better than posterior means.

In the newly developed criteria, the deviance statistic is proposed as the average of posterior median and denoted as $D(\tilde{\boldsymbol{\theta}}_{mdn})$. It is defined as the value of the estimate that separates the higher half of estimated values in the posterior distribution from the lower half across iterations. Finding a balance between model complexity and model fit is critical when developing and choosing IC based measures. The penalty parts in the new criteria are proposed to be $2 \times k$, $log(N \times k$, $log(\frac{N+2}{24}) \times k$ and $log(N+1) \times k$, where N is sample size and k represents effective model parameters. The effective model parameter k is be on the same level as the value of likelihood. The proposed penalty part corresponds to penalty terms in AIC, BIC, $CAIC$ and $SABIC$, respectively. Table 5.1 is a list of different DIC based model selection criteria.

The current study differs from Lu et al. (2015)'s study in several ways. First, the proposed BNP selection GCM is a type of longitudinal infinite mixture models, where the FMM criteria have not been studied in the in-

finite mixture framework before. Second, the study extended the deviance statistics, namely, using the posterior median to reflect multi-modal shapes in posterior distributions. Only using the posterior mean as the parameter estimate in the deviance statistic can be inconclusive to select mixture type models. As illustrated in Lu et al. (2015), $\overline{D(\boldsymbol{\theta})}$ based criteria performed slightly better than $D(\tilde{\boldsymbol{\theta}})$ based criteria in mixture data. Third, the new selection criteria are evaluated differently from the previous study. Although an infinite mixture model, the focus of the BNP selection GCM is still the estimate of parameters of interest (e.g., the growth parameter). Therefore, unlike evaluating whether the FMM criteria can correctly identify number of latent classes as in the previous study, the current study compares among the traditional normal-distribution-based GCM, BNP GCM and the BNP selection GCM in terms of parameter estimation.

Chapter 6

A Simulation Study

The study develops a nonparametric Bayesian selection growth curve model (BNP selection GCM). The proposed model is a type of infinite mixture model and can simultaneously handle longitudinal nonnormal and missing data with the ignorable or non-ignorable missing data. The study further develops Bayesian model selection criteria to select the best fitting model in the context of longitudinal infinite mixture modeling. The performance of the BNP selection GCM is compared to two other GCMs (i.e., the traditional normal-distribution-based GCM and BNP GCM) and evaluated through a Monte Carlo simulation study.

6.0.1 Data Conditions

In the simulation, six potentially influential factors are studied (see Table 6.1), including types of nonnormal data, missing data mechanisms, missing data rates, covariances between the latent growth parameters (latent intercept and slope) and an auxiliary variable and variance of intraindividual measurement errors.

Longitudinal data are generated based on an unconditional linear GCM with four measurement occasions. Although the unconditional GCM is used, conclusions can be easily generalized to conditional models with multiple covariates. Population fixed effects values are prespecified a priori as $\boldsymbol{\beta} = (\beta_{00}, \beta_{10})' = (6, 0.3)'$ and $\boldsymbol{\Psi} = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{SL} & \sigma_S^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.1 \end{pmatrix}$. Details of six influential factors are explained as below.

First, three sample sizes are considered ($N = 60$, 200, *and* 600). Second,

Table 6.1: DESIGN OF INFLUENTIAL FACTORS IN THE SIMULATION

Influential Factors	# of factor levels	Levels
Sample Size (N)	3	60, 200, 600
Measurement Error Variance (σ_e^2)	2	0.5, 0.7
Missingness Mechanism	2	Ignorable; Non-ignorable
Strength(Missing, Auxiliary) (r)	2	0.4, 0.8
Missing Rate (mr)	3	0, 18%, 36%
Nonnormal Data (nn)	3	Normal data; Data with outliers; Nonnormally distributed data

the variance of intraindividual measurement errors is manipulated to be 0.5 or 0.7 to study the influence of intraindividual measurement errors and the reliability of the longitudinal outcome. In the current setting, the reliability ranges from 0.50 to 0.96.

Third, the MAR and MNAR missing data mechanisms are examined. All data at the first measurement occasion are complete and may miss in following measurement occasions. To resemble real world settings, number of missing values increases from the second measurement occasion on as $(1-(1-mr_t)^{(t-1)})$, where mr_t is a prespecified missing rate. For example, suppose $mr_t = 12\%$, the missing rate in time points two, three and four will be around 12%, 24% and 32%, respectively, which will result in a total missing rate of around 18% for the entire observations. A cutoff score is predetermined for measurement occasion t as the $(1-mr_t/\kappa)$th quantile of a reference variable, where κ is a probability threshold for missing values ($\kappa = 0.8$ in this design). For each measurement occasion, $100 \times \kappa\%$ of data larger than the cutoff score are missing. This way sets the missing probability function as a step function and guarantees that the desired missing rate is generated. Depending on the choice of the reference variable, the generated missing data are MAR or MNAR. Specifically, data are MAR when the reference variable is the first measurement occasion variable. Data are MNAR if the reference variable is chosen to be an auxiliary variable, which can help explain the missingness but will not be included in the model. An auxiliary variable is generated for each individual as $z_i = \gamma \times b_{1i} + \epsilon_i$, where γ is the correlation between the latent slope parameter and the auxiliary variable. ϵ_i is generated from a normal distribution $N(0, \sqrt{1-\gamma^2})$. Fourth, to investigate further about missing data, the covariance between the latent slope parameter and the auxiliary variable γ is manipulated to be 0.4 or 0.8, reflecting strength between the missingness and latent traits. Fifth, the missing rate for the entire data MR

is manipulated to be 0, 18% or 36%, where the corresponding mr_t will be 0, 12% and 24%, respectively. .

Sixth, normally-distributed data and two types of nonnormal data (data with outliers and nonnormally-distributed data) are generated on the measurement errors of the model. The outliers are generated from the same distributions as the normally-distributed data, except that the mean is six standard deviations away from the normal data mean, with outlier proportion being 10%. For the nonnormally-distributed data, the third and fourth moments (skewness=7 and kurtosis=110) are added to the same distribution as the normally-distributed data to form the multivariate nonnormal distribution. The error terms in the linear GCM are generated from this multivariate nonnormal distribution.

6.0.2 Three Bayesian GCMs

The three Bayesian GCMs are fitted to each data condition, including the traditional normal-distribution-based GCM (see Equation 2.1), BNP GCM (see Equations 2.1 and 2.2) and BNP selection GCM (see Equations 2.1, 2.2 and 4.3), resulting in a total of 540 conditions (432 conditions with missing data and 108 conditions with complete data).

Bayesian methods are used for all model estimations. For each condition, 100 data sets are generated and analyzed using software OpenBUGS (Spiegelhalter et al., 2003) and R (Team et al., 2013). For prior distributions of parameters for the three Bayesian models (see Table 6.2), informative priors of multivariate normal distributions are used for parameters including latent intercept β_L, latent slope β_S and coefficients for the selection model $\boldsymbol{\tau}$. Precision of the random effects errors follows a Wishart distribution. For the variances of intraindividual measurement errors, DP mixture priors are used of the two nonparametric Bayesian approaches whereas the inverse Gamma distribution is used for the normal-distribution-based GCM. In the nonparametric Bayesian approaches, the concentration parameter α controls the dispersion of the random probability measure G_e and is assigned an informative Gamma distribution with location and scale hyperparameters as 100.

Table 6.2: PRIOR DISTRIBUTIONS FOR THREE BAYESIAN GCMS

	Normal-based GCM	BNP GCM	BNP Selection GCM
β_L, β_S	Multivariate Normal[1]	Multivariate Normal	Multivariate Normal
$\boldsymbol{\Phi}$	inverse Gamma[2]	DP Mixture[3]	DP Mixture
$1/\boldsymbol{\Psi}$	Wishart[4]	Wishart	Wishart
α	-	Gamma(100,100)[5]	Gamma(100,100)
$\boldsymbol{\tau}$	-	-	Multivariate Normal

Note:
[1]Multivariate normal represents the multivariate normal distribution with hyperparameters $(6, 1.0E{+}2)$ and $(0.3, 1.0E{+}2)$ for latent intercept and slope, respectively.
[2]Precision of the measurement errors $1/\boldsymbol{\Phi}$ follows a Gamma distribution with hyperparameters $(0.001, 0.001)$.
[3]DP mixture represents the Dirichlet process mixture priors.
[4]Precision of the random effects errors $1/\boldsymbol{\Phi}$ follows a Wishart distribution.
[5]Concentration parameter α follows a Gamma distribution with hyperparameters $(100, 100)$.

6.0.3 Evaluations

The performance of the proposed modeling framework is evaluated through the performance of the parameter estimates. Specifically, the relative bias, standard error (SE) and mean squared error (MSE) for the six major parameters $\boldsymbol{\theta} = (\beta_L, \beta_S, \sigma_L^2, \sigma_S^2, \sigma_{LS}, \sigma_e^2)$ in GCM are obtained and compared. Let $\hat{\theta}_r$ denote its estimate from the rth simulation replication and let $\hat{s}_r$ denote the posterior standard deviation of the parameter estimate from the rth simulation replication, $r = 1, \ldots, 100$. The replication estimate is calculated as the average of parameter estimates over 100 simulation replications,

$$Estimate = \frac{1}{100}\Sigma_{r=1}^{100}\hat{\theta}_r.$$

Absolute relative bias captures the absolute value of the relative distance between the replication estimate and its population value, and is defined as

$$Absolute\,relative\,bias = \begin{cases} absolute((\frac{Estimate}{\theta} - 1)) \times 100 & \theta \neq 0, \\ absolute(Estimate) \times 100 & \theta = 0. \end{cases}$$

The SD is the average posterior standard deviation and defined as

$$SD = \frac{1}{100}\Sigma_{r=1}^{100}\hat{s}_r.$$

MSE is a combined measure of the bias and standard deviation and defined as

$$MSE = E[(\frac{1}{100}\Sigma_{r=1}^{100}\hat{\theta}_r - \theta)^2].$$

For each condition, the thirteen model selection criteria discussed in Chapter 5 (see also Table 5.1) are calculated. The proportion of the correct model being selected is calculated from the 100 replications to evaluate the performance of the model evaluation criteria in this study.

All computations use MCMC methods with 40,000 iterations for the normal-distribution-based GCM, 80,000 iterations for the BNP GCM and 100,000 iterations for the BNP selection GCM, with the first half of iterations in all conditions as the burn-in period. The model convergence is assessed for each simulation replication using Geweke tests (Geweke, 1991). The percentage of the converged replications for each simulation condition is reported.

6.0.4 Results

6.0.4.1 Parameter Estimation

This part compares and evaluates the performance of the three Bayesian GCMs in parameter estimation. The absolute relative bias, posterior SDs, MSEs and convergence rates are reported for the six major parameters (two fixed effects and four random effects) in growth curve modeling. Geweke tests (Geweke, 1991) are used for convergence diagnostics. All results reported in the tables are based on converged estimations. Main results are summarized and presented in this section. The complete results of parameter estimates for the complete and ignorable missing data are available in Appendix A and for the non-ignorable missing data are available in Appendix B , respectively, on this webpage
https://drive.google.com/drive/folders/1FvNKGUt54k6VzCgY47tC9SBHSjre6p13

In general, the BNP selection GCM and BNP GCM may outperform the normal-distribution-based GCM, and BNP selection GCM may outperform BNP GCM. This is as expected because BNP is used to handle nonnormal data and the added-on selection model is used to explain the non-ignorable missingness. Although the relative performance of the three models has a similar pattern across conditions, specific results and conclusions vary for normal data, nonnormally distributed data, and data containning outliers. The detailed results are discussed below. Hereafter nonnormal data refer to

data generated from nonnormal distributions and outlier data refer to data being contaminated with outliers.

6.0.4.1.1 Normal Data When data are complete and normally distributed, the three models perform almost equally well. Table 6.3 presents the absolute relative bias, posterior SDs, MSEs and convergence rates for six parameter estimates (β_L, β_S, σ^2_L, σ^2_S, σ_{LS}, σ^2_e) when $N = 600$ and $\sigma^2_e = 0.5$. The three models yield similar absolute relative bias, posterior SDs and MSEs for almost all parameter estimates. The BNP selection GCM has a higher absolute relative bias of the estimated σ^2_e than the other two GCMs. The results display similar patterns with smaller sample sizes or larger measurement errors. See Table 6.0.4.1.1 for the results when $N = 60$ and $\sigma^2_e = 0.5$ and Table 6.5 for the results when $N = 60$ and $\sigma^2_e = 0.7$.

When data are normal and include missing values, similar patterns are observed. For example, as shown in Table 6.6, for the normal and ignorable missing data when $N = 200$, $\sigma^2_e = 0.5$ and $mr = 0.36$, the three GCMs provide similar absolute relative bias, posterior SDs, MSEs and convergence rates of the five parameter estimates of the GCMs, except that the absolute relative bias of the estimated σ^2_e from the BNP selection GCM is slightly higher than the other two GCMs. For the normal and non-ignorable missing data, Table 6.0.4.1.1 shows the absolute relative bias, posterior SDs, MSEs and convergence rates of the six parameter estimates when $N = 200$, $r = 0.8$, $mr = 0.18$ and $\sigma^2_e = 0.7$. There is no obvious difference among three GCMs for five model parameters (β_L, β_S, σ^2_L, σ^2_S and σ_{LS}). In terms of σ^2_e, the absolute relative bias and MSEs of the estimated σ^2_e from the BNP selection GCM is larger than those from the other two GCMs. The posterior SD of the estimated σ^2_e from the BNP GCM is consistently smaller than those from the normal-distribution-based GCM and the BNP selection GCM. In practice, the measurement error σ^2_e is typically not what researchers are interested in so the BNP selection GCM can be a good alternative model to analyze the normal data with missing values.

6.0.4.1.2 Nonnormal Data Next, we evaluate the performance of the three models when data are nonnormally distributed. Across all conditions, the three GCMs provide similarly unbiased and efficient estimates to the latent intercept parameter β_L. This is probably because the second level random effects parameters were generated from multivariate normal distri-

Table 6.3: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NORMAL AND COMPLETE DATA WHEN $N = 600$, $\sigma_e^2 = 0.5$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.045	0.002	1.000	0.001	0.045	0.002	0.970	0.001	0.045	0.002	0.930
β_S	0.003	0.020	0.000	1.000	0.003	0.020	0.000	0.940	0.001	0.020	0.000	0.910
σ_L^2	0.061	0.087	0.010	0.980	0.076	0.088	0.012	0.920	0.065	0.088	0.013	0.940
σ_S^2	0.277	0.013	0.001	0.970	0.273	0.013	0.001	0.950	0.273	0.013	0.001	0.930
σ_{LS}	0.150	0.023	0.003	1.000	0.151	0.024	0.002	0.940	0.159	0.024	0.003	0.920
σ_e^2	0.355	0.024	0.032	1.000	0.350	0.015	0.031	0.940	0.598	0.020	0.090	0.900

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.4: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NORMAL AND COMPLETE DATA WHEN $N = 60$, $\sigma_e^2 = 0.5$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.084	0.002	1.000	0.001	0.085	0.002	0.930	0.001	0.085	0.002	0.980
β_S	0.021	0.059	0.001	1.000	0.024	0.059	0.001	0.950	0.017	0.058	0.001	0.980
σ_L^2	0.142	0.290	0.099	1.000	0.139	0.291	0.091	0.940	0.100	0.285	0.078	0.930
σ_S^2	0.931	0.052	0.010	0.920	0.936	0.052	0.010	0.910	0.891	0.051	0.009	0.970
σ_{LS}	0.298	0.084	0.016	1.000	0.234	0.084	0.012	0.920	0.296	0.082	0.014	0.950
σ_e^2	0.300	0.076	0.027	0.440	0.313	0.049	0.030	0.940	0.640	0.071	0.109	0.960

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.5: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NORMAL AND COMPLETE DATA WHEN $N = 60$, $\sigma_e^2 = 0.7$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.000	0.085	0.002	1.000	0.000	0.086	0.002	0.980	0.001	0.086	0.002	0.950
β_S	0.010	0.061	0.003	1.000	0.011	0.060	0.003	0.980	0.020	0.060	0.003	0.960
σ_L^2	0.174	0.310	0.123	1.000	0.195	0.315	0.138	0.970	0.164	0.309	0.127	0.940
σ_S^2	0.971	0.055	0.011	0.920	0.949	0.054	0.011	0.950	0.955	0.054	0.011	0.900
σ_{LS}	0.334	0.088	0.017	1.000	0.297	0.088	0.013	0.920	0.329	0.087	0.015	0.960
σ_e^2	0.116	0.091	0.016	0.470	0.107	0.057	0.015	0.890	0.341	0.080	0.068	0.970

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.6: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NORMAL AND IGNORABLE-MISSING DATA WHEN $N = 200$, $\sigma_e^2 = 0.5$ AND $mr = 0.36$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.000	0.066	0.003	0.950	0.001	0.067	0.003	0.930	0.001	0.067	0.003	0.960
β_S	0.061	0.049	0.002	0.970	0.063	0.049	0.002	0.970	0.064	0.049	0.003	0.900
σ_L^2	0.063	0.162	0.027	0.870	0.075	0.164	0.028	0.930	0.082	0.165	0.027	0.900
σ_S^2	0.794	0.040	0.007	0.940	0.783	0.039	0.007	0.930	0.776	0.039	0.007	0.950
σ_{LS}	0.267	0.057	0.009	0.980	0.264	0.058	0.009	0.930	0.260	0.058	0.009	0.880
σ_e^2	0.306	0.060	0.028	0.920	0.310	0.053	0.028	0.920	0.659	0.063	0.114	0.930

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.7: PARAMETER ESTIMATION FOR THREE GCMS WITH NORMAL AND NON-IGNORABLE MISSING DATA WHEN $N = 200$, $r = 0.8$, $mr = 0.18$ AND $\sigma_e^2 = 0.7$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.067	0.002	0.950	0.001	0.067	0.002	0.940	0.001	0.067	0.002	0.960
β_S	0.115	0.041	0.003	0.900	0.102	0.041	0.002	0.930	0.101	0.041	0.002	0.940
σ_L^2	0.088	0.165	0.027	0.970	0.098	0.168	0.032	0.900	0.093	0.167	0.029	0.950
σ_S^2	0.637	0.032	0.005	0.930	0.640	0.032	0.005	0.940	0.623	0.032	0.005	0.920
σ_{LS}	0.244	0.051	0.007	0.990	0.245	0.051	0.008	0.920	0.245	0.050	0.007	0.960
σ_e^2	0.123	0.057	0.010	0.880	0.130	0.043	0.011	0.950	0.335	0.052	0.058	0.870

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.8: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NONNORMAL AND COMPLETE DATA WHEN $N = 600$ AND $\sigma_e^2 = 0.7$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.046	0.002	1.000	0.008	0.040	0.004	0.960	0.008	0.041	0.004	0.970
β_S	0.001	0.022	0.000	1.000	0.050	0.016	0.000	0.940	0.048	0.016	0.000	0.910
σ_L^2	0.050	0.092	0.013	0.990	0.002	0.067	0.004	0.920	0.015	0.068	0.004	0.960
σ_S^2	0.389	0.015	0.002	0.950	0.140	0.009	0.000	0.900	0.144	0.009	0.000	0.930
σ_{LS}	0.144	0.025	0.004	1.000	0.045	0.020	0.001	0.900	0.034	0.020	0.000	0.880
σ_e^2	0.171	0.029	0.042	1.000	0.220	0.036	0.050	0.850	0.388	0.043	0.105	0.860

butions and the latent intercept parameter β_L was unrelated to the missingness in the data. Even the normal-distribution-based GCM is sufficient to analyze β_L. Therefore, the estimates of β_L will not be discussed in detail for the nonnormal data and outlier data conditions.

When nonnormal data are completely observed, the two nonparametric Bayesian GCMs in general outperform the traditional normal-distribution-based GCM in parameter estimation. For example, Table 6.8 displays the absolute relative bias, posterior SDs, MSEs and convergence rates of the parameter estimates of the GCMs for the complete and nonnormal data when $N = 600$ and $\sigma_e^2 = 0.7$. The two nonparametric Bayesian GCMs lead to smaller absolute relative bias, posterior SDs and MSEs of the parameter estimates than the normal-distribution-based GCM. The convergence rates of the two nonparametric Bayesian GCMs are slightly lower than the normal-distribution-based GCM but still reasonably high (above 85%).

When nonnormal data have ignorable missing values, the BNP GCM and BNP selection GCM lead to less biased parameter estimates, smaller posterior SDs and smaller MSEs than the normal-distribution-based GCM. Table 6.0.4.1.2 illustrates the parameter estimation results from the three GCMs when $N = 200$, $\sigma_e^2 = 0.7$ and $mr = 0.36$. For the fixed effect latent growth parameter β_S, the absolute relative bias of β_S drops from 16.7% in the normal-distribution-based GCM to around 1.0%s in the two nonparametric Bayesian GCMs. The posterior SDs and MSEs from the nonparametric Bayesian GCMs are smaller than those from the normal-distribution-based GCM, respectively. For the random effects related parameters σ_L^2, σ_S^2 and σ_{LS}, the nonparametric Bayesian GCMs outperform the normal-distribution-based GCM in terms of the absolute relative bias, posterior SDs and MSEs.

Comparing between the BNP GCM and BNP selection GCM, the bias and efficiency of the fixed and random effect parameter estimates do not differ much. This is as expected because the data are missing at random. The MI technique in the Bayesian method handles the ignorable missingness and provides reliable parameter estimates. The outperformance of the two nonparametric Bayesian GCMs over the normal-distribution-based GCM shows similar patterns with larger sample sizes and smaller missing rates for the nonnormal and ignorable missing data. As shown in Table 6.0.4.1.2, for the nonnormal data when $N = 600$, $\sigma_e^2 = 0.7$ and $mr = 0.18$, the absolute relative bias, posterior SDs and MSEs are almost uniformly smaller from the nonparametric Bayesian GCMs than those from the normal-distribution-based GCM. Similarly, because the ignorable missingness is handled by the MI in the Bayesian estimation, the BNP GCM and BNP selection GCM perform similarly in parameter estimation.

When nonnormal data have non-ignorable missing values, the BNP selection GCM provides the most accurate and efficient estimates of four GCM parameters (β_S, σ_L^2, σ_S^2 and σ_{LS}) in almost all conditions. Table 6.0.4.1.2 reports the absolute relative bias, posterior SDs, MSEs and convergence rates of the six major parameter estimates for the nonnormal and non-ignorable missing data when $N = 200$, $r = 0.8$, $mr = 0.36$ and $\sigma_e^2 = 0.7$. The BNP selection GCM outperforms the other two GCMs by producing the smallest absolute relative bias, MSEs and posterior SDs to the four parameters (highlighted in blue). In specific, the absolute relative bias of β_S, σ_L^2, σ_S^2 and σ_{LS} are 0.140, 0.202, 1.296 and 0.422, respectively, from the normal-distribution-based GCM. They drop to 0.001, 0.004, 0.390 and 0.100, respectively, when applied to the BNP selection GCM. Similarly, the MSEs of the four parameters are the smallest from the BNP selection GCM. The nonparametric Bayesian GCMs provide smaller posterior SDs than the normal-distribution-based GCM. Note that estimates of the measurement errors σ_e^2 are 1.11, 1.06 and 1.30, respectively, from the three GCMs, which lead to large absolute relative bias and MSEs of σ_e^2. We notice that this pattern is similar to what is shown when data are normal and complete (e.g., see Tables 6.3, 6.0.4.1.1 or 6.5). In those conditions, the absolute relative bias and MSE of σ_e^2 are also higher from the BNP selection GCM than those from the other two GCMs. This may be due to the data generating mechanism of the nonnormal data. Future work can consider generating nonnormal data from other nonnormal distributions such as gamma distributions.

Table 6.9: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NONNORMAL AND IGNORABLE MISSING DATA WHEN $N = 200$, $\sigma_e^2 = 0.7$ AND $mr = 0.36$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.068	0.003	0.950	0.009	0.062	0.006	0.920	0.009	0.062	0.006	0.940
β_S	0.167	0.048	0.005	0.970	0.010	0.036	0.001	0.960	0.001	0.036	0.001	0.960
σ_L^2	0.230	0.180	0.231	0.970	0.014	0.127	0.012	0.890	0.001	0.126	0.010	0.910
σ_S^2	0.938	0.041	0.014	0.960	0.357	0.023	0.002	0.920	0.336	0.023	0.002	0.940
σ_{LS}	0.332	0.064	0.036	0.970	0.111	0.043	0.003	0.900	0.130	0.042	0.003	0.950
σ_e^2	0.062	0.061	0.089	0.940	0.032	0.128	0.088	0.850	0.125	0.124	0.084	0.860

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.10: PARAMETER ESTIMATION FOR THE THREE GCMS WITH NONNORMAL AND IGNORABLE MISSING DATA WHEN $N = 600$, $\sigma_e^2 = 0.7$ AND $mr = 0.18$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.002	0.047	0.002	0.960	0.008	0.041	0.003	0.880	0.008	0.041	0.003	0.970
β_S	0.121	0.024	0.002	0.960	0.030	0.017	0.000	0.930	0.029	0.017	0.000	0.940
σ_L^2	0.185	0.102	0.080	0.980	0.002	0.072	0.005	0.920	0.007	0.072	0.004	0.950
σ_S^2	0.201	0.016	0.001	0.980	0.117	0.010	0.000	0.920	0.112	0.010	0.000	0.960
σ_{LS}	0.206	0.031	0.007	0.950	0.073	0.022	0.001	0.910	0.079	0.022	0.001	0.920
σ_e^2	0.064	0.031	0.035	0.900	0.315	0.087	0.093	0.850	0.485	0.095	0.162	0.780

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.11: PARAMETER ESTIMATION FOR GCMS WITH NONNORMAL AND NON-IGNORABLE MISSING DATA WHEN $N = 200$, $r = 0.8$, $mr = 0.36$ AND $\sigma_e^2 = 0.7$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.001	0.068	0.003	0.980	0.008	0.061	0.005	0.940	0.007	0.062	0.004	0.940
β_S	0.140	0.050	0.004	0.960	0.029	0.035	0.001	0.940	0.001	0.035	0.001	0.880
σ_L^2	0.220	0.188	0.399	0.890	0.013	0.123	0.009	0.850	0.004	0.125	0.008	0.940
σ_S^2	1.296	0.052	0.029	0.920	0.384	0.023	0.002	0.940	0.390	0.023	0.002	0.930
σ_{LS}	0.422	0.073	0.064	0.930	0.119	0.042	0.002	0.890	0.103	0.043	0.002	0.880
σ_e^2	0.590	0.074	0.199	0.860	0.519	0.131	0.152	0.840	0.858	0.152	0.285	0.800

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

Table 6.12: PARAMETER ESTIMATION FOR THREE GCMS WITH OUTLIER AND COMPLETE DATA WHEN $N = 200$, AND $\sigma_e^2 = 0.5$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.043	0.099	0.066	1.000	0.045	0.098	0.071	0.910	0.045	0.098	0.072	0.940
β_S	0.086	0.035	0.002	1.000	0.065	0.026	0.002	0.940	0.059	0.026	0.002	0.950
σ_L^2	18.608	2.074	346.990	1.000	17.003	1.957	290.226	0.860	16.910	1.946	286.947	0.910
σ_S^2	0.175	0.020	0.001	1.000	0.284	0.019	0.001	0.820	0.247	0.018	0.001	0.880
σ_{LS}	0.185	0.168	0.042	1.000	1.671	0.189	0.321	0.840	1.636	0.187	0.301	0.880
σ_e^2	0.186	0.028	0.010	1.000	0.099	0.033	0.005	0.820	0.107	0.037	0.006	0.880

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

6.0.4.1.3 Outlier Data We then investigate the performance of the three GCMs for the outlier data. In general, when data contain outliers, the nonparametric Bayesian GCMs have good performance in estimating the average latent slope parameter β_S and the measurement error σ_e^2.

Table 6.12 presents the absolute relative bias, posterior SDs, MSEs and convergence rates of the parameter estimates for complete data with outliers when $N = 200$ and $\sigma_e^2 = 0.5$. The two nonparametric Bayesian GCMs outperform the normal-distribution-based GCM with smaller absolute relative bias, posterior SDs and MSEs of β_S and σ_e^2. The convergence rates of the two nonparametric Bayesian GCMs are lower than the normal-distribution-based GCM.

When data are ignorably missing and contain outliers, results patterns are similar to the above-discussed complete and outlier condition in that the two nonparametric Bayesian GCMs outperform the normal-distribution-based GCM in estimating β_S and σ_e^2. For example, the absolute relative bias, posterior SDs, MSEs and convergence rates of the parameter estimates for the ignorable missing and outlier data when $N = 200$, $\sigma_e^2 = 0.7$ and $mr = 0.36$ are presented in Table 6.0.4.1.3. Both BNP GCM and BNP selection GCM provide smaller absolute relative bias, posterior SDs and MSEs of β_S and σ_e^2 than the normal-distribution-based GCM. The BNP GCM and BNP selection GCM have lower convergence rates. Comparing between the BNP GCM and BNP selection GCM, the performance of the two nonparametric Bayesian GCMs are similar under the ignorable missing data. This is expected as the data are missing at random and are handled by the MCMC technique in the nonparametric Bayesian approach.

Table 6.13: PARAMETER ESTIMATION FOR THREE GCMS WITH OUTLIER AND IGNORABLE MISSING DATA WHEN $N = 200$, $mr = 0.36$ AND $\sigma_e^2 = 0.7$

	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.042	0.098	0.065	0.990	0.043	0.098	0.067	0.980	0.043	0.098	0.067	0.940
β_S	0.157	0.039	0.004	0.970	0.055	0.029	0.003	0.870	0.072	0.029	0.003	0.840
σ_L^2	19.265	2.116	371.931	0.860	18.191	2.058	332.414	0.840	18.184	2.048	331.628	0.830
σ_S^2	1.037	0.034	0.016	0.960	0.780	0.036	0.013	0.860	0.648	0.032	0.008	0.860
σ_{LS}	0.302	0.242	0.154	0.990	2.619	0.291	0.967	0.830	2.368	0.279	0.748	0.830
σ_e^2	0.843	0.014	0.350	0.820	0.599	0.081	0.206	0.790	0.432	0.082	0.108	0.810

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

When data contain outliers and are non-ignorably missing, the BNP selection GCM provides the smallest absolute relative bias and MSEs of the estimated latent growth parameter β_S and measurement error σ_e^2. Table 6.0.4.1.3 presents the converged absolute relative bias, posterior SDs, MSEs and convergence rates of the six GCM parameter estimates for the outlier and non-ignorable missing data with $N = 600$, $r = 0.8$, $mr = 0.36$ and $\sigma_e^2 = 0.7$. There is a decreasing absolute relative bias of β_S as 0.20, 0.095 to 0.083 from the normal-distribution-based GCM, BNP GCM and BNP selection GCM, respectively. Moreover, the absolute relative bias of the estimated σ_e^2 drop substantially from 0.44, 0.36 to 0.19 from the normal-distribution-based GCM, BNP GCM to the BNP selection GCM, respectively. Furthermore, the MSEs of the estimated β_S and σ_e^2 are smaller from the BNP selection GCM than those from the other two GCMs. In sum, the BNP selection GCM provides less biased and more efficient estimates of the growth parameter β_S and measurement error σ_e^2 (highlighted in blue) than the normal-distribution-based GCM and the BNP GCM do. The same pattern applies to all other conditions with outlier and non-ignorable missing data. Note that in the presence of outlier data, the variance of σ_e^2 is hard to be controlled and there is no true value for this parameter. Therefore, the absolute relative bias, posterior SDs and MSEs of the estimates of σ_e^2 may not be trusted. However, the estimates of σ_e^2 are still trustworthy.

Note that for the outlier data, estimates of σ_L^2 are substantially different

Table 6.14: PARAMETER ESTIMATION FOR THREE GCMS WITH OUTLIER AND NON-IGNORABLE MISSING DATA WHEN $N = 600$, $r = 0.8$, $mr = 0.36$ AND $\sigma_e^2 = 0.7$

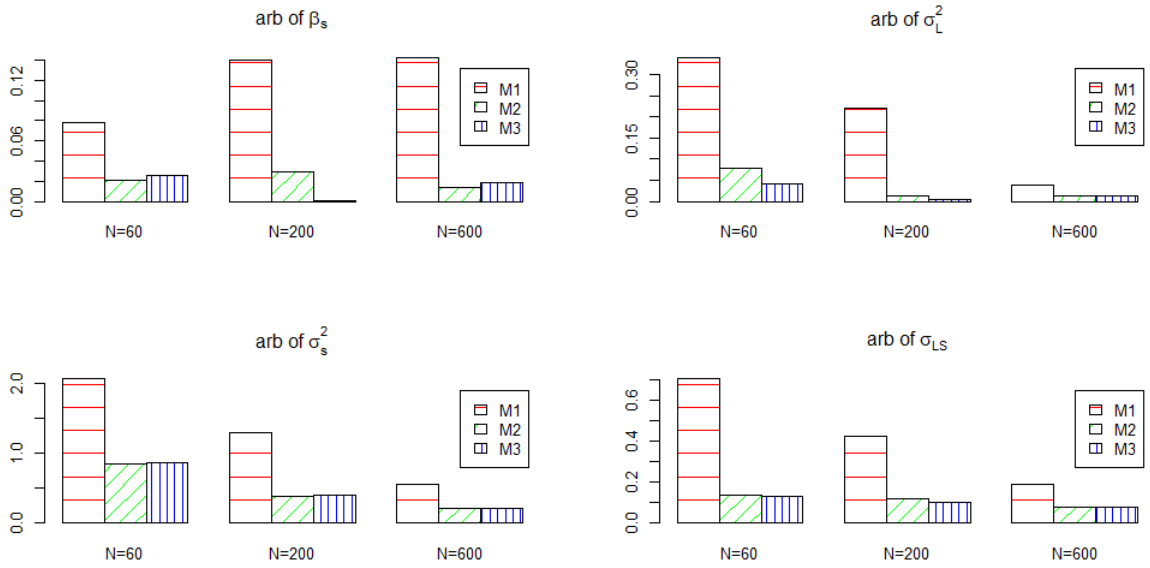
	Normal-distribution-based GCM				BNP GCM				BNP Selection GCM			
	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]	ARB[1]	SD[2]	MSE[3]	CR[4]
β_L	0.117	0.092	0.493	0.980	0.119	0.091	0.507	0.930	0.119	0.091	0.513	0.950
β_S	0.204	0.026	0.004	0.960	0.095	0.018	0.002	0.890	0.083	0.018	0.001	0.880
σ_L^2	16.697	1.114	279.049	0.960	15.529	1.065	241.567	0.850	15.397	1.060	237.507	0.790
σ_S^2	0.117	0.017	0.001	0.910	0.227	0.013	0.001	0.860	0.232	0.013	0.001	0.870
σ_{LS}	0.128	0.118	0.026	0.920	1.727	0.129	0.301	0.830	1.786	0.131	0.314	0.830
σ_e^2	0.435	0.023	0.095	0.940	0.357	0.037	0.065	0.840	0.193	0.041	0.021	0.850

Note 1. absolute relative bias; 2. posterior standard deviation; 3. mean squared error; 4. convergence rate

from the population value, which have also affected the estimates of σ_{LS} and σ_S^2, all of which are related to random effects. Although the nonparametric Bayesian approach is used in general in the two nonparametric Bayesian GCMs, only the level-1 measurement error is modeled with the DP mixture prior while the level-2 random residuals, which are related to random effects, are still modeled with traditional parametric distributions. This could make the likelihood harder to converge to a global maximum. The estimates of the three random effects parameters (σ_L^2, σ_S^2 and σ_{LS}) need further investigation under this situation.

6.0.4.1.4 Other simulation factors Previously, we compared the performance of the three GCMs under different nonnormal and missing data conditions. This subsection takes a close look at the effects of other simulation factors, particularly on the performance of the BNP selection GCM, which is the focus of the study. The performance of the BNP selection GCM is affected by sample size, missing rate and strength between missing data and the auxiliary variable. The effect of the BNP selection GCM increases with a larger proportion of the non-ignorable missingness. As sample size increases, the advantage of the BNP selection GCM is more obvious. Figures 6.1 and 6.2 demonstrate the absolute relative bias and MSEs, respectively, of GCM parameters for the nonnormal and non-ignorable data across all sample sizes when $mr = 0.36$, $r = 0.8$ and $\sigma_e^2 = 0.5$. Based on previous findings, we take a close look at results from four parameters (β_S, σ_L^2, σ_S^2 and σ_{LS}). The performance of the BNP selection GCM progresses with increas-

Figure 6.1: ABSOLUTE RELATIVE BIAS OF GCM PARAMETERS FOR NONNORMAL AND NON-IGNORABLE MISSING DATA WHEN $mr = 0.36$, $r = 0.8$ AND $\sigma_e^2 = 0.5$



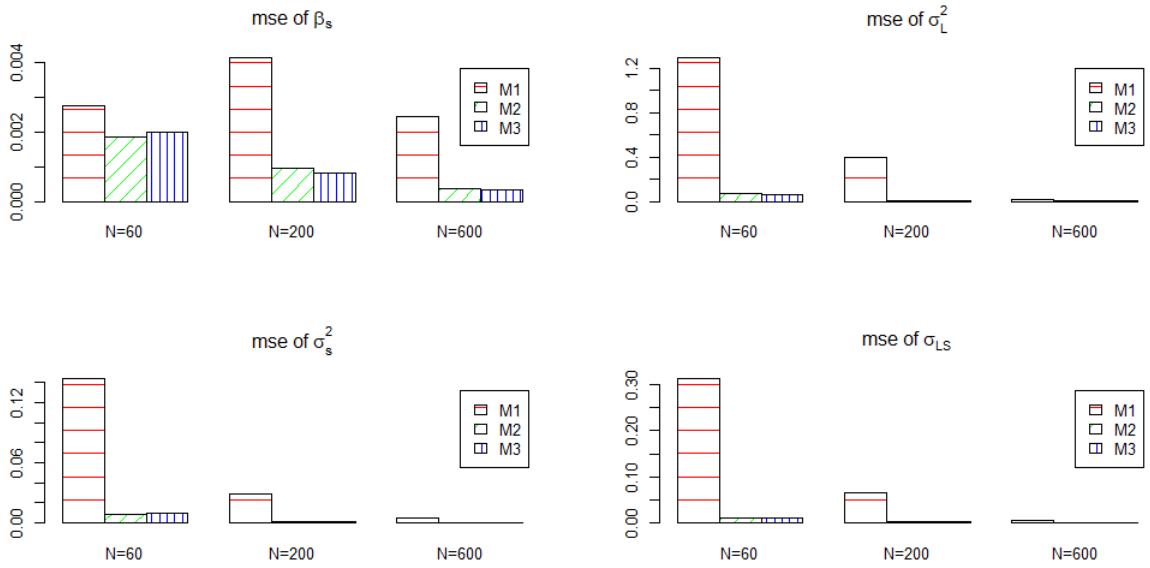
Note: 1. M1=Normal-distribution-based GCM; M2=BNP GCM; M3=BNP Selection GCM
2. Y-axis scales are different in each cell.

ing sample sizes in estimating the four parameters. The absolute relative bias and MSEs from the BNP selection GCM are the lowest among the three models in almost all conditions. In addition, the effect of the BNP selection GCM is more apparent with a stronger association between the missing data and the auxiliary variable that explains the missingness. Tables 6.3 and 6.4 present the absolute relative bias and MSEs of six parameter estimates with the nonnormal and non-ignorable missing data when $N = 60$, $mr = 0.36$ and $\sigma_e^2 = 0.7$. We see that overall the BNP selection GCM has a better performance in parameter estimation when the strength between the missing data and the auxiliary variable is stronger.

6.0.4.2 Model Selection Criteria

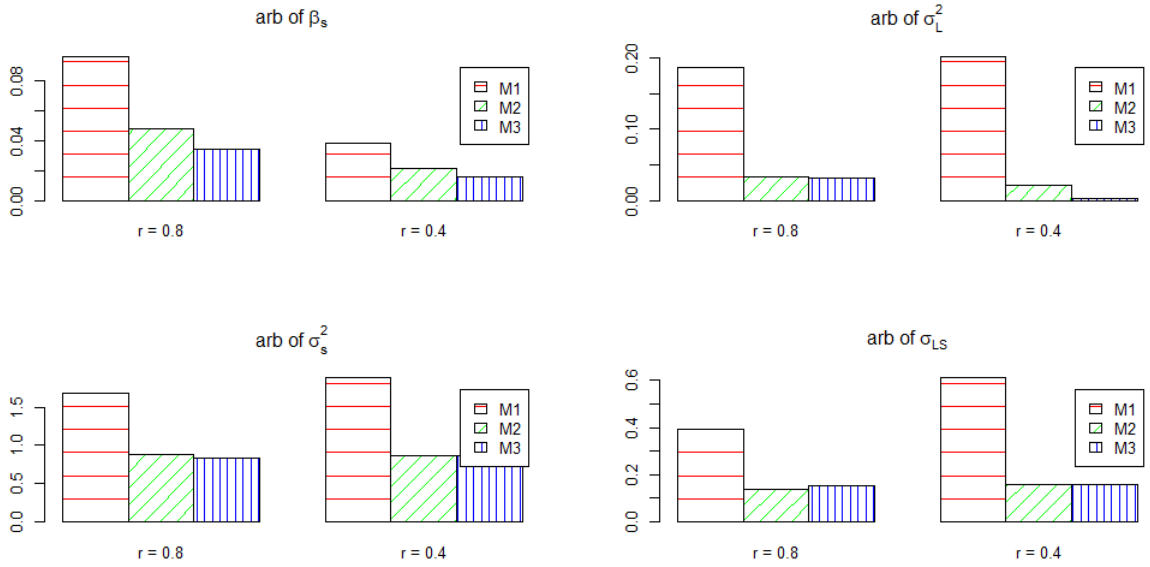
This part compares and investigates the performance of the thirteen Bayesian model selection criteria in the three Bayesian GCMs. The nonparametric

Figure 6.2: MEAN SQUARED ERROR OF GCM PARAMETERS FOR NONNORMAL AND NON-IGNORABLE MISSING DATA WHEN $mr = 0.36$, $r = 0.8$ AND $\sigma_e^2 = 0.5$



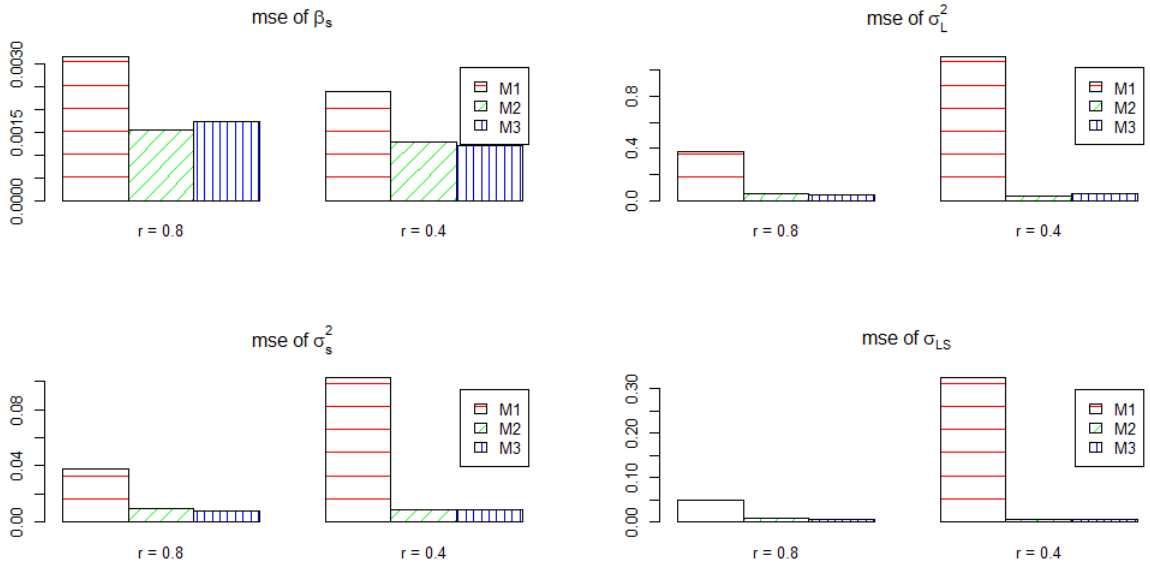
Note: 1. M1=Normal-distribution-based GCM; M2=BNP GCM; M3=BNP Selection GCM
2. Y-axis scales are different in each cell.

Figure 6.3: ABSOLUTE RELATIVE BIAS OF GCM PARAMETERS FOR NONNORMAL AND NON-IGNORABLE MISSING DATA WHEN $N = 60$, $mr = 0.36$ AND $\sigma_e^2 = 0.7$



Note: Y-axis scales are different in each cell.

Figure 6.4: MEAN SQUARED ERROR OF GCM PARAMETERS FOR NONNORMAL AND NON-IGNORABLE MISSING DATA WHEN $N = 60$, $mr = 0.36$ AND $\sigma_e^2 = 0.7$



Note: Y-axis scales are different in each cell.

Bayesian growth curve modeling is a type of the infinite mixture model, so the Bayesian criteria are evaluated to select Bayesian models in the infinite mixture modeling framework. The proportion of the correct model being selected from the converged replications is reported in each table. The largest proportion across three Bayesian GCMs is highlighted in blue for each criterion. Model selection results for all simulation conditions can be found in Appendix C, available on this webpage
https://drive.google.com/drive/folders/1FvNKGUt54k6VzCgY47tC9SBHSjre6p13

In summary, all model selection criteria, except the conventional DIC, are able to correctly select the nonparametric Bayesian GCM under the nonnormal data with an average sensitivity above 95%. DIC can always correctly select the normal-distribution-based GCM when data are normal. When data are nonnormal and have missing values, the performance of DIC in terms of model selection depends on sample size and missing rate.

We first investigate the performance of the model fit criteria under normal data. For each criterion, the proportion of selecting among three models across converged replications is reported. Results from the normally-distributed population with the complete data (Table 6.0.4.2), the ignorable missingness (Table 6.0.4.2) and the non-ignorable missingness (Table 6.17) are presented. For the normal data, the normal-distribution-based GCM should be the true model. Based on the parameter estimation results discussed above, the BNP GCM also well describes the normal data. For the non-ignorable missing data, the BNP selection GCM should be the best model.

When the population data are normally distributed, regardless of whether data have missing values or which mechanism the missingness is, the conventional DIC always selects the normal-distribution-based GCM. Among the other twelve indices, the Dmedian.BIC, Dmedian.CAIC, Dbar.BIC and Dbar.CAIC select the normal-distribution-based GCM when sample size is small (i.e., $N = 60$) and select the BNP GCM as sample sizes increase to 200 and 600. The Dmedian.AIC, Dmedian.ABIC, Dbar.AIC, Dbar.ABIC as well as all four Dhat based criteria consistently select the BNP GCM or the infinite mixture model regardless of sample size. When data are normal and non-ignorably missing, the thirteen indices display similar patterns to previously discussed normal data conditions. The Dmedian.ABIC and Dbar.ABIC differ slightly from previous conditions in that they select the normal-distribution-based GCM even when $N = 200$. Nevertheless, the two indices are able to select the BNP GCM when sample size reaches to 600.

Table 6.15: MODEL SELECTION FOR THE THREE GCMS WITH NORMAL AND COMPLETE DATA WHEN $N = 60$, 200, 600 AND $\sigma_e^2 = 0.7$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	35%	65%	0%	4%	96%	0%	0%	100%	0%
Dmedian.BIC	79%	21%	0%	39%	61%	0%	13%	87%	0%
Dmedian.SABIC	4%	96%	0%	4%	96%	0%	0%	100%	0%
Dmedian.CAIC	89%	11%	0%	51%	49%	0%	21%	79%	0%
DIC	100%	0%	0%	100%	0%	0%	100%	0%	0%
Dbar.AIC	30%	70%	0%	2%	98%	0%	0%	100%	0%
Dbar.BIC	76%	24%	0%	37%	63%	0%	11%	89%	0%
Dbar.SABIC	3%	97%	0%	3%	97%	0%	0%	100%	0%
Dbar.CAIC	84%	16%	0%	50%	50%	0%	19%	81%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

Note that none of the criteria picked the BNP selection GCM. This is different from what we expected, especially because DIC usually select more complicated models. We later figured out that BNP selection GCM and the other two models are not comparable because they used different data. Specifically, for the BNP selection model, missing data indicators were created and included in the data. With different data (i.e., with or without missingness indicators) being used in these models, the BNP selection GCM cannot be compared directly to the other two models using the model fit criteria. In order to compare the selection-structured and non-selection-structured models in the future, we can include missing data indicators for the other two models but fix certain parameters to zero.

We then evaluate the performance of the model selection criteria when data are nonnormal. Under nonnormality, the nonparametric Bayesian GCMs are better models for the analyses. Table 6.0.4.2 presents the model selection proportions for nonnormal and complete data with different sample sizes and $\sigma_e^2 = 0.5$. DIC typically favors more complicated models. However, all indices, except the DIC, select the BNP GCM. The percentage of these indices correctly choosing the nonparametric Bayesian GCM is nearly 100%

Table 6.16: MODEL SELECTION FOR THE THREE GCMS WITH NORMAL AND IGNORABLE MISSING DATA WHEN $N = 60, 200, 600$, $mr = 0.18$ AND $\sigma_e^2 = 0.7$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	38%	62%	0%	7%	93%	0%	0%	100%	0%
Dmedian.BIC	74%	26%	0%	49%	51%	0%	18%	82%	0%
Dmedian.SABIC	6%	94%	0%	7%	93%	0%	0%	100%	0%
Dmedian.CAIC	84%	16%	0%	59%	41%	0%	23%	77%	0%
DIC	100%	0%	0%	100%	0%	0%	100%	0%	0%
Dbar.AIC	34%	66%	0%	4%	96%	0%	0%	100%	0%
Dbar.BIC	72%	28%	0%	49%	51%	0%	13%	87%	0%
Dbar.SABIC	5%	95%	0%	6%	94%	0%	0%	100%	0%
Dbar.CAIC	82%	18%	0%	57%	43%	0%	21%	79%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

Table 6.17: MODEL SELECTION FOR THE THREE GCMS WITH NORMAL AND NON-IGNORABLE MISSING DATA WHEN $N = 60$, 200, 600, $mr = 0.36$, $r = 0.4$ AND $\sigma_e^2 = 0.5$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	38%	62%	0%	1%	99%	0%	0%	100%	0%
Dmedian.BIC	74%	26%	0%	45%	55%	0%	21%	79%	0%
Dmedian.SABIC	12%	88%	0%	1%	99%	0%	0%	100%	0%
Dmedian.CAIC	87%	13%	0%	59%	41%	0%	36%	64%	0%
DIC	100%	0%	0%	100%	0%	0%	100%	0%	0%
Dbar.AIC	35%	65%	0%	1%	99%	0%	0%	100%	0%
Dbar.BIC	71%	29%	0%	42%	58%	0%	18%	82%	0%
Dbar.SABIC	11%	89%	0%	1%	99%	0%	0%	100%	0%
Dbar.CAIC	84%	16%	0%	54%	46%	0%	29%	71%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

in almost all conditions. DIC selects the normal-distribution-based GCM when $N = 60$. DIC starts to change its selection decision when $N = 200$. When sample size increases to 600, DIC has 100% rate of correctly selecting the nonparametric Bayesian GCM. Table 6.0.4.2 lists the model selection proportions when data contain outliers and ignorable missingness with $mr = 0.36$ and $\sigma_e^2 = 0.5$. In general, the patterns of all thirteen indices are similar to the nonnormal and complete conditions. With a larger missing rate and smaller sample size, although the median-based deviance indices (Dmedian.AIC, Dmedian.BIC, Dmedian.ABIC and Dmedian.CAIC) still correctly select the nonparametric Bayesian GCM, in small sampled conditions ($N = 60$), their percentages of correctly selecting those models are smaller, ranging from 75% to 83%.

Table 6.0.4.2 shows the model selection proportions when the population data are nonnormal and non-ignorably missing when $mr = 0.18$, $r = 0.8$ and $\sigma_e^2 = 0.7$. On one hand, all twelve indices, except DIC, uniformly choose the BNP GCM. Although the BNP selection GCM should be more appropriate to analyze the non-ignorable missing data, all indices choose the non-selection-structured nonparametric Bayesian GCM over the added-on selection model. This shows that these DIC-related indices are able to correctly select the BNP GCM under nonnormal data. Note that results for the indices are not comparable between the selection-structured and non-selection-structured models because different data sets were used in these models. On the other hand, the performance of DIC in correctly choosing the nonparametric Bayesian GCM for the nonnormal data is affected by missing rate and sample size. With a higher missing rate, the proportion of DIC to correctly select the nonparametric Bayesian GCM with nonnormal data is smaller. Table 6.0.4.2 reports the same conditions as Table 6.0.4.2 and differs only in missing rates. When $N = 200$ and $\sigma_e^2 = 0.7$, the proportion of DIC that correctly selects the BNP GCM for the nonnormal data is 90% when the missing rate $mr = 0.18$ as shown in Table 6.0.4.2. From Table 6.0.4.2, this proportion falls to 36% when the missing rate rises to $mr = 0.36$. Furthermore, the performance of DIC to correctly select the nonparametric Bayesian GCM or the infinite mixture model is affected by sample size. For example, when $N = 60$ and $mr = 0.18$, DIC correctly selects BNP GCM with a proportion of 54%. When sample size increases to 600, the correct model selection proportion increases to 100%. Additionally, the impact of sample size on DIC to correctly select the infinite mixture model is more noticeable when the missing rate is higher. For example, from Table 6.0.4.2, when $mr = 0.36$, the proportions

Table 6.18: MODEL SELECTION FOR THE THREE GCMS WITH OUTLIER AND COMPLETE DATA WHEN $N = 60$, 200, 600 AND $\sigma_e^2 = 0.5$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
DIC	75%	25%	0%	5%	95%	0%	0%	100%	0%
Dbar.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	99%	1%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	99%	1%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	99%	1%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	99%	1%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

of DIC that correctly select the BNP GCM are 12% for $N = 60$, 36% for $N = 200$ and 57% for $N = 600$, respectively. Comparing among the model selection criteria defined based on Dmedian, Dbar and Dhat, the Dhat-based criteria perform slightly better as they consistently have larger proportions of correctly selecting the nonparametric Bayesian GCM for nonnormal data, particularly when sample size is small and missing rate is large.

Table 6.19: MODEL SELECTION FOR THE THREE GCMS WITH OUTLIER AND IGNORABLE DATA WHEN $N = 60, 200, 600$, $mr = 0.36$ AND $\sigma_e^2 = 0.5$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	19%	81%	0%	0%	100%	0%	0%	100%	0%
Dmedian.BIC	24%	76%	0%	0%	100%	0%	0%	100%	0%
Dmedian.SABIC	17%	83%	0%	0%	100%	0%	0%	100%	0%
Dmedian.CAIC	25%	75%	0%	0%	100%	0%	0%	100%	0%
DIC	100%	0%	0%	100%	0%	0%	100%	0%	0%
Dbar.AIC	19%	81%	0%	0%	100%	0%	0%	100%	0%
Dbar.BIC	24%	76%	0%	0%	100%	0%	0%	100%	0%
Dbar.SABIC	16%	84%	0%	0%	100%	0%	0%	100%	0%
Dbar.CAIC	24%	76%	0%	0%	100%	0%	0%	100%	0%
Dhat.AIC	0%	99%	1%	0%	100%	0%	0%	99%	1%
Dhat.BIC	0%	99%	1%	0%	100%	0%	0%	99%	1%
Dhat.SABIC	0%	99%	1%	0%	100%	0%	0%	99%	1%
Dhat.CAIC	0%	99%	1%	0%	100%	0%	0%	99%	1%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

Table 6.20: MODEL SELECTION FOR THE THREE GCMS WITH NON-NORMAL AND NON-IGNORABLE DATA WHEN $N = 60, 200, 600$, $mr = 0.18$, $r = 0.8$ AND $\sigma_e^2 = 0.7$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
DIC	46%	54%	0%	10%	90%	0%	0%	100%	0%
Dbar.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%

Note 1. Normal-distribution-based GCM; 2. BNP GCM; 3. BNP selection GCM

Table 6.21: MODEL SELECTION FOR THE THREE GCMS WITH NON-NORMAL AND NON-IGNORABLE DATA WHEN $N = 60, 200, 600$, $mr = 0.36$, $r = 0.8$ AND $\sigma_e^2 = 0.7$

	N=60			N=200			N=600		
	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]	M1[1]	M2[2]	M3[3]
Dmedian.AIC	2%	98%	0%	0%	100%	0%	0%	100%	0%
Dmedian.BIC	4%	96%	0%	0%	100%	0%	0%	100%	0%
Dmedian.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dmedian.CAIC	4%	96%	0%	0%	100%	0%	0%	100%	0%
DIC	88%	12%	0%	64%	36%	0%	43%	57%	0%
Dbar.AIC	2%	98%	0%	0%	100%	0%	0%	100%	0%
Dbar.BIC	3%	97%	0%	0%	100%	0%	0%	100%	0%
Dbar.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dbar.CAIC	4%	96%	0%	0%	100%	0%	0%	100%	0%
Dhat.AIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.BIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.SABIC	0%	100%	0%	0%	100%	0%	0%	100%	0%
Dhat.CAIC	0%	100%	0%	0%	100%	0%	0%	100%	0%

Chapter 7

Real Data Analysis

This chapter illustrates the application of the three GCMs using a real data set about students' longitudinal mathematics achievement scores. The data set contains Peabody Individual Achievement Test (PIAT) math scores from 399 students from the National Longitudinal Survey of Youth 1997 (NLSY97) Cohort. In this study, students were measured yearly from grades 7 to 10. The individual growth trajectory plot (Figure 7.1) suggests a linear growth pattern of mathematics abilities, with highlighted red line being the average growth pattern for all participants. The plot also shows that participants do not follow the same pattern, which means in addition to growth patterns, there are individual differences in growth. Data will be analyzed using linear growth curve modeling to investigate intraindividual change and interindividual differences in change over time.

Observed data from each grade are plotted. The histograms (Figure 7.2) and QQ-plots (Figure 7.3) show that observed scores from each grade deviate from normality. Descriptive statistics of the observed data (Table 7.1) show that skewness and kurtosis of the observed data in Grades 7-10 are different from normal distribution. The Shapiro-Wilk normality tests were conducted (Table 7.2) to provide additional evidence that observed data are nonnormal in each grade. Summary statistics (Table 7.1) shows that each grade has missing scores. Because of the existence of missing values, the distributions of the data can be affected by where the missing data are located. For example, the negatively skewed score distribution in Grade 10 of Figure 7.2 can be associated with nonnormal data or can be due to missing values in the lower range of the math achievement scores. Thus, the true data distribution is unknown. But since the nonnormality is suspected, we consider using

Figure 7.1: Individual Growth Trajectory Plot for the PIAT Data

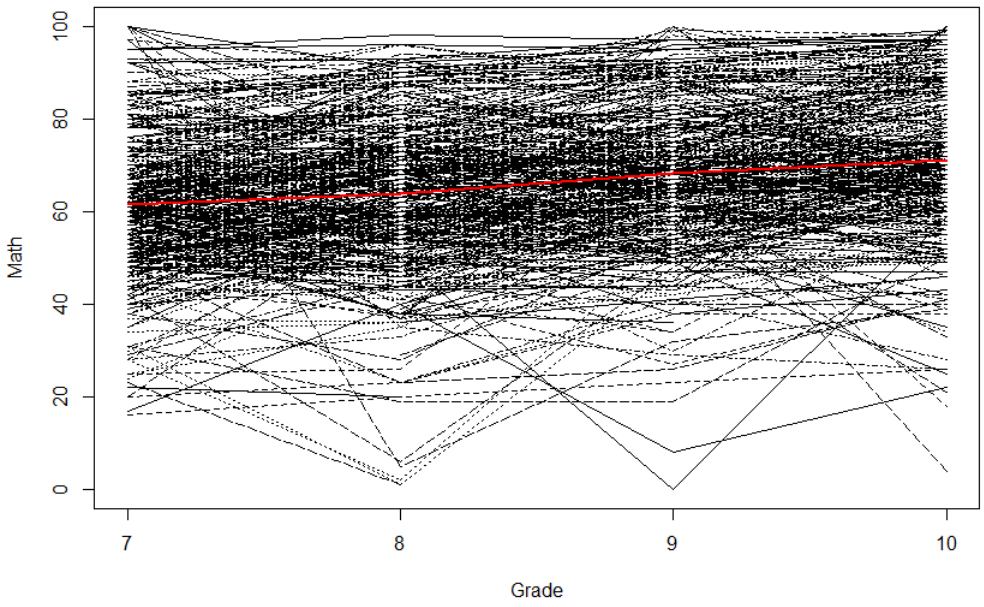
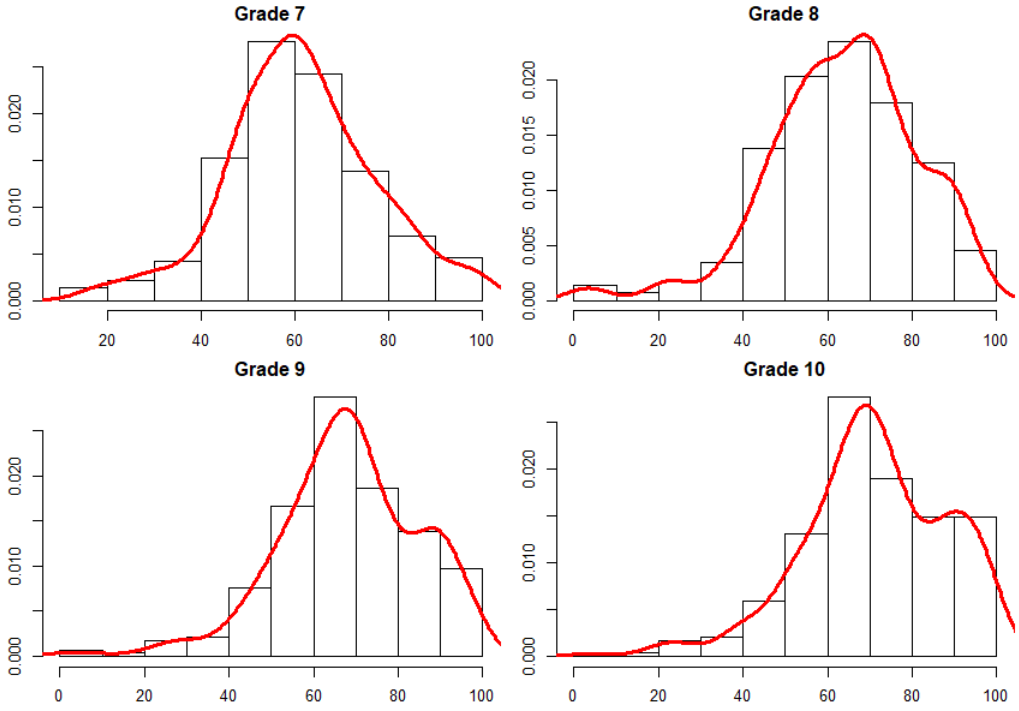


Figure 7.2: HISTOGRAM OF THE LONGITUDINAL PIAT DATA BY GRADE



robust methods to analyze the data.

Because of the untenability of the missingness mechanisms (Davey et al., 2009), the missingness mechanism in the data set can not be determined. Given undetermined missingness mechanism and possible nonnormality of the data, we apply the nonparametric Bayesian GCMs (i.e., BNP GCM and BNP selection GCM) to the empirical longitudinal PIAT data and compare to the results from the traditional normal-distribution-based GCM. All models apply the same priors (refer to Table 6.2) as those from the simulation study. From the simulation results, the nonparametric Bayesian GCMs should be better models to fit the nonnormal and missing data.

Table 7.3 presents six major parameter estimates from the three GCMs. Estimates from the two nonparametric Bayesian GCMs differ substantially from the normal-distribution-based GCM, particularly in the estimates of σ_L^2, σ_S^2 and σ_{LS}. Results between the BNP GCM and BNP selection GCM are similar and we may infer that the data are ignorably missing. Given the sim-

Figure 7.3: QQ-PLOT OF THE LONGITUDINAL PIAT DATA BY GRADE

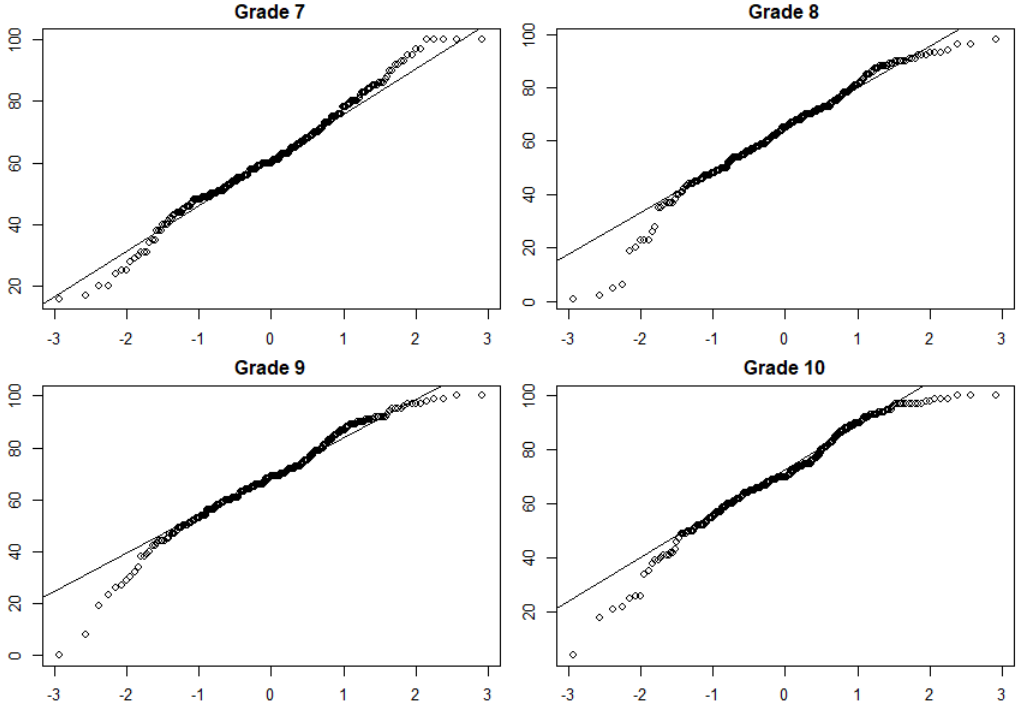


Table 7.1: DESCRIPTIVE STATISTICS OF THE PIAT DATA WITH MISSING VALUES

Grade	Mean	S.D.	Skewness	Kurtosis	# of Missing Values
7	61.50	15.85	0.01	3.33	24
8	63.80	17.27	-0.64	4.05	22
9	68.21	16.65	-0.55	3.97	42
10	71.00	17.02	-0.58	3.68	49

Table 7.2: SHAPIRO-WILK NORMALITY TEST FOR THE PIAT DATA BY GRADE

Grade	Statistic	p-value
7	0.988	0.016
8	0.969	0.000
9	0.973	0.000
10	0.967	0.000

Table 7.3: PARAMETER ESTIMATES OF THE PIAT MATH STUDY

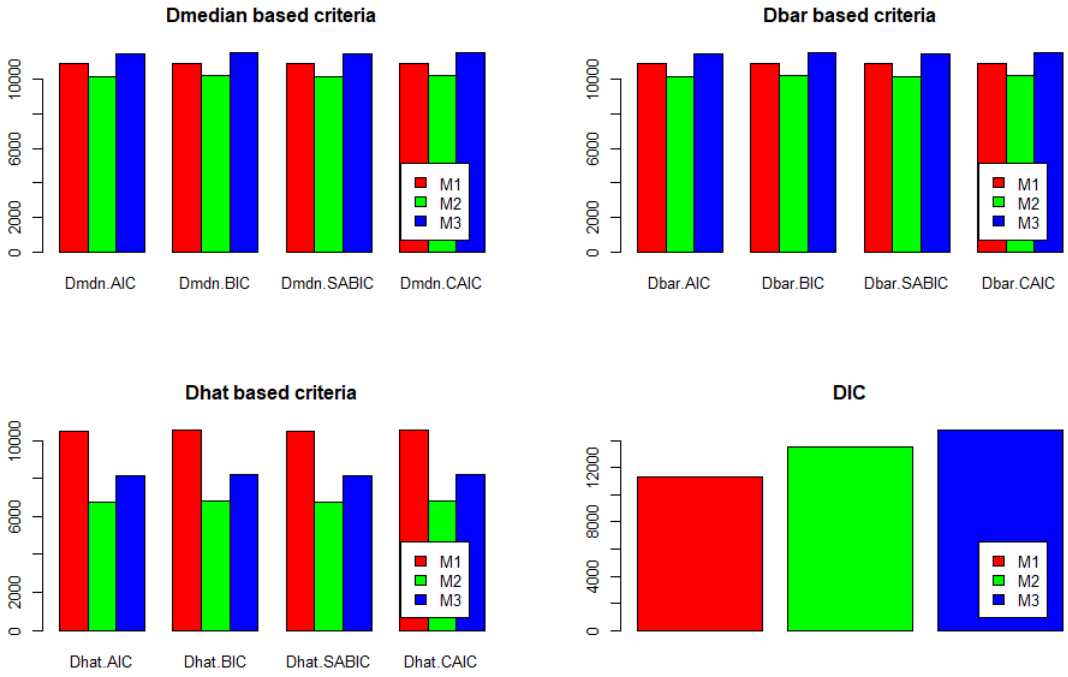
	M1				M2				M3			
	Mean	SD	CI.L	CI.U	Mean	SD	CI.L	CI.U	Mean	SD	CI.L	CI.U
β_L	60.99	0.10	60.80	61.19	61.00	0.10	60.81	61.20	61.00	0.10	60.81	61.20
β_S	3.01	0.09	2.82	3.19	3.02	0.09	2.85	3.20	3.02	0.09	2.85	3.19
σ_L^2	166.70	18.39	133.10	205.20	177.20	16.07	147.70	210.80	177.30	16.02	148.10	210.50
σ_S^2	4.02	2.74	0.38	9.93	1.90	1.14	0.27	4.48	1.92	1.15	0.23	4.46
σ_{LS}	0.56	5.31	-10.75	9.68	-1.14	2.78	-6.78	4.08	-1.19	2.76	-6.90	3.88
σ_e^2	102.70	6.29	90.91	115.10	105.70	4.64	97.00	115.30	105.70	4.65	97.16	115.40

Note: M1=Normal-distribution-based GCM; M2=BNP GCM; M3=BNP Selection GCM

ulation study results, we rely on the BNP GCM for the interpretation. The average true math achievement score in Grade 7 is about 61, with an average true annual growth rate of about 3 from Grades 7 to 10. The measurement error indicates the population variability in an average student's math score around its own true change trajectory, which is 106, and is statistically significant. The population variation in true individual intercepts around the averages is 171 and the between-person variability in the true rate of change is about 1.9, both of which are statistically significant. There is no significant magnitude or direction of association between where the student starts and its rate of change in math abilities. Comparing among the three GCMs, the nonparametric Bayesian GCMs produce substantially higher estimates of σ_L^2 and σ_e^2 and lower estimates of σ_S^2 than the normal-distribution-based GCM does. The posterior SDs from the nonparametric Bayesian GCMs are smaller than those from the normal-distribution-based GCM. These show that under nonnormal and incomplete longitudinal data, the effects of the nonparametric Bayesian GCMs are different than those from the traditional normal-distribution-based GCM.

Furthermore, thirteen model selection criteria are calculated and compared for each of the three models. Results are shown in Table 7 and Figure 7. All the model selection criteria, except DIC, choose the BNP GCM under the nonnormal and missing data condition. The DIC index, on the contrary, chooses the normal-distribution-based GCM. Because based on the simulation study, DIC has a low sensitivity, we trust the other twelve model selection criteria and select the BNP GCM as our final model.

Figure 7.4: MODEL SELECTION OF THE PIAT MATH STUDY



Note: M1=Normal-distribution-based GCM; M2=BNP GCM; M3=BNP Selection GCM

Table 7.4: MODEL SELECTION OF THE PIAT MATH STUDY

	M1	M2	M3
Dmedian.AIC	10912.00	10158.00	11470.00
Dmedian.BIC	10935.93	10193.90	11529.83
Dmedian.SABIC	10916.90	10165.34	11482.24
Dmedian.CAIC	10941.93	10202.90	11544.83
DIC	11290.00	13542.55	14762.41
Dbar.AIC	10907.01	10161.29	11465.24
Dbar.BIC	10930.95	10197.19	11525.08
Dbar.SABIC	10911.91	10168.63	11477.48
Dbar.CAIC	10936.95	10206.19	11540.08
Dhat.AIC	10508.51	6762.03	8138.07
Dhat.BIC	10532.45	6797.93	8197.90
Dhat.SABIC	10513.41	6769.38	8150.31
Dhat.CAIC	10538.45	6806.93	8212.90

Note: M1=Normal-distribution-based GCM; M2=BNP GCM; M3=BNP Selection GCM

Chapter 8

Discussion

Growth curve modeling is a commonly used technique to analyze longitudinal data. Applying Bayesian methods to growth curve analysis has gained popularity in recent years. While many Bayesian methods apply parametric assumptions to growth curve analysis, the nonparametric Bayesian approaches bring modeling flexibility to GCMs. This study highlights two novelties to current literature in nonparametric Bayesian growth curve analysis. First, the study proposes a nonparametric Bayesian GCM with an added-on selection structure to simultaneously analyze nonnormal and missing data. Traditional parametric distribution assumptions of measurement errors are replaced by an unknown distribution, which was constructed by Dirichlet process (DP). The nonparametric Bayesian growth curve analysis using DP flexibly models nonnormal residuals and thus analyzes longitudinal nonnormal data. In addition to handling the ignorable missing data with MCMC methods, a selection model structure is added to the nonparametric Bayesian growth curve analysis to accommodate the non-ignorable missing data as well. The study further develops model selection criteria in nonparametric Bayesian GCMs. Because the nonparametric Bayesian GCM is a type of infinite mixture model, the Bayesian criteria can also be used to evaluate Bayesian models in infinite mixture context. A Monte Carlo simulation was conducted to systematically evaluate the performance of the nonparametric Bayesian GCMs as well as to investigate the performance of the Bayesian model selection criteria. A real data example using NLSY79 data was used to illustrate the applications of the proposed methods and model selection criteria.

Major findings for parameter estimation are summarized as below. In

general, when data are normal and complete, the normal-distribution-based GCM, the BNP GCM and the BNP selection GCM perform almost equally well. The nonparametric Bayesian GCMs perform better than the normal-distribution-based GCM when data are nonnormal and ignorably missing, as the ignorable missingness is handled by MCMC method. Comparing among the three models for the non-ignorable missing data, the performance of BNP selection GCM depends specifically on types of nonnormal data. When data are nonnormal with specific skewness and kurtosis and are non-ignorably missing, the BNP selection GCM provides the most accurate and efficient estimates for the growth parameter β_S and parameters related to random effects (σ_L^2,σ_S^2 and σ_{LS}). Estimates of the latent intercept parameter β_L are similar among three models as this parameter is unrelated to the missingness. The BNP selection GCM tends to overestimate the σ_e^2, which may be related to how the multivariate nonnormal data are generated. When data are non-ignorably missing and contain outliers, the BNP selection GCM produces the smallest absolute relative bias and mean squared errors of the estimated growth parameter β_S and the measurement error σ_e^2. Further investigation regarding estimates of the three random effects related parameters σ_L^2, σ_S^2 and σ_{LS} is needed in this data condition. When data are normal and non-ignorably missing, the BNP selection GCM demonstrates similar performance to the other two GCMs in parameter estimation except to the estimates of σ_e^2. Moreover, factors including sample size, missing rate and strength between missing data and the auxiliary variable have impact on the performance of the parameter estimation. With a larger sample size, higher missing rate and stronger association between the missingness and the auxiliary variable that explains the missingness, the performance of the BNP selection GCM is more distinctive.

Note that the performance of the BNP selection GCM does not exceed other two GCMs in some cases under the non-ignorable missing data. For example, the outperformance of BNP selection GCM to other two GCMs is not as obvious when sample size is small. The performance pattern of the BNP selection GCM is clearer as sample size becomes larger. There may be several reasons for this. First, this may be due to the non-ignorable missingness generating mechanism. The non-ignorable missing data are generated based on a criterion from the auxiliary variable and then within that criterion, the missing data are generated with a probability. This could affect the actual presence of the non-ignorable missing data. If the non-ignorable missingness generating mechanism were more definite (i.e., having a larger probability

for generating the non-ignorable missing data or a more strict criterion), we expect the pattern that the BNP selection GCM largely outperforms the non-selection GCMs for the non-ignorable missing data to be clearer even in small samples. Second, the non-ignorable missingness is further complicated with the nonnormal data. We first generated nonnormal data and then removed values. This may affect the distribution of the data and its missing value mechanisms. In some situations, the adverse effect of the non-ignorable missing data may be lessened with nonnormality. The third reason is due to the limitation of the current study with only 100 replications. Having complicated models and data in this study, the patterns and effects of the BNP selection GCM should be clearer with more simulation replications.

In regard to the evaluation of the Bayesian model selection criteria, the following conclusions can be drawn. Among the thirteen model selection criteria being studied, twelve of the indices (the first four are the posterior median (Dmedian) based, the second four are the posterior mean (Dbar) based and the last four are based the deviance evaluated at posterior mean (Dhat)) can correctly select the nonparametric Bayesian GCM when data are nonnormal. The Dhat based criteria perform slightly better than the Dmedian as well as Dmean based criteria, particularly with small samples and large missing rates. The conventional DIC can always correctly select the normal-distribution-based GCM when data are normal. Under nonnormal data, the performance of DIC is affected by the sample size and the missing rate. With a larger sample and lower missing rate, DIC can correctly select the nonparametric Bayesian GCM with a larger certainty.

DIC is widely used to compare models in Bayesian methods. However, the performance of DIC in mixture models is less known and studies related to model selection in infinite mixture models are even less. This study demonstrates the relatively poor performance of DIC to correctly select the nonparametric Bayesian model, or the infinite mixture model under nonnormal data. Spiegelhalter et al. (2002) introduced DIC in linear and generalized linear models. Later literature shows that the calculation of DIC can take on different forms, such as in random effects models or mixture models (Celeux et al., 2006), which are akin to the growth curve analyses in this study. This study develops a number of ways to compute Bayesian criteria in infinite mixture models. Based on results from the simulation study, these DIC-variant criteria are promising tools to select Bayesian models. They can in particular help inform the selection of Bayesian growth models in infinite mixture context. Future studies may also test the new Bayesian criteria in traditional

finite mixture modeling. In addition to the IC-family criteria discussed in the study, other effective Bayesian model evaluation criteria, such as the Watanabe Akaike information criterion (WAIC, Watanabe, 2010) or leave-one-out cross validation (LOO, Gelfand, 1996) can be studied to evaluate and compare with the IC-family criteria.

Although the Bayesian model selection criteria are able to select nonparametric Bayesian GCM or the infinite mixture model in general, the BNP selection GCM as another infinite mixture model is never selected by these criteria. This is because the nonparametric Bayesian GCMs with the added-on selection structure and without the selection structure can not be directly compared due to the difference in data. Several strategies can be extended in future work. With missing variables being treated as parameters, a different computation of likelihood function may be needed. For example, Celeux et al. (2006) developed conditional DICs, which are based on computing the conditional likelihood that is conditioned on additional missing data parameters. Future work can consider using additional data information (e.g., missingness indicators) in the non-selection models to make the selection and non-selection structures comparable. Additionally, one can compare the BNP selection GCM with the selection model structure added to the normal-distribution-based GCM in future work. Comparing the BNP selection GCM with the normal-distribution-based selection GCM have at least two advantages. First, in terms of parameter estimation, for the normal and non-ignorable missing data, the simulation study finds that the BNP selection GCM has a similar performance to the other two GCMs studied. Alternatively, the normal-distribution-based GCM with an added-on selection structure may be a better model since it can correctly model both the normal part and the non-ignorable missingness part. Second, as the IC family criteria penalize complicated models, it is reasonable to compare models with the same complexity and further evaluate the performance of the Bayesian criteria discussed here.

As mentioned above, this study finds that the outperformance of the BNP selection GCM for the non-ignorable missing data is more apparent with a larger sample size. Although the largest sample size manipulated in the study is 600, this finding shows that the BNP selection GCM has promising potentials to analyze nonnormality and missingness with big data. Future studies can further investigate the performance of the proposed models with large sample size and explore potentials for these methods in big data analysis.

Note that convergence rates are high across all models in this study, even

for complicated nonparametric Bayesian approaches. We think an important reason is that informative priors are used for the DP concentration parameter α. The DP concentration parameter α affects the dispersion of the distributions and thus it affects the construction of the nonparametric models. How the nonparametric distribution is being constructed can directly affect how well the model converges. Therefore, we think the specification of the α parameter may affect the convergence rate. The performance of model convergence is always essential in Bayesian methods and how the model converges has practical significance. With complicated nonparametric Bayesian approaches, future work can assess the impact of the concentration parameter α on the convergence of the nonparametric Bayesian approach.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer.

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological methods & research*, 28(3):301–309.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, 112(4):545.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pages 1152–1174.

Arbuckle, J. L., Marcoulides, G. A., and Schumacker, R. E. (1996). Full information estimation in the presence of incomplete data. *Advanced structural equation modeling: Issues and techniques*, 243:277.

Azuero, A., Pisu, M., McNees, P., Burkhardt, J., Benz, R., and Meneses, K. (2010). An application of longitudinal analysis with skewed outcomes. *Nursing research*, 59(4):301.

Bernier, J., Feng, Y., and Asakawa, K. (2011). Strategies for handling normality assumptions in multi-level modeling: a case study estimating trajectories of health utilities index mark 3 scores. *Health Rep*, 22:45–51.

Best, N. G., Spiegelhalter, D. J., Thomas, A., and Brayne, C. E. (1996). Bayesian analysis of realistically complex models. *Journal of Royal Statistical Society, A*, 159:323–342.

Black, A. C., Harel, O., and Matthews, G. (2012). Techniques for analyzing intensive longitudinal data with missing values.

Brandt, H. and Klein, A. G. (2015). A heterogeneous growth curve model for nonnormal data. *Multivariate behavioral research*, 50(4):416–435.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(4):287–316.

Buuren, S. v. and Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, pages 1–68.

Cain, M. K., Zhang, Z., and Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior research methods*, 49(5):1716–1735.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). Deviance information criteria for missing data models. *Bayesian analysis*, 1(4):651–673.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Collins, L. M., Schafer, J. L., and Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4):330.

Davey, A. et al. (2009). *Statistical power analysis with missing data: A structural equation modeling approach*. Routledge.

DeIorio, M. and Robert, C. P. (2002). Discussion of spiegelhalter et al. *Journal of the Royal Statistical Society, Series B*, 64:629–630.

Demirtas, H., Freels, S. A., and Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-gaussian continuous outcomes: a simulation assessment. *Journal of Statistical Computation and Simulation*, 78(1):69–84.

Demirtas, H. and Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in medicine*, 22(16):2553–2575.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):49–73.

Donner, A. (1982). The exclusion of patients from a clinical trial. *Statistics in medicine*, 1(3):261–265.

Enders, C. K. (2001a). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological methods*, 6(4):352.

Enders, C. K. (2001b). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8(1):128–141.

Enders, C. K. (2011a). Analyzing longitudinal data with missing values. *Rehabilitation psychology*, 56(4):267.

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

Enders, C. K. and Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, 8(3):430–457.

Escobar, M. D. (1994). Estimating normal means with a dirichlet process prior. *Journal of the American Statistical Association*, 89(425):268–277.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230.

Ferguson, T. S. et al. (1974). Prior distributions on spaces of probability measures. *The annals of statistics*, 2(4):615–629.

Ferron, J., Dailey, R., and Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, 37(3):379–403.

Gelfand, A. E. (1996). Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, pages 145–161.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.

Geman, S. and 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. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.

Ghahramani, Z. (2013). Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110553.

Ghosh, S. and Pahwa, P. (2008). Assessing bias associated with missing data from joint canada/us survey of health: An application. *JSM Biometrics Section*, pages 3394–3401.

Gold, M. S. and Bentler, P. M. (2000). Treatments of missing data: A monte carlo comparison of rbhdi, iterative stochastic regression imputation, and expectation-maximization. *Structural equation modeling*, 7(3):319–355.

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, Inc., New York.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pages 475–492. NBER.

Hsu, C., Yen, A., Chen, L., and Chen, H. (2015). Analysis of household data on influenza epidemic with bayesian hierarchical model. *Mathematical biosciences*, 261:13–26.

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc, New York.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

Karahalios, A., Baglietto, L., Carlin, J. B., English, D. R., and Simpson, J. A. (2012). A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC medical research methodology*, 12(1):96.

Kass, R. E. (1993). Bayes factors in practice. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 42(5):551–560.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

Kelava, A. and Brandt, H. (2014). A general non-linear multilevel structural equation mixture model. *Frontiers in psychology*, 5:748.

Kleinman, K. P. and Ibrahim, J. G. (1998a). A semi-parametric bayesian approach to generalized linear mixed models. *Statistics in Medicine*, 17(22):2579–2596.

Kleinman, K. P. and Ibrahim, J. G. (1998b). A semiparametric bayesian approach to the random effects model. *Biometrics*, pages 921–938.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):1458–1468.

Lange, K. L., Little, R. J., and Taylor, J. M. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896.

Lavori, P. W., Dawson, R., and Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in medicine*, 14(17):1913–1925.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, 83(404):1198–1202.

Little, R. J. and Rubin, D. B. (1986). Statistical analysis with missing data.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Lu, Z. L., Zhang, Z., and Cohen, A. (2015). Model selection criteria for latent growth models using bayesian methods. In *Quantitative psychology research*, pages 319–341. Springer.

Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., and Thomas, A. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.

McArdle, J. J. and Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association.

Meng, X.-L. et al. (1994). Posterior predictive p-values. *The annals of statistics*, 22(3):1142–1160.

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin*, 105(1):156.

Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes' factors. *Biometrics*, 65(3):962–969.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G. (2014). *Handbook of missing data methodology*. Chapman and Hall/CRC.

Muthén, B., Asparouhov, T., Hunter, A. M., and Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: alternative analyses of the star* d antidepressant trial. *Psychological methods*, 16(1):17.

Muthén, B., Kaplan, D., and Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3):431–462.

Muthén, L. K. and Muthén, B. O. (2004). *Mplus user's guide: Statistical analysis with latent variables: User'ss guide*. Muthén & Muthén.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., and Boker, S. M. (2016). Openmx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81(2):535–549.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4):372–411.

Olinsky, A., Chen, S., and Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151(1):53–79.

Osborne, J. (2005). Notes on the use of data transformations. *Practical Assessment, Research and Evaluation*, pages 42–50.

Osborne, J. W. and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1):6.

Peng, C.-Y. J., Harwell, M., Liou, S.-M., Ehman, L. H., et al. (2006). Advances in missing data methods and implications for educational research. *Real data analysis*, 3178.

Peugh, J. L. and Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research*, 74(4):525–556.

Pigott, T. D. (2001). A review of methods for missing data. *Educational research and evaluation*, 7(4):353–383.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, 9(3):523–539.

Raftery, A. E. (1999). Bayes factors and bic: Comment on a critique of the bayesian information criterion for model selection. *Sociological Methods & Research*, 27(3):411–427.

Richardson, S. (2002). Discussion of spiegelhalter et al. *Journal of the Royal Statistical Society, Series B*, 64:626–627.

Roth, P. L. and Switzer III, F. S. (1995). A monte carlo analysis of missing data techniques in a hrm setting. *Journal of Management*, 21(5):1003–1023.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall/CRC.

Schafer, J. L. and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2):147.

Schafer, J. L. and Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate behavioral research*, 33(4):545–571.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3):333–343.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650.

Shi, D. and Tong, X. (2017). The impact of prior information on bayesian latent basis growth model estimation. *SAGE Open*, 7(3):2158244017727039.

Shi, D. and Tong, X. (2020). Mitigating selection bias: a bayesian approach to two-stage causal modeling with instrumental variables for nonnormal missing data. *Sociological Methods & Research*, page 0049124120914920.

Shin, T., Davison, M. L., and Long, J. D. (2009). Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1):70–98.

Shriner, D. and Yi, N. (2009). Deviance information criterion (dic) in bayesian multiple qtl mapping. *Computational statistics & data analysis*, 53(5):1850–1860.

Song, X.-Y., Pan, J.-H., Kwok, T., Vandenput, L., Ohlsson, C., and Leung, P.-C. (2010). A semiparametric bayesian approach for structural equation models. *Biometrical journal*, 52(3):314–332.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). Winbugs user manual.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and 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.

Switzer III, F. S., Roth, P. L., and Switzer, D. M. (1998). Systematic data loss in hrm settings: A monte carlo analysis. *Journal of Management*, 24(6):763–779.

Taljaard, M., Donner, A., and Klar, N. (2008). Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical journal*, 50(3):329–345.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Tong, X. and Zhang, Z. (2019). Robust bayesian approaches in growth curve modeling: Using students t distributions versus a semiparametric method. *Structural Equation Modeling: A Multidisciplinary Journal*, pages 1–17.

Wang, L., Zhang, Z., McArdle, J. J., and Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate behavioral research*, 43(3):476–496.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data.

Yuan, K.-H., Marshall, L. L., and Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika*, 67(1):95–121.

Yuan, K.-H. and Zhang, Z. (2012). Structural equation modeling diagnostics using r package semdiag and eqs. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(4):683–702.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American statistical association*, 86(413):79–86.

Zhang, Z. (2016). Modeling error distributions of growth curve models through bayesian methods. *Behavior research methods*, 48(2):427–444.

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

Zhang, Z., Wang, L., et al. (2012). A note on the robustness of a full bayesian method for nonignorable missing data analysis. *Brazilian Journal of Probability and Statistics*, 26(3):244–264.

Zu, J. and Yuan, K.-H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research*, 45(1):1–44.