Robust Bayesian Approaches in Growth Curve Modeling: Using Student's t Distributions versus

a Semiparametric Method

Xin Tong

University of Virginia

Zhiyong Zhang

University of Notre Dame

Author Note

Correspondence concerning this article should be addressed to Xin Tong, Department of

Psychology, University of Virginia, Charlottesville, VA 22903. Email: xt8b@virginia.edu.

Abstract

Despite broad applications of growth curve models, few studies have dealt with a practical issue -- nonnormality of data. Previous studies have used Student's t distributions to remedy the nonnormal problems. In this study, robust distributional growth curve models are proposed from a semiparametric Bayesian perspective, in which intraindividual measurement errors follow unknown random distributions with Dirichlet process mixture priors. Based on Monte Carlo simulations, we evaluate the performance of the robust semiparametric Bayesian method and compare it to the robust method using Student's t distributions as well as the traditional normal-based method. We conclude that the semiparametric Bayesian method is more robust against nonnormal data. An example about the development of mathematical abilities is provided to illustrate the application of robust growth curve modeling, using school children's Peabody Individual Achievement Test mathematical test scores from the National Longitudinal Survey of Youth 1997 Cohort.

Keywords: Semiparametric Bayesian, Robust method, Dirichlet process mixture, Growth curve modeling.

Robust Bayesian Approaches in Growth Curve Modeling: Using Student's *t* Distributions versus

a Semiparametric Method

Growth curve modeling is one of the most frequently used analytical techniques for longitudinal data analysis (e.g., McArdle and Nesselroade, 2014; Meredith and Tisak, 1990). In growth curve modeling, repeated measures of dependent variables are represented as a function of time and possible covariates, and the function mean is the mean growth. Individual variations around the mean growth curve are due to random effects and intraindividual measurement errors. Traditional growth curve analysis typically assumes that the random effects and intraindividual measurement errors are normally distributed. Although the normality assumption makes growth curve models easy to estimate, data in social and behavioral sciences are commonly collected using surveys or questionnaires and thus often are nonnormal (Cain et al., 2017; Micceri, 1989) because of nonnormal population distributions or data contamination. Ignoring the nonnormality of data may lead to inefficient or even biased parameter estimates, and statistical inferences based on common test statistics and fit indices could be misleading (e.g., Maronna, Martin, and Yohai, 2006; Yuan and Bentler, 2001). In this article, we propose a semiparametric Bayesian method to handle the nonnormality issue in growth curve modeling and compare the proposed method to existing robust Bayesian approaches using Student's *t* distributions.

Researchers have become more and more keenly aware of the large influence that nonnormality has upon model estimation (e.g., Hampel, Ronchetti, Rousseeuw, and Stahel, 1986; Huber, 1981; Yuan, Bentler, and Chan, 2004) and have developed strategies aiming to provide reliable parameter estimates and inferences when the normality assumption is violated. A straightforward and feasible strategy is to either transform the data so that they are close to being normally distributed, or directly delete potential outliers before data analysis. However, data transformation often makes the interpretation of the model estimation results complicated. Simply deleting outliers may reduce efficiency as the resulting inferences may fail to reflect uncertainty in the exclusion process (e.g., Lange, Little, and Taylor, 1989). Moreover, diagnosing multivariate outliers is a challenging task (e.g., Filzmoser, 2005; Peña and Prieto, 2001). Tong

and Zhang (2017) proposed six methods to detect outlying observations in growth curve modeling and concluded that the greatest chance of success comes from the use of multiple methods, comparing their results and making a decision based on research purposes. Therefore, alternatively, many researchers (e.g., Savalei and Falk, 2014; Yuan and Zhang, 2012) have recommended the application of robust methods and statistics to protect data from being distorted by the presence of outliers or nonnormality. These methods either downweight the potential outliers as a transformation technique (e.g., Yuan and Bentler, 1998) or assume that the data come from certain nonnormal distributions such as a t distribution or a mixture of normal distributions (e.g., Asparouhov and Muthén, 2016; Muthén and Shedden, 1999; Tong and Zhang, 2012).

Recently, robust methods from Bayesian perspectives have drawn growing interest because Bayesian methods have many advantages. First, estimating models with complex structures often involves high dimensional integration and thus is computationally intensive. Sampling methods such as Markov Chain Monte Carlo (MCMC) under the Bayesian framework can handle this problem relatively easily. Second, Bayesian estimation can conveniently infer parameters that do not have symmetric distributions (e.g., variance parameters), whereas it is difficult or computationally intensive to capture the asymmetric nature for parameters using frequentist methods. Third, with Bayesian methods, prior information can be incorporated via informative priors to make parameter estimates more efficient. Furthermore, Bayesian methods naturally accommodate missing data without requiring new techniques for inference and missing data can be taken into account at the same time as parameter estimation. Because of these strengths, more and more Bayesian estimation methods are employed in robust analysis.

From the Bayesian perspective, one approach of robust methods to account for nonnormality is to replace normal distributions by Student's t distributions in the model as the degrees of freedom of t distribution can control the robustness (Lange, Little, and Taylor, 1989). For example, Pinheiro, Liu, and Wu (2001) proposed a robust version of the linear mixed-effect model to remedy the distributional deviation from the normality assumption, in which normal distributions for the random effects and measurement errors were both replaced by multivariate t

distributions. A few studies have directly discussed this approach in growth curve analysis. Tong and Zhang (2012) and Zhang, Lai, Lu, and Tong (2013) suggested modeling heavy-tailed data and outliers in growth curve modeling using t distributions and provided online software to conduct the robust analysis. The two articles demonstrated that the robust growth curve modeling based on t distributions is easy to understand and implement, and thus potentially would greatly promote the adoption of robust growth curve analyses. However, although there are many advantages in using t distributions for robust data analysis (e.g., Tong and Zhang, 2012), Student's t distribution has a parametric form and thus still has a restriction on the distribution of data. For example, using t distributions may be sensitive to skewed data or mixture data, or even break down under some circumstances (e.g., Azzalini and Genton, 2008; Zhang, 2013). Note that researchers have proposed robust methods based on skew-normal distributions (e.g., Asparouhov and Muthén, 2016; Zhang, 2013) to overcome the problem of skewed data. However, again, skew-normal distributions have parametric forms and are limited to describing certain shape of distributions. Growth mixture models, first introduced by Muthén and Shedden (1999), provide another useful approach to reduce the influence of the nonnormality problem. These models assume that individuals can be grouped into a finite number of classes having distinct growth trajectories. Although growth mixture models are flexible, some difficult issues, including choice of the number of latent classes and selection of growth curve models within each class, have to be tackled. The typical strategy fixes the number of latent classes in advance at a small value (e.g., 2-4), models the growth trajectories parametrically with a polynomial function, and assesses model fits using criteria such as AIC, BIC, and likelihood-based tests (Nylund, Asparoubov, and Muthén, 2007). Semiparametric Bayesian methods, sometimes referred to as nonparametric Bayesian methods (e.g., Gershman and Blei, 2012; Müller and Mitra, 2004), provide a different approach to this problem. Rather than comparing models that vary in complexity, semiparametric Bayesian methods are to fit a single model that can adapt its complexity to the data and allow the complexity to grow as more data are observed.

While parametric models can only capture a bounded amount of information from data,

semiparametric Bayesian models allow for a richer and larger class of models. Müller and Mitra (2004) pointed out that restriction to a parametric family can mislead investigators into an inappropriate illusion of posterior certainty. On the contrary, semiparametric Bayesian methods are adaptive and have proven to be a valuable tool for discovering complicated patterns in data due to their great flexibility. There are two typical building blocks for semiparametric Bayesian models: Gaussian process and Dirichlet process, where Gaussian process is a distribution over functions that can be used for modeling functions and classification, and Dirichlet process is a distribution over probability measures that can be used for density estimation and clustering. For robust analysis against nonnormality, semiparametric Bayesian methods with Dirichlet process priors are desirable, by viewing latent variables or measurement errors as from unknown random distributions. Fueled by the MCMC ideas and the development of Bayesian software (e.g, Lunn, Jackson, Best, Thomas, and Spiegelhalter, 2013), many researchers have discussed the advantages and flexibility of using the semiparametric Bayesian methods (e.g., Fahrmeir and Raach, 2007; Ghosal et al., 1999; Hjort, 2003; Hjort et al., 2010; Müller and Mitra, 2004; MacEachern, 1999) and have applied these methods to models with complex structures. For example, Bush and MacEachern (1996), Kleinman and Ibrahim (1998), and Brown and Ibrahim (2003) used Dirichlet process mixtures for random effects distributions. Ansari and Iyengar (2006) used Dirichlet components to define a semiparametric dynamic choice model. Burr and Doss (2005) used a conditional Dirichlet process for the random effects distribution within a meta-analysis application. Dunson (2006) used dynamic mixtures of Dirichlet process to allow a latent variable distribution to change nonparametrically across groups. For categorical data analysis, Dirichlet process mixtures of multinomial distributions have been studied and applied to missing data through multiple imputation techniques to capture complex dependencies especially in high dimensions (Si and Reiter, 2013; Si et al., 2015). Semiparametric Bayesian approach has also been developed for structural equation models to relax the assumption that the distribution of the latent variables is normal (e.g., Lee, Lu, and Song, 2008; Yang and Dunson, 2010). As far as we are aware of, measurement errors or residuals in these models are still normally distributed

although it is pointed out in Müller and Mitra (2004) that the nonparametric model extension can go towards the direction of nonparametric residual distributions.

Despite the popularity of growth curve modeling, the prevalence of nonnormal data, and the flexibility of semiparametric Bayesian methods, no study has been directly conducted on semiparametric Bayesian growth curve modeling. Therefore, the main contributions of this article is to (1) propose a semiparametric Bayesian approach for growth curve modeling to relax the normality assumption imposed in traditional normal-based analysis (especially on measurement errors); and (2) evaluate the performance of the semiparametric Bayesian method and compare it with the robust method using Student's t distributions which is a parametric analysis. The robust method using Student's t distributions is selected for comparison because it is relatively more broadly used in practice as statistical software has been developed for it to facilitate the implementation. In the next section, we briefly review the traditional Bayesian growth curve modeling and the robust approach using Student's t distributions. After that, robust semiparametric Bayesian growth curve modeling is proposed. In addition, the comparison between semiparametric Bayesian models and finite growth mixture models is discussed. Then in the subsequent section, simulation studies are carried out to show the effectiveness of semiparametric Bayesian methods in comparison to the traditional growth curve modeling as well as the robust growth curve modeling using Student's t distributions. Finally, we illustrate the application of the semiparametric Bayesian methods through an example with the Peabody Individual Achievement Test math data from the National Longitudinal Survey of Youth 1997 Cohort (Bureau of Labor Statistics, U.S. Department of Labor, 2005). We end the article by summarizing our findings with recommendations.

Robust Bayesian Growth Curve Modeling

Bayesian growth curve modeling: a brief review

Growth curve models are used to analyze longitudinal data in which the same subjects are observed repeatedly over time on the same tests. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be a $T \times 1$ random

vector and y_{ij} be an observation for individual i at time j ($i = 1, \ldots, N; j = 1, \ldots, T$). Here N is the sample size and T is the total number of measurement occasions. A typical form of unconditional growth curve models can be expressed as

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

where $\mathbf{\Lambda}$ is a $T \times q$ factor loading matrix determining the growth trajectories, $\mathbf{b}_i$ is a $q \times 1$ vector of random effects, and $\mathbf{e}_i$ is a vector of intraindividual measurement errors. The vector of random effects $\mathbf{b}_i$ varies for each individual, and its mean, $\boldsymbol{\beta}$, represents the fixed effects. The residual vector $\mathbf{u}_i$ represents the random component of $\mathbf{b}_i$. Because $\boldsymbol{\beta}$ is constant across individuals, $\mathbf{b}_i$ and $\mathbf{u}_i$ share the same type of distribution with different means.

Traditional growth curve models typically assume that both $\mathbf{e}_i$ and $\mathbf{u}_i$ follow multivariate normal distributions such that $\mathbf{e}_i \sim MN_T(\mathbf{0}, \mathbf{\Phi})$ and $\mathbf{u}_i \sim MN_q(\mathbf{0}, \mathbf{\Psi})$, where MN denotes a multivariate normal distribution and the subscript denotes its dimension. The $T \times T$ matrix $\mathbf{\Phi}$ and the $q \times q$ matrix $\mathbf{\Psi}$ represent the covariance matrices of $\mathbf{e}_i$ and $\mathbf{u}_i$, respectively. The intraindividual measurement error structure is usually simplified to $\mathbf{\Phi} = \sigma_e^2 \mathbf{I}$ where σ_e^2 is a scale parameter. By this simplification, we assume homogeneity of error variance across time and measurement errors are uncorrelated at different time points. Given the current specification of $\mathbf{u}_i$, $\mathbf{b}_i \sim MN_q(\boldsymbol{\beta}, \mathbf{\Psi})$.

Special forms of growth curve models can be derived from the preceding form. For example, if

$$
\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix}, \mathbf{b}_i = \begin{pmatrix} L_i \\ S_i \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_L \\ \beta_S \end{pmatrix}, \mathbf{\Psi} = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix},
$$

the model represents a linear growth curve model with random intercept (initial level) L_i and random slope (rate of change) S_i. The average intercept and slope across all individuals are β_L and β_S, respectively. In $\mathbf{\Psi}$, σ_L^2 and σ_S^2 represent the variability (or interindividual differences)

around the mean intercept and the mean slope, respectively, and σ_{LS} represents the covariance between the latent intercept and slope.

To estimate growth curve models, Bayesian methodology can be applied. Bayesian methods for complex data analysis have been made popular in the past few decades because of its advantages as described previously (e.g., Lee and Shi, 2000; Lee and Song, 2004; Lee and Xia, 2008; Serang et al., 2015; Zhang et al., 2007b; Tong and Zhang, 2012). The basic idea of Bayesian methods is to obtain the posterior distributions of model parameters by combining the likelihood function and the priors. When priors are uninformative or weakly informative, the likelihood dominates and thus results from Bayesian estimation are similar to those from maximum likelihood estimation. For a typical unconditional growth curve model, we define the joint prior distribution of model parameters by $p(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ and denote the likelihood function as L. The joint posterior distribution of model parameters is

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

where $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_N')'$. This integral is difficult to solve in practice. Instead, MCMC methods (e.g., Gibbs sampling; Robert and Casella, 2004) are often applied to obtain parameter estimates and statistical inferences. We first obtain conditional posterior distributions for the parameters, then by iteratively drawing samples from the conditional posterior distributions, we obtain empirical marginal distributions of the model parameters and make statistical inferences based on the empirical marginal distributions (Geman and Geman, 1984). Detailed algorithm can be found in Song and Lee (2012), Zhang et al. (2013), etc.

Robust Bayesian growth curve modeling using Student's t distributions

The traditional growth curve analysis discussed above is based upon the normality assumption of random effects and intraindividual measurement errors. However, practical data in social and behavioral sciences are rarely normal due to data contamination or nonnormal population distributions. Without taking the nonnormality problem into consideration, we may obtain inefficient or even incorrect parameter estimates in model estimation (e.g., Yuan and

Bentler, 2001; Yuan and Zhang, 2012). Studies to deal with the adverse effects of nonnormality on parameter estimates, standard errors, and test statistics have been carried out in growth curve analysis. In a growth curve model, the nonnormality may occur in the random effects, in the intraindividual measurement errors, or both (Pinheiro et al., 2001). Motivated by a real dataset from an orthodontic study, Pinheiro et al. proposed a robust hierarchical linear mixed-effects model in which the random effects and the intraindividual errors follow multivariate t distributions, with known or unknown degrees of freedom.

By using multivariate t distributions, extreme values in a dataset can be downweighted. Suppose k dimensional data $\mathbf{x}$ follows a multivariate t distribution, with $k \times 1$ location vector $\boldsymbol{\mu}$, $k \times k$ shape matrix $\boldsymbol{\Sigma}$, and degrees of freedom ν, denoted by $MT(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. The probability density function of $\mathbf{x}$ is

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left[(\nu+k)/2\right]}{\Gamma(\nu/2)\nu^{k/2}\pi^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{-(\nu+k)/2}.$$

The maximum likelihood estimates of model parameters $\hat{\boldsymbol{\theta}}$ satisfy $\Sigma_{i=1}^{N} w_i \boldsymbol{A}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})_i^{-1}(\mathbf{x}_i - \boldsymbol{\mu}(\hat{\boldsymbol{\theta}})_i) = 0$ (Lange et al., 1989), where N is the sample size, $\boldsymbol{A}_i$ is the matrix of partial derivatives of $\boldsymbol{\mu}(\boldsymbol{\theta})_i$ with respect to $\boldsymbol{\theta}$, and $w_i = \frac{\nu+\tau_i}{\nu+\delta_i^2}$ is the weight assigned to case i (τ_i is the dimension of $\boldsymbol{\theta}$ for case i, and δ_i^2 is the squared Mahalanobis distance $\delta_i^2 = (\mathbf{x}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_i)$ for case i). Thus, potential outliers can be downweighted in the model estimation process because lower weights will be assigned to cases with large squared Mahalanobis distances, given fixed degrees of freedom ν and dimensions τ_i.

From a Bayesian perspective, Tong and Zhang (2012) proposed four types of distributional growth curve models where the random effects $\mathbf{u}_i$ and intraindividual measurement errors $\mathbf{e}_i$ may follow either multivariate normal or t distributions. They concluded that four types of distributional growth curve models imply very different patterns in growth trajectories, and thus given an empirical data set, it is very important to specify the correct type of growth curve models. Lu and Zhang (2014) expanded the study to further conduct a robust growth mixture analysis with nonnormal missing data.

Although there are many advantages in using t distributions for robust data analysis (e.g.,

Lange, Little, and Taylor, 1989), a t distribution is symmetric and thus still has a restriction on the distribution of the data. Constraining inference to a specific parametric form may limit the scope of inferences in many situations. For example, the shape of the assumed population distribution is restricted so that it may not approximate the data well enough. Also, the complexity of the model is fixed. If more data are collected later bringing in diversity and more characteristics, a different model may need to be adopted. This could impose technical challenges when researchers want to synthesize research findings from different studies (e.g., meta-analysis). Consequently, we propose to use semiparametric Bayesian methods/models in this article because there is no need to make arbitrary and unverified distributional assumptions for the latent variables or the measurement errors as in the parametric modeling. We expect that semiparametric Bayesian models perform equally well as parametric models when the parametric models are correctly specified, and outperform the parametric models when they are mis-specified.

Semiparametric Bayesian growth curve modeling

Semiparametric Bayesian methods, sometimes also called nonparametric Bayesian methods in the literature, are based on distributions over spaces of distributions and are flexible in modeling the nonnormality (e.g., Tong, 2014; Lee et al., 2008). Unlike typical classical nonparametric methods such as rank and permutation tests, semiparametric Bayesian methods can provide full probability models for the data-generating process and provide posterior distributions of model parameters. The semiparametric Bayesian approach to solving problems is rapidly gaining popularity among practitioners as theoretical properties become increasingly better understood and computational hurdles are being removed (e.g., Neal, 2000). Below we first introduce Dirichlet process priors and then develop semiparametric Bayesian methods in growth curve modeling.

Dirichlet process (DP) priors. Within the semiparametric Bayesian scope, the traditional parametric assumption of a random vector $\boldsymbol{\xi}$ [e.g., $\boldsymbol{\xi} \sim N(\boldsymbol{\mu}_{\xi}, \boldsymbol{\Phi}_{\xi})$] is replaced by

$$\boldsymbol{\xi} \;\sim\; G,$$

where G is an unknown distribution function and its prior is a distribution over the distribution G. The Dirichlet process (DP, Ferguson, 1973, 1974) is a widely used prior for unknown distributions. DP generates random distributions and is characterized by two hyperparameters, α and G_0. G_0 is a base distribution, which represents the central or "mean" distribution in the distribution space, while the precision parameter α governs how close realizations of G are to G_0. For any measurable partitions $P_1, \ldots, P_k$ of the sample space $\mathcal{X}$, $(G(P_1), \ldots, G(P_k))$ follows a Dirichlet distribution $Dirichlet(\alpha G_0(P_1), \ldots, \alpha G_0(P_k))$. For example, if $\mathcal{X}$ is the real space and $P = (-\infty, x]$ where x is a real number, then

$$G(x) \sim Dirichlet(\alpha G_0(x), \alpha(1 - G_0(x))).$$

Thus,

$$
\begin{aligned}
E(G(x)) &= G_0(x), \\
Var(G(x)) &= \frac{G_0(x)(1 - G_0(x))}{\alpha + 1}.
\end{aligned}
$$

Ferguson (1973) pointed out that DP is a conjugate prior and has two desirable properties: (1) its support is sufficiently large, and (2) the posterior distribution is analytically manageable. He further explained that the posterior of G is $DP(\tilde{\alpha}, \tilde{G}_0)$, where $\tilde{\alpha} = \alpha + N$ and

$$\tilde{G}_0 = \frac{\alpha}{\alpha + N}G_0 + \frac{N}{\alpha + N}G_N,$$

where G_N is the empirical distribution function of the data. Thus, the posterior point estimate of G, $E(G|data) = \tilde{G}_0$, is a weighted average of two distributions: G_0 and G_N. If $\alpha = 0$, the posterior point estimate is G_N, which is nonparametric. When α approaches to infinity, the posterior point estimate approaches to G_0, which is parametric. In practice, the hyperparameter α is usually specified to follow $Gamma(a_1, a_2)$, which is neither 0 nor infinity. As suggested by Ishwaran (2000), $a_1 = a_2 = 2$ is usually selected to encourage both small and large values of α. Thus, we consider the posterior point estimate of G as semiparametric.

Stick-breaking construction.   There are different representations of DP and the most common one is based on the stick-breaking construction, developed by Sethuraman (1994). Let

$q_1, q_2, \ldots, q_k, \ldots \sim Beta(1, \alpha)$. Define

$$p_k = q_k \prod_{j=1}^{k-1} (1 - q_j).$$

Then,

$$G = \sum_{k=1}^{\infty} p_k \delta_{\boldsymbol{\xi}_k^*},$$

where $\delta_{\boldsymbol{\xi}_k^*}$ is the Dirac probability measure and $\boldsymbol{\xi}_k^* \sim G_0$. Note that $\sum_{k=1}^{\infty} p_k = 1$ so it guarantees G to be a distribution. Sethuraman showed that draws from the stick-breaking are indeed DP distributed under general conditions. From this constructive way, we can tell that $G(\cdot)$ is a discrete distribution as

$$G(\cdot) = \begin{cases} \boldsymbol{\xi}_1^*, & with \ p = p_1 \\ \\ \boldsymbol{\xi}_2^*, & with \ p = p_2 \\ \\ \vdots & \vdots \\ \\ \boldsymbol{\xi}_k^*, & with \ p = p_k \\ \\ \vdots & \vdots \end{cases}.$$

The discrete nature of DP was also described by Blackwell and MacQueen (1973) through the Polya urn scheme. Thus, in a density estimation problem, where $y_i \sim G(\cdot)$, $i = 1, \ldots, N$, and y_i is exchangeable (i.e., independently drawn from some unknown distribution), it would be inappropriate to assume $G \sim DP$ if the distribution is known to be absolutely continuous. To define a continuous distribution, a convenient way is to use DP as the basis of a mixture model, for example, a mixture of $N(\mu_k, \sigma_k^2)$ with mixing proportions defined by p_k. This model is known as the DP mixture model (Antoniak, 1974; Escobar and West, 1995):

$$\begin{aligned} y_i | \theta_i &\sim F(\theta_i), \\ \theta_i | G &\sim G, \\ G &\sim DP(\alpha, G_0). \end{aligned}$$

Theoretically, there are infinite number of mixture components as $k = 1, \ldots, \infty$, given an arbitrarily flexible choice of distributional shapes. In practice, a finite number of mixture

components would be good enough, and this number is taken into account by DP. Smaller values of DP precision parameter α result in a smaller number of mixture components. If α approaches infinity, there would be N mixture components, one associated with each observation.

Semiparametric Bayesian methods in growth curve modeling. Because semiparametric Bayesian methods are flexible in modeling the nonnormality, we propose to use them for robust growth curve modeling. The parametric assumptions of the intraindividual measurement errors $\mathbf{e}_i$ (either normal or t distributed) are replaced by an unknown random distribution G_e. From a Bayesian perspective, a prior needs to be specified for G_e. Because intraindividual measurement errors $\mathbf{e}_i$ follow a continuous distribution, we use a DP mixture prior:

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

That is $\mathbf{e}_i \sim G_e$, and $G_e \sim DPM$ is a mixture of multivariate normal distributions where the mixing measure has a Dirichlet process prior. Note that the mean of intraindividual measurement errors $\mathbf{e}_i | \mathbf{\Phi}_i$ is set to be $\mathbf{0}$ when the growth curve model is correctly specified. We can obtain the distribution G_e by the truncated stick-breaking construction (e.g., Lunn, Jackson, Best, Thomas, and Spiegelhalter, 2013; Sethuraman, 1994). Assume that the data can be represented by a maximum of C ($1 \leq C \leq N$) possible mixture components. For $q_1, q_2, \ldots, q_C \sim Beta(1, \alpha)$, define $p'_k = q_k \prod_{j=1}^{k-1}(1 - q_j)$, $k = 1, \ldots, C$. Then, the mixing proportion p_k is obtained by

$$
p_k = \frac{p'_k}{\sum_{j=1}^{C} p'_j},
$$

to satisfy that $\sum_{k=1}^{C} p_k = 1$. Thus, the unknown distribution G_e can be constructed below, which is a mixture of multivariate normal distributions:

$$G_e = \begin{cases} MN(\mathbf{0}, \mathbf{\Phi}^{(1)}), & with\ p = p_1 \\ MN(\mathbf{0}, \mathbf{\Phi}^{(2)}), & with\ p = p_2 \\ \vdots & \vdots \\ MN(\mathbf{0}, \mathbf{\Phi}^{(C)}), & with\ p = p_C \end{cases},$$

where $\mathbf{\Phi}^{(k)}, k = 1, \ldots, C$ are parameters of the multivariate normal distribution in the kth component. Inverse Wishart priors for $\mathbf{\Phi}^{(k)}$, $p(\mathbf{\Phi}^{(k)}) = IW(n_0, W_0)$, are used, where n_0 and W_0 are hyperparameters. Following Lunn et al. (2013, page 294), we can fix the shape parameter n_0 at a specific number and assign the uncertain scale matrix W_0 an inverse Wishart prior.

In general, G_e is a symmetric distribution centered at $\mathbf{0}$. Samples drawn from G_e can be multimodal. G_e can approximate a normal distribution, a t distribution, or other distributions. So G_e may fit the data better than a t distribution. The measurement error for the ith individual, $\mathbf{e}_i$, comes from $MN(\mathbf{0}, \mathbf{\Phi}^{(k)})$ with the probability p_k. We should point out that the measurement errors for different individuals may be drawn from the same component. If $\mathbf{e}_i, i = 1, \ldots, N$ are from K different distributions among $MN(\mathbf{0}, \mathbf{\Phi}^{(k)}), k = 1, \ldots, C$, K is the number of clusters for $\mathbf{e}_i$. Clearly, $K \leq C$, and within each cluster, $\mathbf{e}_i$s come from the same distribution.

Recall that in the traditional growth curve model, $\boldsymbol{\beta}, \mathbf{\Phi}$, and $\mathbf{\Psi}$ are the model parameters. In the semiparametric Bayesian growth curve model, $\boldsymbol{\beta}$ and $\mathbf{\Psi}$ are still model parameters and can be estimated in the same way. However, instead of estimating $\mathbf{\Phi}$ (the covariance matrix of $\mathbf{e}_i$) as in the traditional model, we obtain $\mathbf{e}_i$ and K. The estimate of K indicates the heterogeneity of intraindividual measurement errors $\mathbf{e}_i$. With a larger value of K, we are more confident to conclude that different individuals' measurement errors are distributed differently. West (1992) pointed out that K asymptotically follows a Poisson distribution

$$K = 1 + x, \ x \sim Poisson\left(\alpha\left(\gamma + logN\right)\right),$$

where γ is Euler's constant. Table 1 gives different percentiles of the distribution of K. With a

larger value of α and a larger sample size N, it is more likely to have more clusters. To obtain an estimate of $\mathbf{\Phi}$, we let $\mathbf{e}_{i(s)}, i = 1, \ldots, N$ be the observations of $\mathbf{e}_i$ simulated from the posterior distribution in the sth Gibbs sampler iteration, and let $\mathbf{\Phi}_{(s)}$ be the corresponding sample covariance matrix. An estimate of $\mathbf{\Phi}$ can be taken as the mean of $\mathbf{\Phi}_{(s)}$, averaging over all the Gibbs sampler iterations after the burn-in period.

In the present study for the semiparametric Bayesian growth curve modeling, only the distribution of intraindividual measurement errors is replaced by an unknown distribution with the DP mixture prior. A similar strategy can be taken for the random effects $\mathbf{u}_i$. In this article, we focus on the model described above because it is more comparable to the robust Bayesian method proposed in Zhang, Lai, Lu, and Tong (2013) and the comparison will be discussed in the next section. Information about semiparametric Bayesian growth curve models with random effects following a random distribution is available in Tong (2014). Tong (2014) concluded that it is important to specify the correct type of model for practical data analyses. Outlying observation diagnostic methods in growth curve modeling (Tong and Zhang, 2017) have been developed to distinguish outliers (extreme scores in measurement errors) and leverage observations (extreme scores in random effects). Such methods can be used first to figure out whether the random effects or the measurement errors follow normal distributions, so that we can then select the correct type of models and conduct data analysis.

Semiparametric Bayesian growth curve modeling versus growth mixture modeling. For the semiparametric Bayesian growth curve models, the traditional normal distribution of intraindividual measurement errors is replaced by a random distribution with a DP mixture prior. Basically, the random distribution is a mixture of multivariate normal distributions with the mixing proportions generated following certain rules (e.g., stick-breaking construction). Similar to the growth mixture modeling, semiparametric Bayesian growth curve modeling also identify multiple unobserved sub-populations/clusters and describing longitudinal change within each unobserved sub-population/cluster. However, the semiparametric Bayesian growth curve modeling is different from the growth mixture modeling in several ways.

First, the traditional mixture modeling approach to clustering requires the number of clusters to be determined by comparing models with different number of clusters. Although substantive researchers typically use a combination of criteria (e.g., AIC and BIC) to guide the decision, there is not common acceptance of the best criteria despite various suggestions (Nylund et al., 2007). The Bayesian semiparametric approach sidesteps the problem of finding the correct number of clusters by assuming infinitely many mixture components. It provides a compelling alternative to the traditional finite mixture model paradigm and can automatically infer an adequate model size/complexity from the data, without explicitly doing Bayesian model comparisons. The risks of using too few classes are reduced while fully estimating uncertainty in posterior distributions. Second, in growth mixture modeling, the number of clusters is fixed, whereas in the semiparametric Bayesian growth curve modeling, the actual number of clusters used to model data is not fixed, and can be automatically inferred from data using the Bayesian posterior inference framework. So the model complexity is part of the posterior distribution. Semiparametric Bayesian modeling provides a way of getting very flexible models since inflexible models (e.g., a mixture of 5 Gaussians) may yield unreasonable inferences. Third, for growth mixture models, adding one additional cluster brings in more parameters to be estimated. Thus, it is not practically possible to have many clusters when we conduct growth mixture analyses. In contrast, it is often not a problem to obtain a large number of clusters if we use the semiparametric Bayesian method. Because of the hierarchical feature of Bayesian modeling, parameters for the Dirichlet process are of interest since they can determine the clusters no matter how many clusters there are. Fourth, semiparametric Bayesian growth curve modeling allows future data to exhibit previously unseen clusters. As Dunson (2013) pointed out, when semiparametric Bayesian methods are applied in practice, the prior and the penalty that comes in through integrating over the prior in deriving the marginal likelihood tends to lead to allocation of all the individuals in the sample to relatively few clusters. Individuals in the population may potentially come from more clusters. When more data become available, some data may show patterns that do not belong to previous data. The semiparametric Bayesian model automatically adjusts itself. However, for

growth mixture modeling, we may need to fit a different set of models when more data are included in the study because a larger sample size probably indicates more clusters.

Note that the comparison between the semiparametric Bayesian growth curve modeling and finite growth mixture modeling is not conducted in our simulation study. As we explained, the semiparametric Bayesian method is essentially a mixture modeling with infinite number of latent classes. Therefore, as demonstrated in Dahlin et al. (2016), finite mixture modeling can be seen as an approximation of the infinite mixture modeling. Their numerical illustration showed that when the data record is finite, the two approaches reach the same conclusion.

Comparison of the Two Robust Approaches through a Simulation Study

The purpose of this simulation study is to compare the semiparametric Bayesian method with the existing robust method using Student's t distributions (Zhang et al., 2013) for growth curve modeling, in analyzing symmetric data, skewed data, mixture data, and data with outliers. The traditional normal-based method serves as a baseline for the comparison.

Study design

We focus on the linear growth curve model as discussed previously. In the model (see Figure 1), the fixed effects are given by $\boldsymbol{\beta} = (\beta_L, \beta_S)' = (6.2, 0.3)'$. These numbers are reasonable as they are the estimated fixed effects in the real data example. From Tong and Zhang (2012) and (Tong, 2014), it is known that the number of measurement occasions, the covariance between the latent intercept and slope, and the variance of the measurement errors do not significantly affect the relative performance of the robust methods. Therefore, we fix $T = 4$, $\sigma_e^2 = 0.5$, and $\sigma_{LS} = 0$ in this study.

We study the effects of two factors - sample size and population distribution of data. Three different sample sizes are evaluated: $N = 50, 200,$ and 500. The influence of population distribution on the three methods is evaluated by manipulating the distributions of intraindividual measurement errors of the growth curve model. In total, seven distributional conditions are considered, as shown in Table 2. As shown in Cain et al. (2017), these distributional conditions

have been observed in practice. For condition C1, data are generated from normal distributions as the baseline condition in comparison to nonnormal cases. For condition C2, data are still symmetrically distributed. However, they have longer-than-normal tails as the kurtosis is bigger. From conditions C3 to C4, the intraindividual measurement errors are generated from nonnormal distributions, with an increase of skewness and kurtosis. For condition C5, we generate mixture data with half of individuals' measurement errors following $N(0, \sigma_e^2)$ and the other half following $N(0, 5\sigma_e^2)$. Data in conditions C6 and C7 are generated with outliers in the intraindividual measurement errors. The number of outliers is 10 at each measurement occasion. We consider the effects of the geometry of outliers on the performance of the two models. Outliers can locate on one side or both sides of the population distribution, resulting in asymmetric or symmetric sample distributions, respectively. Outliers on one side of the population distribution are generated from the normal distribution $e_j \sim N(7\sigma_e, \sigma_e^2)$, and outliers on both sides of the population distribution are generated from either $e_j \sim N(7\sigma_e, \sigma_e^2)$ or $e_j \sim N(-7\sigma_e, \sigma_e^2)$ with equal probability. The outliers are generated following the definition of outliers in growth curve modeling (Tong and Zhang, 2017; Yuan and Zhong, 2008).

For each simulation condition, a total of 500 data sets are generated. For each data set, the traditional normal-based method, robust method using t distributions, and semiparametric Bayesian method[1] are applied separately using free software R and OpenBUGS (Lunn et al., 2013). The programming code is available on this webpage: https://www.dropbox.com/sh/njhhnm4tdi4ahp9/AADXXBbB79qvp0x37YZSNeK7a?dl=0. We obtain the parameter estimate, bias, relative bias[2], standard error, mean squared error (MSE), and

---

[1] Notice that for semiparametric Bayesian method, the maximum of the potential number of mixture components C is set at 20 in the simulation. This is reasonable because the estimated number of clusters is always below 20 based Tong (2014). So setting C at 20 or a larger value such as 50 does not affect the model estimation. But a very small value of C (e.g., 5) may reduce the accuracy of the estimation. In addition, the hyperparameter α is specified to follow $Gamma(2, 2)$, as suggested by Ishwaran (2000).

[2] $relative\ bias = \begin{cases} \hat{\theta} \times 100\% & \theta = 0 \\ \dfrac{\hat{\theta} - \theta}{\theta} \times 100\% & \theta \neq 0 \end{cases}$, where θ is the true parameter value and $\hat{\theta}$ is the estimate of θ.

coverage probability (CP) of the 95% highest posterior density (HPD) credible intervals[3] for each parameter. In addition, Geweke tests (Geweke, 1992) are used to assess the convergence of Markov chains for all simulation replications.

Note that for the traditional growth curve modeling, priors of the model parameters are selected following Zhang et al. (2013). Fixed effects β has a noninformative diffuse prior $N(0, 10^6)$. The covariance matrix of the latent intercept and slope Ψ has an inverse-Wishart prior with an indentity scale matrix and degrees of freedom being 2. The variance of the measurement errors σ_e^2 also has a diffuse prior $inverse - gamma(0.001, 0.001)$. For the robust growth curve modeling using t distributions, the prior for the degrees of freedom of the t distribution is a uniform distribution $U(0, 100)$. For the semiparametric Bayesian growth curve modeling, DP mixture prior is used for the intraindividual measurement error $\mathbf{e}_i$ with $\alpha \sim Gamma(2, 2)$ and $\mathbf{\Phi}^{(k)} \sim IW(n_0, W_0)$ where n_0 is fixed at 2 and the diagonal elements of W_0 have inverse-gamma distributions. Detailed prior information is also provided in the shared OpenBUGS code.

Main results.   Table 3 presents the detailed parameter estimation results from the two robust estimation methods as well as the baseline normal-based method for conditions C1 when the sample size is 200. When sample size is 50 or 500, similar patterns are observed. For conditions C2-C7, data are not normally distributed. Both robust methods outperform the normal-based method, which is consistent with the existing literature (e.g., Tong and Zhang, 2012). In this section, we mainly compare the two robust methods based on the average of MSE and CP across parameters. Geweke tests suggest that all cases are converged. Detailed estimation results for all parameters in all conditions are available on the webpage: https://www.dropbox.com/sh/njhhnm4tdi4ahp9/AADXXBbB79qvp0x37YZSNeK7a?dl=0. We summarize our main findings below.

---

[3] Posterior credible interval, also called credible interval or Bayesian confidence interval, is analogical to the frequentist confidence interval. The 95% HPD credible interval $[l, u]$ satisfies: 1. $Prob(l \leq \theta \leq u | data) = 0.95$; 2. for $\theta_1 \in [l, u]$ and $\theta_2 \notin [l, u]$, $Prob(\theta_1 | data) > Prob(\theta_2 | data)$. In general, HPD intervals have the smallest volume in the parameter space of θ, and numerical methods have to be used to find HPD intervals.

For condition C1 ($e_1, \ldots, e_T \sim N(0, \sigma_e^2)$), Table 3 shows that the two robust methods and the traditional normal-based method perform equally well because the bias, empirical standard errors, MSEs and CPs for the estimated parameters are similar. For the robust method using t distributions, the estimated degrees of freedom is 55.661, indicating that the intraindividual measurement errors $\mathbf{e}_i$ are most likely to be normally distributed. For the semiparametric Bayesian method, the estimated number of clusters of $\mathbf{e}_i$ is 5.150, meaning that there are 5 different clusters for the measurement errors. The estimated K_e is related to the DP precision parameter α. In this study, the estimated value of α is 0.928 given the prior $\alpha \sim Gamma(2, 2)$. Notice that the estimated number of clusters of $\mathbf{e}_i$ is larger than 1 for normally distributed data. We may suspect whether there are overfitting problems. Since semiparametric Bayesian methods are data driven and α governs the model complexity to some degree, it is not prone to overfitting in theory. Since parameter estimations are very similar across the three methods applied to analyze the data, we do not further investigate overfitting problems in this article. More will be discussed in the discussion section.

For conditions C2-C7, both robust methods outperform traditional normal-based method in terms of bias and standard error. So we only summarize results from robust methods into Table 4[4]. Results from condition C1 are also included in this table as a benchmark. We calculate MSE and CP for each model parameter and average them over certain model parameters for each simulation condition to compare the two types of robust growth curve models. For conditions C1-C5, MSEs and CPs are averaged over all the six model parameters β_L, β_S, σ_L^2, σ_S^2, σ_{LS}, and σ_e^2. For conditions C6 and C7, MSEs and CPs are only averaged over β_L, β_S, σ_L^2, σ_S^2, and σ_{LS}, since the population parameter value for σ_e^2 is unkown due to the existence of outliers for the intraindividual measurement errors. In Table 4, on the rows are the different types of generated data and on the columns are the two types of robust methods used to analyze the generated data.

(1) As we discussed, for condition C1, the two robust methods perform about the same. When sample size is large, the semiparametric Bayesian method leads to CPs closer to the

---

[4] Note that all the detailed parameter estimation results are provided on our webpage in the same structure as Table 3.

nominal level 0.95.

(2) From conditions C2 to C4, the performance of the semiparametric Bayesian method becomes better than the performance of the robust method based on t distributions, and this pattern gets more and more prominent as data deviate more from normal distributions. For condition C2 ($e_1, \ldots, e_T \sim t_{(5)}(0, \sigma_e^2)$), the two methods perform similarly in estimating fixed effects and random effects related parameters β_L, β_S, σ_L^2, σ_S^2, and σ_{LS}. For the estimated σ_e^2, the robust method based on t distributions leads to a larger bias, a larger standard error, and thus a larger MSE and a worse CP. This is because the intraindividual measurement errors are generated from rescaled t distributions. Although the distribution of $\mathbf{e}_i$ is still symmetric, it is not distributed as the regular Student's t distribution. The estimated degrees of freedom is far from precise as the standard error of it is 9.58. For the semiparametric Bayesian method, the estimated number of clusters of $\mathbf{e}_i$ is 7.60, which is bigger than that in condition C1, meaning that there are more clusters for the measurement errors when they are not normally distributed. For conditions C3 and C4, the distribution of $\mathbf{e}_i$ is not symmetric anymore. The semiparametric Bayesian method performs consistently better than the t-distribution-based method for all the parameter estimates, especially the variance of measurement errors σ_e^2. Even for the estimation of the fixed effects parameter β_L, the difference between the performance of the two robust methods is obvious. For example, by switching from the robust t-distribution-based method to the semiparametric Bayesian method in condition C3, the bias of $\hat{\beta}_L$ is reduced by (0.081-0.021)/0.081=74%, and the CP increases from 0.80 to 0.94. Therefore, the semiparametric Bayesian method is more robust to the skewed data than the robust method using t distributions. For highly skewed data with heavy tails (e.g., condition C4), the bias of some parameter estimates (e.g., $\hat{\sigma}_L^2$) from the t-distribution-based method can be as high as 10 times bigger than that from the semiparametric Bayesian method. Given such results, the robust Bayesian method using t distributions should not be used for highly skewed data when comparing to the semiparametric Bayesian method.

Note that the df_e can control the robustness for the t-distribution-based method (e.g., Zhang et al., 2013) and the K_e can control the robustness for the semiparametric Bayesian method. From

conditions C1-C4, data distributions are getting more severely different from normal distributions with an increase of skewness and kurtosis. As a result, the estimated df_e decreases from 55.661 to 1.667 for the t-distribution-based method, and the estimated K_e increases from 5.150 to 9.812 for the semiparametric Bayesian method. Although both robust models have their accommodations to the distributional deviance from normal distributions, the semiparametric Bayesian method seems to be more robust to data generated from nonnormal populations.

(3) For condition C5 with mixture data, the semiparametric Bayesian method also outperforms the t-distribution-based robust method. Particularly, the estimated parameters from the semiparametric Bayesian method are more accurate and efficient. The CP is improved substantially when the semiparametric Bayesian method is used. For the semiparametric Bayesian growth curve model, the estimated K_e is 7.5557, bigger than that in condition C1, indicating that more clusters are needed for mixture data analysis.

(4) For conditions C6 and C7 where data contain outliers, the two robust methods provide similar MSEs and CPs regardless of the geometry of the outliers. By comparing conditions C3 and C4 with condition C6, we notice that although data under those conditions are all skewed, the comparisons between the performance of the two robust methods are different. This is consistent with the existing literature (Zhang et al., 2013). Although Student's t distribution is symmetric and sensitive to skewed data, practically, we find that if the skewness in the data is caused by outliers, the t-distribution-based robust method can still perform very well. However, if the skewness is because of skewed distributions such as the Gamma distribution and the lognormal distribution, the t-distribution-based robust method may break down. The t-distribution-based method can downweight potential outliers, but cannot be used to approximate systematical skewed distributions. In contrast, the semiparametric Bayesian model is more robust to skewed data and performs stably well.

General conclusions

In sum, we draw the following conclusions. First, the two robust methods outperform the traditional normal-based growth curve modeling when data are nonnormal. When data are normally distributed, the performance of the robust methods is as good as that of the traditional normal-based method. Second, the semiparametric Bayesian method can analyze skewed data and mixture data more accurately and precisely than the robust method using t distributions in growth curve modeling. Third, in modeling data with outliers, the two robust approaches perform equally well. Fourth, the increase of the sample size can often improve the performance of all three estimation methods, but it does not affect their relative performance.

An Illustrative Example

To demonstrate the application of the two robust approaches, we investigate a subset of data from the National Longitudinal Survey of Youth 1997 (NLSY97) Cohort (Bureau of Labor Statistics, U.S. Department of Labor, 2005)[5]. In the study, 512 students' Peabody Individual Achievement Test (PIAT) mathematics scores were collected yearly from the 7th grade to the 10th grade. The trajectory plot for all the individuals (Figure 2) suggests a linear growth pattern for the development of math abilities. The descriptive statistics of the data (Table 5) reveal that the skewness and kurtosis of the data at grades 9 and 10 are significantly different from those of normal distributions. The PIAT math scores at each year are skewed to the left. Thus, it is reasonable to consider the data as nonnormal. As a consequence, we will use this dataset to illustrate the application of the robust methods.

A linear growth curve model is fitted to the data and two robust methods are used for model

---

[5] In addition to the illustrative example in this article, we also provide another example to help substantive researchers understand the application of the proposed robust method in growth curve modeling, using data from the Virginia Cognitive Aging Project. The detailed description of this example as well as the programming code are available on our webpage: https://www.dropbox.com/sh/njhhnm4tdi4ahp9/AADXXBbB79qvp0x37YZSNeK7a?dl=0.

estimation, including the t-distribution-based robust method and the semiparametric Bayesian method. The traditional normal-based method is also applied to serve as a baseline. The parameter estimates are provided in Table 6. Differences among the three methods are shown. Given the conclusions from the simulation study, the results from semiparametric Bayesian methods are more reliable. The average initial mathematical ability at grade 7 is about 6.168 with an average growth rate of 0.314 from grade 7 to grade 10. The credible intervals for estimated σ_L^2 and σ_S^2 do not cover zero, indicating that there are interindividual differences in the initial ability and the rate of change, respectively. Contrary to the traditional growth curve modeling with normal assumptions which fails to detect a correlation between latent intercept and slope, the semiparametric Bayesian method (as well as the t-distribution-based robust method) reports a negative association between initial math abilities and math ability growth rates since the credible interval of the estimate σ_{LS} does not cover zero. Specifically, estimation results suggest that children initially with lower math abilities exhibit higher growth rates in their math abilities from grade 7 to 10. The contradictory results between the traditional normal-based method and the robust methods are likely related to the width of the estimated credible intervals. In the presence of nonnormality, robust methods typically provide more precise parameter estimates, and thus the credible intervals obtained from robust methods are likely to be narrower than the corresponding intervals obtained from the traditional normal-based method, which is exactly the case in this example. Given that the correlation between the latent intercept and slope parameters across sessions are interested in many studies (e.g., Zhang, Davis, Salthouse, and Tucker-Drob, 2007a), the traditional normal-based growth curve modeling is not recommended to use when the nonnormality is suspected. The example here documents evidence favoring the semiparametric Bayesian approach as it detects the existence of an effect more often than the traditional method.

Discussion

In this article, we proposed a semiparametric Bayesian approach for growth curve analysis with nonnormal data. The normal distributions of the intraindividual measurement errors of the

traditional growth curve model were replaced by random distributions with DP mixture priors. In general, being Bayesian allows the model to avoid having to choose a single value for each parameter, and being semiparametric enables the model to increase the dimensionality of the parameters as more data become available. The performance of the semiparametric Bayesian method was systematically evaluated through a simulation study. We further compared the semiparametric Bayesian method to the robust method using Student's t distributions (e.g., Tong and Zhang, 2012) as well as the traditional normal-based method to illustrate the intrinsic characteristics of the three approaches.

From Tong and Zhang (2012) and Tong (2014), it was known that the number of measurement occasions, the potential number of clusters, the covariance between the latent intercept and slope, and the variance of the measurement errors did not affect the performance of the robust method using Student's t distributions and the semiparametric Bayesian method. For simplification, we fixed these four factors and only considered the effect of sample size and distribution of the population in this study. Overall, the traditional normal-based method should not be used when data are nonnormal. In general, the semiparametric Bayesian method outperformed or performed as well as, the robust method using Student's t distributions. The semiparametric Bayesian approach is more robust to the nonnormality, especially when the nonnormality of the data is caused by a nonnormal population. For example, when the measurement errors followed Gamma or lognormal distributions, a random distribution with DP mixture priors captured data better than a t distribution because the t distribution has a parametric form and the shape of it is far away from the shape of a Gamma or a lognormal distribution. Thus, we prefer to use robust semiparametric Bayesian method to analyze such data. If the nonnormality is caused by data contamination such as the existence of outliers, the robust method using t distributions can perform as well as the semiparametric Bayesian method. In practice, it is difficult to distinguish samples with contamination from samples with a nonnormal population distribution. In general, we recommend using semiparametric Bayesian methods when nonnormality is suspected.

Note that we only compared the proposed semiparametric Bayesian method to the robust method using Student's t distributions. In the Bayesian growth curve modeling literature, we are aware of one paper (Zhang, 2013) about modeling measurement errors using different type of distributions including skew-t distributions. Zhang (2013) concluded that one can avoid the loss in the efficiency of standard error estimates when the distribution of the errors is correctly specified. Namely, if data distribution is symmetric and have heavy tails, t distributions should be used. If data distribution is skewed, skew-normal distributions should be used. We can compare the proposed semiparametric Bayesian method to robust methods based on those distributions when skewness of the data is suspected. Overall, our study proposed a semiparametric Bayesian approach and we want to show that it outperforms parametric modeling. The robust growth curve modeling using t distributions is just one representative of parametric robust growth curve analyses. It can be changed to skew-normal distributions which still have parametric forms. Correspondingly, the infinite mixture of normal distributions in the semiparametric Bayesian approach can be changed to the infinite mixture of skew-normal distributions if necessary. Although we did not compare the skew-normal-based robust method to the DP mixture modeling based on a mixture of skew-normal distributions in the simulation study with asymmetric data, we believe the current comparison is still important as to compare parametric and semiparametric Bayesian analyses. In addition, the robust method using Student's t distributions is adopted in our study because it is relatively more broadly used in practice as software has been developed to facilitate the implementation. Therefore, the conclusion from our manuscript has its practical merits.

Each model expresses a data generative process with hidden parameters. Posterior inference is akin to reversing the data generative process to find the hidden structure with which the observed data can most likely be generated. What distinguishes semiparametric Bayesian models from other Bayesian models is that the hidden structure is assumed to grow with the data. Its complexity (e.g., the number of mixture components) is part of the posterior distribution. As illustrated in the simulation study, it is determined as part of the data analysis rather than specified

a priori. Namely, the semiparametric Bayesian method has the advantage of building a Bayesian model on an infinite-dimensional[6] parameter space that can be evaluated on a finite sample in a manner that uses only a finite subset of the available parameters to explain the sample.

The employment of semiparametric Bayesian methods is a field still in its infancy. New Dirichlet process variants and generalizations more suitable to specific applications are being found every year. In our current semiparametric Bayesian growth curve models, only the distribution of intraindividual measurement errors was replaced by an unknown distribution with the DP mixture prior. The similar strategy can be taken for random effects. Based on Tong (2014), if both measurement errors and random effects are allowed to follow unknown random distributions with DP mixture priors, the model convergence rate would be relatively lower. Further study (e.g., using longer Markov chains, picking more informative priors, etc.) needs to be conducted to resolve this issue. We also want to point out that although this study focuses on simple unconditional linear growth curve models, the same methods work for more complicated growth curve models such as linear growth curve models with covariates and nonlinear growth curve models as well.

The DP precision parameter α governs the expected number of clusters. A smaller value of α results in a smaller number of clusters. As suggested by Ishwaran (2000), the prior $\alpha \sim Gamma(2, 2)$ is used in our study to encourage both small and large values. Theoretically, a noninformative prior for α may yield a better estimation of the number of clusters if no prior information is available, but usually causes non-convergence issues or a much longer time for the Markov chain to converge in practice (e.g., Jara, Hanson, Quintana, Müller, and Rosner, 2011; Ohlssen, Sharples, and Spiegelhalter, 2007; West, 1992).

We also want to point out that although normal distributions were replaced by random distributions with DP mixture priors in the robust semiparametric Bayesian method, the distribution used in forming the DP mixture is still normal in this study. However, the base distribution can be nonnormal and chosen in different ways depending on the purpose of analysis.

————

[6] Infinite-dimensional can be interpreted as of finite but unbounded dimension.

If density estimation covers the entire real space, one may use a location-scale kernel such as the normal density or the Student's t density. If the density estimation only covers half of the real space, gamma, log-normal and Weibull mixtures seem to be more appropriate. On the unit interval, mixtures of beta densities can be considered. Special shapes can be produced by special types of mixtures and need to be considered case by case.

For model comparisons, we only compared the parameter estimation. How well the models fit the data is not evaluated. Deviance Information Criterion (DIC) is widely used to evaluate the model fit in Bayesian analysis. Despite the popularity of DIC, it has received much criticism since it was proposed (Spiegelhalter, Best, Carlin, and Linde, 2002). Celeux, Forbers, Robert, and Titterington (2006) argued that the DIC introduced by Spiegelhalter et al. for model assessment and model comparison was directly inspired by linear and generalized linear models, but was open to different possible variations in the setting of models involving random effects, as in our robust growth curve models. Many ways to compute DICs have been proposed in Celeux et al. (2006). However, the calculation of DIC in semiparametric Bayesian analysis has not been studied, but should be considered in the future since DIC is an important index for model comparison.

It is known inflexible models often yield unreasonable inference. For example, complex models may cause overfitting problems. Since semiparametric Bayesian methods provide a way of getting very flexible model and are to fit a single model that can adapt its complexity to the data, it should be an effective guard against overfitting in theory. As stated in the literature (e.g., Niekum, 2015), DP mixture is used to find an appropriate number of mixture components and their associated parameters that best explain the data without overfitting in a fully Bayesian manner. Dunson (2013) pointed out that the concerns about overfitting in semiparametric Bayesian are largely unfounded. In practice, even though there are infinitely many components, only a few of these are allocated and the model behaves like a finite mixture model with sieve behavior in terms of using more components as the sample size increases. "Contrary to the concern about overfitting, the tendency is instead to place a high posterior weight on very few components, potentially underfitting in small sample sizes". As far as we are aware of, no study

has systematically evaluated the overfitting or underfitting problems. More research is needed to

formally address these issues.

References

A. Ansari and R. Iyengar. Semiparametric thurstonian models for recurrent choice: A bayesian analysis. *Psychometrika*, 71:631–657, 2006. doi: 10.1007/s11336-006-1233-5.

C. E. Antoniak. Mixture of dirichlet process with applications to bayesian nonparametric problems. *Annals of Statistics*, 273:1152–1174, 1974. doi: 10.1214/aos/1176342871.

T. Asparouhov and B. Muthén. Structural equation models and mixture models with continous nonnormal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 23:1–19, 2016. doi: 10.1080/10705511.2014.947375.

Adelchi Azzalini and Marc G. Genton. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76:106–129, 2008. doi: 10.1111/j.1751-5823.2007.00016.x.

D. Blackwell and J. B. MacQueen. Ferguson distribution via polya urn schemes. *Annals of Statistics*, 2:353–355, 1973. doi: 10.1214/aos/1176342372.

E. R. Brown and J. G. Ibrahim. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59:221–228, 2003. doi: 10.1111/1541-0420.00028.

Bureau of Labor Statistics, U.S. Department of Labor. *National Longitudinal Survey of Youth 1997 cohort, 1997-2003 (rounds 1-7) [computer file]*. OSU, Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, Ohio, 2005.

D. Burr and H. Doss. A bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association*, 100:242–251, 2005. doi: 10.1198/016214504000001024.

C. A. Bush and S. N. MacEachern. A semiparametric bayesian model for randomised block designs. *Biometrika*, 83:275–285, 1996. doi: 10.1093/biomet/83.2.275.

Meghan K Cain, Zhiyong Zhang, and Ke-Hai Yuan. Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49:1716–1735, 2017. doi: 10.3758/s13428-016-0814-1.

G. Celeux, F. Forbers, C.P. Robert, and D.M. Titterington. Deviance information criteria for missing data models. *Bayesian Analysis*, 1:651–674, 2006. doi: 10.1214/06-BA122.

J. Dahlin, R. Kohn, and T. B. Schon. Bayesian inference for mixed effects models with heterogeneity. Technical Report LiTH-ISY-R-3091, Technical report from Automatic Control at Linköpings universitet, Linköping, Sweden, April 2016.

D. B. Dunson. Bayesian dynamic modeling of latent trait distributions. *Biostatistics*, 7:551–568, 2006. doi: 10.1093/biostatistics/kxj025.

D. B. Dunson. Nonparametric bayes. *https://www.gwern.net/docs/statistics/bayes/2013-dunson.pdf*, 2013.

M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588, 1995. doi: 10.2307/2291069.

Ludwig Fahrmeir and Alexander Raach. A bayesian semiparametric latent variable model for mixed responses. *Psychometrika*, 72:327–346, 2007. doi: 10.1007/s11336-007-9010-7.

T. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1: 209–230, 1973. doi: 10.1214/aos/1176342360.

T. Ferguson. Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2: 615–629, 1974. doi: 10.1214/aos/1176342752.

P. Filzmoser. Identification of multivariate outliers: a performance study. *Austrian Journal of Statistics*, 34:127–138, 2005. doi: 10.17713/ajs.v34i2.406.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984. doi: 10.1109/TPAMI.1984.4767596.

S. J. Gershman and D. M. Blei. A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12, 2012. doi: 10.1016/j.jmp.2011.08.004.

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

S. Ghosal, J. Ghosh, and R. Ramamoorthi. Posterior consistency of dirichlet mixtures in density estimation. *The Annals of Statistics*, 27:143–158, 1999. doi: 10.1214/aos/1018031105.

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

N. L. Hjort. Topics in nonparametric bayesian statistics. In P. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, pages 455–487. Oxford University Press, Oxford, 2003.

N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker. *Bayesian nonparametrics*. Cambridge University Press, Cambridge, 2010.

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

H. Ishwaran. Inference for the random effects in bayesian generalized linear mixed models. In American Statistical Association, editor, *ASA Proceedings of the Bayesian Statistical Science Section*, pages 1–10. 2000.

A. Jara, T. Hanson, F. Quintana, P. Müller, and G. Rosner. Dppackage: Bayesian semi and nonparametric modeling in r. *Journal of Statistical Software*, 40:1–30, 2011. doi: 10.18637/jss.v040.i05.

K. P. Kleinman and J. G. Ibrahim. A semiparametric bayesian approach to the random effects model. *Biometrics*, 54:921–938, 1998. doi: 10.2307/2533846.

Kenneth L. Lange, Roderick J. A. Little, and Jeremy M. G. Taylor. Robust statistical modeling using the t distribution. *Journal of the Americal Statistical Association*, 84:881–896, 1989. doi: 10.2307/2290063.

S. Y. Lee and J. Q. Shi. Joint bayesian analysis of factor scores and structural parameters in the factor analysis model. *Annal of the Institute of Statistical Mathematics*, 52:722–736, 2000. doi: 10.1023/A:1017529427433.

S. Y. Lee and X. Y. Song. Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57:131–150, 2004. doi: 10.1348/000711004849204.

S. Y. Lee and Y. M. Xia. A robust bayesian approach for structural equation models with missing data. *Psychometrika*, 73:343–364, 2008. doi: 10.1007/s11336-008-9060-5.

S. Y. Lee, B. Lu, and X. Y. Song. Semiparametric bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*, 27:2341–2360, 2008. doi: 10.1002/sim.3098.

Z. Lu and Z. Zhang. Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application. computational statistics and data analysis. *Computational Statistics and Data Analysis*, 71:220–240, 2014. doi: 10.1016/j.csda.2013.07.036.

D. Lunn, C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter. *The BUGS book: A practical introduction to Bayesian analysis*. CRC Press, Boca Raton, FL, 2013.

S. MacEachern. Dependent nonparametric processes. In American Statistical Association, editor, *ASA Proceedings of the Section on Bayesian Statistical Science*. 1999.

Ricardo A. Maronna, R. Douglas Martin, and Victor J. Yohai. *Robust statistics: Theory and methods*. John Wiley & Sons, Inc, New York, 2006.

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

W. Meredith and J. Tisak. Latent curve analysis. *Psychometrika*, 55:107–122, 1990. doi: 10.1007/BF02294746.

Theodore Micceri. The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105:156–166, 1989. doi: 10.1037/0033-2909.105.1.156.

P. Müller and R. Mitra. Bayesian nonparametric inference - why and how. *Bayesian Analysis*, 1: 1–33, 2004. doi: 10.1214/13-BA811.

Bengt Muthén and Kerby Shedden. Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55:463–469, 1999. doi: 10.1111/j.0006-341x.1999.00463.x.

R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2000. doi: 10.2307/1390653.

Scott Niekum. A brief introduction to bayesian nonparametric methods for clustering and time series analysis. Technical Report CMU-RI-TR-15-02, Carnegie Mellon University, Pittsburgh, PA, January 2015.

K. L. Nylund, T. Asparoubov, and B. O. Muthén. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling*, 14:535–569, 2007. doi: 10.1080/10705510701575396.

D. I. Ohlssen, L. D. Sharples, and D. Spiegelhalter. Flexible random-effects models using bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*, 26:2088–112, 2007. doi: 10.1002/sim.2666.

D. Peña and F. J. Prieto. Multivariate outlier detection and robust covariance matrix estimation (with discussion). *Technometrics*, 43:286–310, 2001. doi: 10.1198/004017001316975899.

Jose C. Pinheiro, Chuanhai Liu, and Ying Nian Wu. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10:249–276, 2001. doi: 10.1198/10618600152628059.

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2004.

V. Savalei and C. Falk. Robust two-stage approach outperforms robust fiml with incomplete nonnormal data. *Structural Equation Modeling*, 21:280–302, 2014. doi: 10.1080/10705511.2014.882692.

S. Serang, Z. Zhang, Helm J., J. S. Steele, and K. J. Grimm. Evaluation of a bayesian approach to estimating nonlinear mixed-effects mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22:202–215, 2015. doi: 10.1080/10705511.2014.937322.

J. Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

Y. Si and J. P. Reiter. Nonparametric bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *Journal of Educational and Behavioral Statistics*, 38:499–521, 2013. doi: 10.3102/1076998613480394.

Y. Si, J. P. Reiter, and D. S. Hillygus. Semi-parametric selection models for potentially non-ignorable attrition in panel studies with refreshment samples. *Political Analysis*, 23: 92–112, 2015. doi: 10.1093/pan/mpu009.

X.-Y. Song and S.-Y. Lee. A tutorial on the bayesian approach for analyzing structural equation models. *Journal of Mathemetical Psychology*, 56:135–148, 2012. doi: 10.1016/j.jmp.2012.02.001.

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

X. Tong. Robust semiparametric bayesian methods in growth curve modeling. *(Unpublished doctoral dissertation). Notre Dame, IN: University of Notre Dame*, 2014.

X. Tong and Z. Zhang. Diagnostics of robust growth curve modeling using student's t distribution. *Multivariate Behavioral Research*, 47:493–518, 2012. doi: 10.1080/00273171.2012.692614.

X. Tong and Z. Zhang. Outlying observation diagnostics in growth curve modeling. *Multivariate Behavioral Research*, 52:768–788, 2017. doi: 10.1080/00273171.2017.1374824.

M. West. *Hyperparameter estimation in Dirichlet process mixture models*. Technical Report 92-A03, Duke University, ISDS, 1992.

M. Yang and D. B. Dunson. Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, 75:675–693, 2010. doi: 10.1007/s11336-010-9174-4.

K.-H. Yuan and P. M. Bentler. Structural equation modeling with robust covariances. *Sociological Methodology*, 28:363–396, 1998. doi: 10.1111/0081-1750.00052.

K.-H. Yuan and P. M. Bentler. Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 54:161–175, 2001. doi: 10.1348/000711001159366.

K.-H. Yuan and Z. Zhang. Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, 77:803–826, 2012. doi: 10.1007/s11336-012-9282-4.

K.-H. Yuan and X. Zhong. Outliers, high-leverage observations and influential cases in factor analysis: Minimizing their effect using robust procedures. *Sociological Methodology*, 38: 329–368, 2008. doi: 10.1111/j.1467-9531.2008.00198.x.

K.-H. Yuan, P. M. Bentler, and W. Chan. Structural equation modeling with heavy tailed distributions. *Psychometrika*, 69:421—436, 2004.

Z. Zhang. Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics*, 40:1779–1795, 2013. doi: 10.1080/02664763.2013.796348.

Z. Zhang, H. P. Davis, T. A. Salthouse, and E. A. Tucker-Drob. Correlates of individual, and age-related, differences in short-term learning. *Learning and Individual Differences*, 17: 231–240, 2007a. doi: 10.1016/j.lindif.2007.01.004.

Z. Zhang, Keke Lai, Zhenqiu Lu, and Xin Tong. Bayesian inference and application of robust growth curve models using student's t distribution. *Structural equation modeling*, 20:47–78, 2013. doi: 10.1080/10705511.2013.742382.

Zhiyong Zhang, Fumiaki Hamagami, Lijuan Wang, Kevin J. Grimm, and John R. Nesselroade. Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31:374–383, 2007b. doi: 10.1177/0165025407077764.

Table 1

Different percentiles (5%, 50%, 95%) of the distribution of the number of clusters K, given different values of precision parameter α and sample size N

	$\alpha = 0.1$			$\alpha = 1$			$\alpha = 2$		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
$N = 200$	1	1	3	3	7	11	7	13	19
$N = 500$	1	1	3	4	8	12	9	14	21
$N = 1000$	1	2	3	4	8	13	10	16	23

Table 2

Seven distributional conditions used for the Monte Carlo simulation

Condition	Distribution of $e_1 - e_T$	Kurtosis of $e_1 - e_T$	Skewness of $e_1 - e_T$
C1	$N(0, \sigma_e^2)$	0	0
C2	$t_{(5)}(0, \sigma_e^2)$	6	0
C3	$\Gamma_{(1,1)}(0, \sigma_e^2)$	6	2
C4	$LN_{(0,1)}(0, \sigma_e^2)$	110.936	6.185
C5	Mixture of $N(0, \sigma_e^2)$ and $N(0, 5\sigma_e^2)$	-	-
C6	$N(0, \sigma_e^2)$ with 10 outliers on one side	-	-
C7	$N(0, \sigma_e^2)$ with 10 outliers on both sides	-	-

Note. $t_{(df)}(\mu, \sigma^2)$, $\Gamma_{(\alpha,\beta)}(\mu, \sigma^2)$, and $LN_{(\mu_N, \sigma_N^2)}(\mu, \sigma^2)$ represent the rescaled t distribution, the rescaled Gamma distribution, and the rescaled lognormal distribution, respectively, with mean μ and variance σ^2.

Table 3

Parameter estimation for condition C1 ($e_1, \ldots, e_T \sim N(0, \sigma_e^2)$) when $N = 200$

	Normal-based						Robust t						Semiparametric					
	Est.	Bias	RB (%)	SE	MSE	CP	Est.	Bias	RB (%)	SE	MSE	CP	Est.	Bias	RB (%)	SE	MSE	CP
β_L	6.202	0.002	0.030	0.083	0.007	0.942	6.202	0.002	0.040	0.083	0.007	0.944	6.208	0.008	0.123	0.084	0.007	0.935
β_S	0.299	-0.001	-0.448	0.031	0.001	0.962	0.299	-0.001	-0.360	0.031	0.001	0.960	0.292	-0.008	-2.603	0.025	0.001	1.000
σ_L^2	1.015	0.015	1.520	0.134	0.018	0.954	1.018	0.018	1.815	0.134	0.018	0.960	1.017	0.017	1.662	0.129	0.017	0.978
σ_S^2	0.117	0.017	16.710	0.018	0.001	0.940	0.117	0.017	17.080	0.018	0.001	0.940	0.118	0.018	17.581	0.020	0.001	0.935
σ_{LS}	-0.010	-0.010	-0.979	0.036	0.001	0.972	-0.011	-0.011	-1.075	0.036	0.001	0.970	-0.011	-0.011	-1.132	0.039	0.002	0.957
σ_e^2	0.499	-0.001	-0.181	0.033	0.001	0.946	0.498	-0.002	-0.397	0.033	0.001	0.828	0.495	-0.005	-1.018	0.031	0.001	0.848
df_e	-	-	-	-	-	-	55.661	-	-	6.730	-	-	-	-	-	-	-	-
K_e	-	-	-	-	-	-	-	-	-	-	-	-	5.150	-	-	0.785	-	-
α	-	-	-	-	-	-	-	-	-	-	-	-	0.928	-	-	0.127	-	-

Note. Est.: estimate; RB: relative bias; SE: standard error; MSE: mean squared error; CP: coverage probability.

Table 4

Mean squared errors and coverage probabilities for different data conditions

| | | $N = 50$ | | $N = 200$ | | $N = 500$ | |
		Robust t	Semiparametric	Robust t	Semiparametric	Robust t	Semiparametric
C1	MSE	0.019	0.023	0.005	0.005	0.002	0.002
	CP	0.927	0.923	0.934	0.942	0.931	0.944
C2	MSE	0.021	0.021	0.005	0.005	0.002	0.002
	CP	0.907	0.924	0.899	0.908	0.900	0.913
C3	MSE	0.020	0.019	0.006	0.005	0.003	0.002
	CP	0.892	0.912	0.861	0.902	0.825	0.905
C4	MSE	0.037	0.039	0.010	0.008	0.006	0.004
	CP	0.819	0.845	0.785	0.856	0.699	0.828
C5	MSE	0.053	0.043	0.015	0.012	0.007	0.005
	CP	0.883	0.904	0.879	0.903	0.884	0.907
C6	MSE	0.180	0.158	0.006	0.011	0.003	0.003
	CP	0.859	0.828	0.961	0.907	0.950	0.936
C7	MSE	0.037	0.043	0.006	0.007	0.003	0.002
	CP	0.971	0.977	0.962	0.963	0.957	0.967

Table 5

Descriptive statistics of the PIAT math data from NLSY97

Grade	Mean	s.d.	Skewness	Kurtosis
7	6.071	1.312	-0.110	3.392
8	6.590	1.392	-0.168	3.336
9	6.796	1.419	-0.564*	4.814*
10	7.044	1.325	-0.344*	3.708*

Note. s.d. = standard deviation. The "*" sign indicates that the corresponding statistic is signifi-
cantly different from that of a normal distribution. The significance of skewness is tested through
the D'Agostino test, and the significance of kurtosis is tested through the Anscombe-Glynn test.

Table 6

Parameter estimates from the traditional method as well as the two robust approaches

Parameters	Normal-based			Robust t			Semiparametric		
	Est.	CI.L	CI.U	Est.	CI.L	CI.U	Est.	CI.L	CI.U
β_L	6.157	6.047	6.269	6.153	6.043	6.263	6.168	6.058	6.276
β_S	0.312	0.275	0.349	0.319	0.283	0.352	0.314	0.280	0.349
σ_L^2	1.125	0.937	1.337	1.206	1.019	1.414	1.153	0.974	1.357
σ_S^2	0.035	0.024	0.049	0.042	0.029	0.057	0.040	0.028	0.055
σ_{LS}	-0.034	-0.081	0.009	-0.050	-0.096	-0.008	-0.049	-0.093	-0.009
σ_e^2	0.748	0.694	0.806	0.729	0.698	0.762	0.737	0.698	0.781
df_e	-	-	-	4.143	3.222	5.277	-	-	-
# of clusters	-	-	-	-	-	-	11.060	5.000	18.000

Note. Est = estimate; CI.L = lower limit of the 95% credible interval; CI.U = upper limit of the
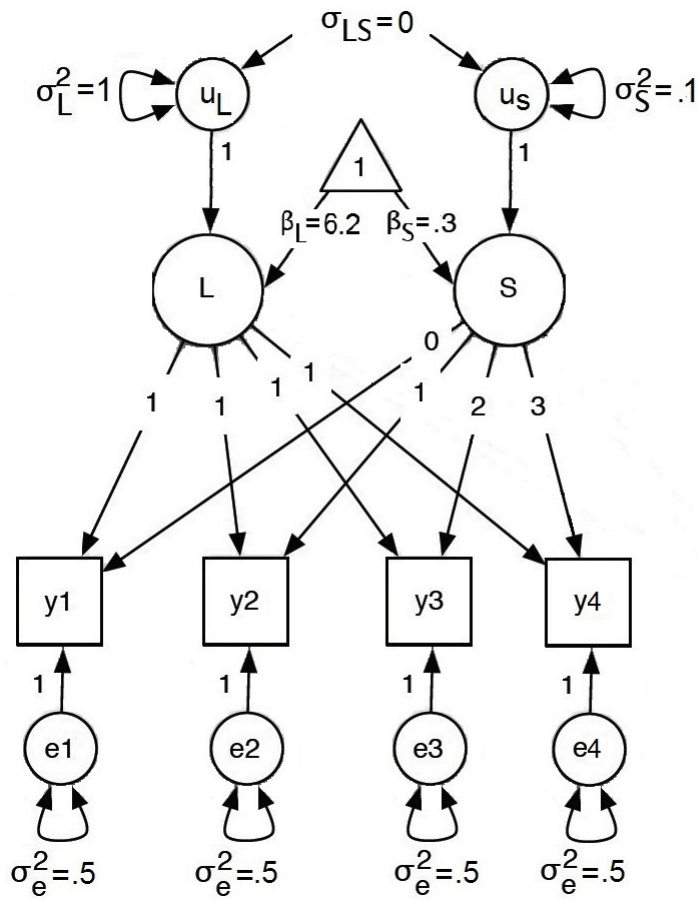
95% credible interval.

Figure 1. Path diagram of a linear growth curve model. The numbers in the path diagram are population parameter values used in the simulation.
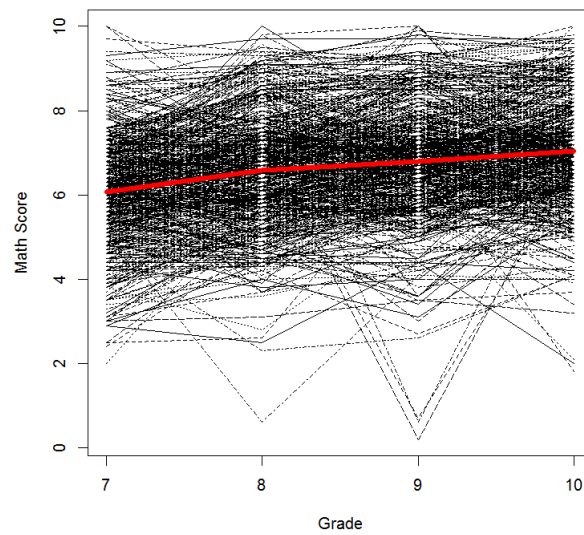
Figure 2. A collection of individual trajectories for the PIAT math data from NLSY97. A total of 512 students are measured at 4 occasions. The red line represents the mean growth trajectory.