Bayesian inference and application of robust growth curve models using Student's t distribution

Zhiyong Zhang, Keke Lai, Zhenqiu Lu, and Xin Tong
University of Notre Dame

Abstract

Despite the widespread popularity of growth curve analysis, few studies have investigated robust growth curve models. In this paper, the t distribution is applied to model heavy-tailed data and contaminated normal data with outliers for growth curve analysis. The derived robust growth curve models are estimated through Bayesian methods utilizing data augmentation and Gibbs sampling algorithms. The analysis of mathematical development data shows that the robust latent basis growth curve model better describes the mathematical growth trajectory than the corresponding normal growth curve model and can reveal the individual differences in mathematical development. Simulation studies further confirm that the robust growth curve models significantly outperform the normal growth curve models for both heavy-tailed t data and normal data with outliers but lose only slight efficiency for normal data. It appears convincing to replace the normal distribution with the t distribution for growth curve analysis. Three information criteria are evaluated for model selection. Online software is also provided for conducting robust analysis discussed in this study.
Keywords: Bayesian inference, robust growth curve models, t distribution, model comparison, mathematical development

Introduction

The need for analysis of change and individual differences in change has advanced the development and application of growth curve models in behavioral and social sciences (e.g., McArdle, 1988; Meredith & Tisak, 1990). Statistical inference for growth curve modeling is typically based on the normal distribution (univariate or multivariate) that is unfortunately vulnerable to longer-than-normal tails or outliers (e.g., Pan & Fang, 2002). Failing to account for the longer-than-normal tails or outliers in data may result in unreliable parameter estimates, incorrect standard errors and confidence intervals, and misleading statistical tests and inference (Maronna et al., 2006; Yuan et al., 2004; Zu & Yuan, 2010). Some convenient (or conventional) methods, for example, simply deleting outliers, often lead to other problems such as under-estimated standard errors and reduced efficiency (e.g., Lange et al., 1989; Yuan & Bentler, 2002).

Robust approaches have been developed to produce reliable parameter estimates, and associated tests and confidence intervals when the statistical assumptions on data distributions are violated. The majority of robust approaches has occurred in the past half century following the fundamental work by Tukey (Tukey, 1962), Huber (Huber, 1981), and Hampel (Hampel et al., 1986). The main theme of robust methods is to downweight the influence of observations that are far away from the majority of the data through carefully chosen weight functions (e.g., Zhong & Yuan, 2010).

Despite the widespread popularity of growth curve analysis, studies on robust growth curve models are still very rare. We are only aware of three studies that directly employed robust methods in growth curve models. Pendergast & Broffitt (1985) discussed two robust estimators and their asymptotic properties for growth curve models. Singer & Sen (1986) proposed an M-method to obtain the parameter estimates for growth curve models by transforming them into standard multivariate liner models. Silvapulle (1992) further extended the M-method to allow asymmetric measurement errors for growth curve analysis.

Considering that growth curve analysis can also be conducted in the framework of mixed-effects modeling, latent variable modeling, and multilevel modeling, it is worth mentioning the robust literature in these areas. For example, Gill (2000) and Yau & Kuk (2002) developed robust methods for linear and generalized linear mixed-effects models (see also, Lachos et al., 2009; Pinheiro et al., 2001; Song et al., 2007). Lee & Xia (2006, 2008); Yuan & Bentler (1998) and Yuan et al. (2000) developed robust methods for linear and nonlinear structural equation models with missing data. Yeap & Davidian (2001) developed robust methods for hierarchical nonlinear models (see also, Rachman-Moore & Wolfe, 1984).

However, the above robust methods developed for both growth curve models and other related models have not been widely adopted in growth curve analysis of longitudinal data. We believe that there are at least three reasons for the lack of enthusiasm in the use of robust methods for growth curve analysis. First, the M-methods used in robust growth curve models proposed in Pendergast & Broffitt (1985), Silvapulle (1992), and Singer & Sen (1986) are nonparametric and not widely used. Second, The parametric robust methods developed for mix-effects models and latent variable models based on t distributions or contaminated normal distributions have hardly been applied to growth curve analysis directly (e.g., Pinheiro et al., 2001; Song et al., 2007). Third, there still lacks easy-to-use software to carry out robust growth curve analysis. To advance the application of robust growth curve analysis, we suggest the use of a parametric robust method based on Student's t distribution and provide software to carry out the analysis.

Robust methods based on Student's t distribution

If a vector $\mathbf{y}$ with p variables follows a multivariate t distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom k, denoted by $MT_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, k)$, its density function can be written as

$$p(\mathbf{y}) = \frac{\Gamma[(k+p)/2]}{\Gamma(k/2)k^{p/2}\pi^{p/2}|\boldsymbol{\Sigma}|^{1/2}}[1 + \frac{(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{k}]^{-\frac{k+p}{2}}. \tag{1}$$

The t distribution has longer tails than the normal distribution and is more robust to outliers than the normal distribution (e.g., Lange et al., 1989). As the degrees of freedom k goes to infinity, the t distribution approaches the normal distribution. For a univariate t distribution, its kurtosis, if existing, is larger than 3, the kurtosis of the normal distribution.

To demonstrate why t distributions are useful for robust analysis, consider a univariate t variable $y_i, i = 1, \ldots, n$ with unknown mean μ, known scale $\Sigma = 1$ and known degrees of freedom k. To estimate μ through maximum likelihood method, we solve the equation with the estimate $\hat{\mu}$,

$$\sum_{i=1}^{n} \frac{y_i - \hat{\mu}}{k + (y_i - \hat{\mu})^2} = \sum_{i=1}^{n} \psi(y_i) = 0. \tag{2}$$

Note that if $(y_i - \hat{\mu})$ is large, indicating outlying observations, the contribution of y_i to the estimation of μ is small. Especially, when $(y_i - \hat{\mu})$ goes to infinity, $\psi(y_i) = 0$. Thus, using t distributions, we can downweight outliers or data on the tails and achieve robust results.

The wide adoption of t distributions in robust statistical modeling has been advanced by Lange et al. (1989) although their earlier applications can be seen in the literature (e.g., Hampel et al., 1986; Little, 1988; Sutradhar & Ali, 1986). Recently, t distributions have been applied in more complex robust data analysis.

For example, Lee & Xia (2006, 2008) applied t distributions in robust structural equation models. Pinheiro et al. (2001) discussed the use of t distributions for linear mixed-effects models (see also, Song et al., 2007). Wang et al. (2004) applied t distributions for robust mixture modeling (see also Shoham, 2002). Although maximum likelihood methods are most widely used, Bayesian methods have also been applied to parameter estimation in robust models (e.g., Lange et al., 1989; Lee & Xia, 2008; Liu, 1996; Song et al., 2007).

There are many advantages in using t distributions for robust data analysis (e.g., Lange et al., 1989). First, t distributions have parametric forms and statistical inference can be carried out relatively easily through both maximum likelihood methods and Bayesian methods. Second, computation based on t distributions is often relatively easier than using other robust techniques. Third, the degrees of freedom of t distributions controls the weight of outliers and can be either set at an *a priori* value or estimated. Fourth, robust methods based on t distributions are easy to understand and can intuitively be considered as natural extensions of the corresponding normal distribution based methods for heavy-tailed data.

In this study, we form a set of robust growth curve models based on t distributions. In the following, we first present a general form of robust growth curve models and illustrate it through several specific robust growth curve models. Then, we discuss how to estimate robust growth curve models through Bayesian methods. After that, we demonstrate the application of robust growth curve models through the analysis of a set of mathematical development data from the National Longitudinal Survey of Youth. The performance of robust growth curve models is then evaluated through carefully designed simulation studies. Simulation studies are also used to evaluate the performance of several information criteria in model selection. Finally, guidelines on the application of robust growth curve models and software implementation are discussed.

Robust Growth Curve Models and Bayesian Estimation

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 general form of growth curve models can be expressed as

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{e}_i \tag{3}$$

where $\boldsymbol{\Lambda}$ is a $T \times q$ factor loading matrix determining the growth trajectory, $\boldsymbol{\eta}_i$ is a $q \times 1$ random vector, and $\mathbf{e}_i$ is a vector of residuals or measurement errors. $\boldsymbol{\eta}_i$ are often called random effects because they are different for each individual. The means of $\boldsymbol{\eta}_i$ are fixed effects so that

$$\boldsymbol{\eta}_i = \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \tag{4}$$

where $\boldsymbol{\epsilon}_i$ follows a q-variate normal distribution as $\boldsymbol{\epsilon}_i \sim MN_q(\mathbf{0}, \boldsymbol{\Psi})$. $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed-effects parameters. We use MN to denote a multivariate normal distribution and the subscript q to denote its dimension. For traditional growth curve models, $\mathbf{e}_i$ is assumed to be normally distributed as $\mathbf{e}_i \sim MN_T(0, \boldsymbol{\Phi})$. We refer to the models based on normal errors as normal growth curve models. In order to deal with long-tailness or outliers in the data, we model the error term with the multivariate t distribution so that

$$\mathbf{e}_i \sim MT_T(0, \boldsymbol{\Phi}, k). \tag{5}$$

We then refer to the resulting models as the robust growth curve models. As for normal growth curve models, the error structure can be simplified to $\boldsymbol{\Phi} = \mathbf{I}_{T\times T}\sigma_e^2$ where σ_e^2 is an unknown scale parameter. With the increase of k, the multivariate t distribution approaches a multivariate normal distribution and therefore the robust growth curve models become the normal growth curve models.

Special robust growth curve models can be derived from the above general form. For example, if

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1/(T-1) \\ \vdots & \vdots \\ 1 & (j-1)/(T-1) \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}, \boldsymbol{\eta}_i = \begin{bmatrix} L_i \\ S_i \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_L \\ \beta_S \end{bmatrix}, \mathbf{\Psi} = \begin{bmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{bmatrix}, \tag{6}$$

the model represents a robust linear growth curve model with random initial level (intercept) L_i and random slope (rate of change) S_i. β_L and β_S are the average intercept and slope across all individuals, respectively. σ_L^2 is the variability around the mean intercept and represents inter-individual differences in the latent initial level and σ_S^2 represents the variability, or individual differences around the mean slope. σ_{LS} represents the covariance between the intercept and slope.

If in the robust linear growth curve model the factor loading matrix is replaced by

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & A[2] \\ \vdots & \vdots \\ 1 & A[j] \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \tag{7}$$

where $A[j], j = 2, \ldots, T-1$, are parameters to be estimated freely, the resulting model is the robust latent basis growth curve model. The robust linear growth curve model can be applied if the growth trajectory is linear. The robust latent basis growth curve model, on the other hand, can fit the nonlinear growth trajectories. As a special case, when $A[j] = (j-1)/(T-1)$, the robust latent basis growth curve model reduces to the robust linear growth curve model. Note that in the robust latent basis growth curve model, β_S is the total change from the first measurement occasion to the last one and σ_S^2 characterizes the individual differences in the total change.

A trivial case is when

$$\mathbf{\Lambda} = [1, 1, \ldots, 1]', \boldsymbol{\eta}_i = L_i, \boldsymbol{\beta} = \beta_L, \mathbf{\Psi} = \sigma_L^2. \tag{8}$$

This is a robust growth curve model with intercept only and can be referred to as robust no growth model following from the normal no growth curve model. The robust no growth model is useful for the analysis of repeated measures data with individual differences in levels but without apparent individual change.

To estimate the robust growth curve models, we use Bayesian methods. The use of Bayesian methods for complex data analysis has been made popular by Lee and colleagues (e.g., Lee, 2007; Lee & Shi, 2000; Lee & Song, 2008; Lee & Xia, 2008). Zhang et al. (2007) also applied Bayesian methods to growth curve modeling. The basic idea of Bayesian methods is to obtain the posterior distributions of model parameters from the likelihood function and the prior distributions. Because the multivariate t distribution can be viewed as a multivariate normal distribution with variance weighted by a Gamma distribution, the data augmentation method is used here to simplify the posterior distribution. Specifically, a Gamma random variable w is augmented with a multivariate normal random variable because if $w_i \sim G(\frac{k}{2}, \frac{k}{2})$ and $\mathbf{y}_i | w_i, \boldsymbol{\eta}_i \sim MN_T(\mathbf{\Lambda}\boldsymbol{\eta}_i, \mathbf{\Phi}/w_i)$, then $\mathbf{y}_i \sim MT_T(\mathbf{\Lambda}\boldsymbol{\eta}_i, \mathbf{\Phi}, k)$, where G denotes the Gamma density function (Press, 1972). The joint distribution of $\mathbf{y}_i$, $\boldsymbol{\eta}_i$, and w_i is

$$\begin{aligned} p(\mathbf{y}_i, \boldsymbol{\eta}_i, w_i | k, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \beta) &= p(\boldsymbol{\eta}_i | \boldsymbol{\Psi}, \beta) p(\mathbf{y}_i | \boldsymbol{\eta}_i, w_i, \boldsymbol{\Phi}, \boldsymbol{\Lambda}) p(w_i | k) \\ &= (2\pi)^{-q/2} |\boldsymbol{\Psi}|^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{\eta}_i - \boldsymbol{\beta})' \boldsymbol{\Psi}^{-1} (\boldsymbol{\eta}_i - \boldsymbol{\beta}) \right] \\ &\quad \times (2\pi)^{-T/2} \left| \frac{\boldsymbol{\Phi}}{w_i} \right|^{-1/2} \exp\left[-\frac{w_i}{2} (\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)' \boldsymbol{\Phi}^{-1} (\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i) \right] \\ &\quad \times w_i^{k/2-1} \exp(-\frac{w_i}{k/2}) \frac{1}{\Gamma(k/2)(k/2)^{k/2}}. \end{aligned} \tag{9}$$

Thus, the likelihood function for the robust growth curve model is

$$\begin{aligned} L &= \prod_{i=1}^{N} p(\mathbf{y}_i, \boldsymbol{\eta}_i, w_i | k, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\beta}) \\ &\propto |\boldsymbol{\Psi}|^{-N/2} \exp\left[-\frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{\eta}_i - \boldsymbol{\beta})' \boldsymbol{\Psi}^{-1} (\boldsymbol{\eta}_i - \boldsymbol{\beta}) \right] \\ &\quad \times \prod_{i=1}^{N} w_i^{T/2} |\boldsymbol{\Phi}|^{-N/2} \exp\left[-\sum_{i=1}^{N} \frac{w_i}{2} (\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)' \boldsymbol{\Phi}^{-1} (\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i) \right]. \\ &\quad \times (\prod_{i=1}^{N} w_i)^{k/2-1} \exp\left(-\frac{\sum_{i=1}^{N} w_i}{k/2} \right) \left[\Gamma(k/2)^{-N} \right] (k/2)^{-Nk/2} \end{aligned} \tag{10}$$

The unknown parameters in the robust growth curve models include $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}$, k, and the possible unknown elements in $\boldsymbol{\Lambda}$ such as $A[j], j = 2, \ldots, T-1$ in the robust latent basis model. Let $p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k)$ denote the joint prior distribution of these parameters. The joint posterior distribution of model parameters is

$$p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k | \mathbf{y}_i) \propto \int \int p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k) \times L d\boldsymbol{\eta} dw. \tag{11}$$

Generally speaking, the integral above is difficult to evaluate. To circumvent the difficulty, Markov chain Monte Carlo (MCMC) methods can be applied (for a comprehensive discussion of MCMC methods, see Robert & Casella, 2004). The conditional posterior distribution $p_{1|2}$ for the model parameters,

$$p_{1|2} = p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k | \mathbf{y}_i, \boldsymbol{\eta}_i, w_i) \propto p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k) \times L, \tag{12}$$

and the conditional posterior distribution $p_{2|1}$ for the auxiliary variables including latent variables,

$$p_{2|1} = p(\boldsymbol{\eta}_i, w_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k) \propto L \tag{13}$$

are often relatively easy to obtain. By iteratively drawing samples from $p_{1|2}$ and $p_{2|1}$, we can obtain the empirical marginal distributions of the model parameters and form parameter estimates through their posterior means.

One way to construct the joint prior distribution is to use the independent priors for each (set of) parameter. In particular, for our model,

$$p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}, k) = p(\boldsymbol{\beta}) p(\boldsymbol{\Lambda}) p(\boldsymbol{\Psi}) p(\boldsymbol{\Phi}) p(k). \tag{14}$$

For robust growth curve models, the individual conditional posterior distribution for model parameters and auxiliary variables can often be obtained that include $p(\boldsymbol{\beta}|\cdot)$, $p(\boldsymbol{\Lambda}|\cdot)$, $p(\boldsymbol{\Psi}|\cdot)$, $p(\boldsymbol{\Phi}|\cdot)$, $p(k|\cdot)$, $p(\boldsymbol{\eta}_i|\cdot)$, and

$p(w_i|\cdot)$, where $\cdot$ represents conditional arguments. With the resulting conditional posterior distributions, Gibbs sampling can be used to generate Markov chains for model parameters and latent variables and construct parameter estimates based on posterior means (e.g., Robert & Casella, 2004). In the appendix, we give the prior and posterior distributions for the robust growth curve models. The Gibbs sampling algorithm is also outlined.

Model comparisons

For the analysis of empirical data, deviance information criterion (DIC, Spiegelhalter et al., 2002) can be used to compare and select models, for example, among robust no growth models, robust linear growth models and robust latent basis growth models. DIC is a widely used criterion for model selection in the Bayesian framework. DIC is defined as a Bayesian measure of the goodness of model fit with a penalty of model complexity,

$$\text{DIC} = \overline{D(\boldsymbol{\Theta})} + p_D = D(\overline{\boldsymbol{\Theta}}) + 2p_D, \tag{15}$$

where $\boldsymbol{\Theta}$ represents a vector of all the unknown parameters in the model, $\overline{D(\boldsymbol{\Theta})}$ is the posterior mean of -2(LogLikehood function) and $D(\overline{\boldsymbol{\Theta}})$ is the value of -2(LogLikehood function) calculated at the posterior mean of $\boldsymbol{\Theta}$. The complexity measure, p_D, is defined as the difference between the posterior mean of deviance ($\overline{D(\boldsymbol{\Theta})}$) and the deviance evaluated at the posterior mean of the parameters ($D(\overline{\boldsymbol{\Theta}})$). In other words,

$$p_D = \overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}}). \tag{16}$$

The model with the minimum DIC will make the best short-term predictions and thus indicates the best model among evaluated ones (Spiegelhalter et al., 2002).

In addition to DIC, we also used two other model comparison criteria - the extended BIC (EBIC) and the extended AIC (EAIC) (see discussion by Brooks in Spiegelhalter et al., 2002). The EBIC is calculated by

$$\text{EBIC} = D(\overline{\boldsymbol{\Theta}}) + \log N \times p \tag{17}$$

and the EAIC is calculated by

$$\text{EAIC} = D(\overline{\boldsymbol{\Theta}}) + 2 \times p \tag{18}$$

where N is the sample size and p is the number of unknown random-effects and fixed-effects parameters in the robust models. For both EBIC and EAIC, a smaller value indicates a better fit.

An Example on Mathematical Development Analysis

To demonstrate the application of the robust growth curve models using t distributions, we analyze a set of data from the National Longitudinal Survey of Youth 1997 Cohort. A sample of $N = 310$ school children were administered the Peabody Individual Achievement Test (PIAT) mathematics subtest yearly from grade 7 to grade 11. The boxplots for the PIAT math at all grades are given in Figure 1. From the boxplots, there seems to be outliers at each grade. Note that although the data appear to skew to the left from the boxplots, the skewness from all grades is not statistically different from 0 with the maximum negative skewness 0.127 at the 10th grade when outliers identified by the boxplots are removed. The kurtoses range from 3.13 to 4.30 for the observed data from grade 7 to grade 11, and the Anscombe-Glynn test (Anscombe & Glynn, 1983) shows that the kurtoses at grades 8, 9, and 11 are significantly larger than 3, the kurtosis of the normal distribution. Both the boxplots and the large kurtoses suggest that robust growth curve models based on t distribution could be more appropriate than normal growth curve models.

To select an appropriate model to analyze mathematical development, we fitted three robust growth curve models, including the robust no growth model, the robust linear growth curve model, and the robust latent basis growth curve model, to the data. The following priors are specified in these models: $p(\boldsymbol{\beta}) =$

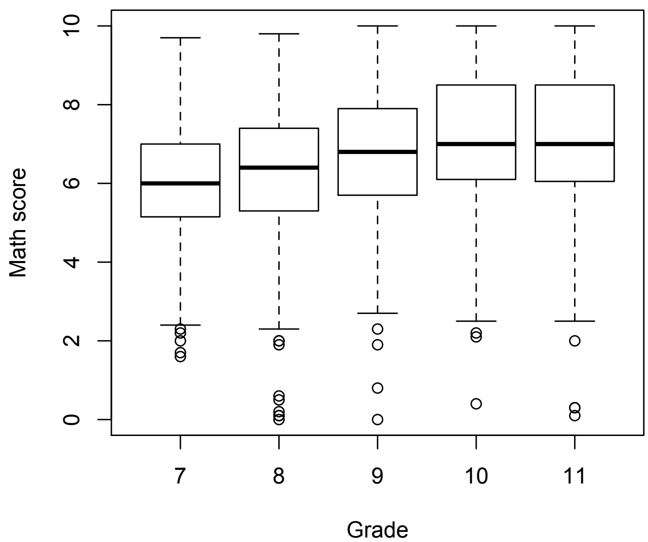

Figure 1 Boxplots of Peabody Individual Achievement Test (PIAT) math data

Table 1 Anscombe-Glynn test of kurtosis of math data at each grade

Grade	Mean	Standard deviation	Kurtosis	Anscombe-Glynn test (p-value)
7	6.08	1.53	3.13	.257
8	6.28	1.80	4.21	.001
9	6.77	1.70	3.66	.021
10	7.06	1.73	3.48	.055
11	7.05	1.77	4.30	.001

$MN(\mathbf{0}_{d\times 1}, \mathbf{I}_{d\times d} \times 10^6)$, $p(\mathbf{\Psi}) = IW(d, \mathbf{I}_{d\times d})$, $\mathbf{\Phi} = \mathbf{I}_{T\times T}\sigma_e^2$ and $p(\sigma_e^2) = IG(.001, .001)$, and $p(k) = U(0, 100)$, where MN, IG, IW, and U denote the density functions for multivariate normal distribution, inverse Gamma distribution, inverse Wishart distribution, and uniform distribution, respectively. For the intercept only model, $d = 1$; and for the robust linear and latent basis growth curve models, $d = 2$. Furthermore, for the robust latent basis model, the prior for the unknown factor loadings is set as $A[2 : T-1]' = MN(\mathbf{0}_{(T-2)\times 1}, \mathbf{I}_{(T-2)\times(T-2)} \times 10^6)$. Those priors are usually considered as uninformative priors in the Bayesian literature (Congdon, 2003).

For each fitted model, the convergence of the Markov chain for each parameter was diagnosed both visually by inspecting the history plot of each parameter and statistically based on the Geweke test (Geweke, 1992). We first visually checked the history plot of each parameter to identify the burn-in period. For example, the history plots of the robust latent basis growth curve model parameters are presented in Figure 2. The history plots indicate that the Markov chains for all parameters seemed to converge after a few hundred of iterations. We then threw away the first 1000 iterations and further tested the convergence of each Markov chain based on the Geweke statistics. The Geweke statistics for the parameters in Figure 2 are summarized in Table 3. All Geweke statistics are smaller than 2 and suggest the convergence of the Markov chains. After convergence, in order to control Monte Carlo error, we generated the Markov chains long enough so that for all parameters, the Monte Carlo errors are smaller than 5% of their corresponding standard deviations, which indicates that the parameter estimates are accurate. After observing convergence and controlling Monte Carlo error, we then constructed model parameter estimates and fit indices for each model.

After convergence, DIC and its components, EBIC and EAIC for each model were obtained and summarized in Table 2. Based on $D(\overline{\Theta})$, the deviances became smaller from the no growth model to the latent basis growth curve model. However, at the same time, the models became more complex with increasing p_D. Overall, the robust latent basis growth curve model had the smallest DIC (3898), EBIC (2814), and EAIC (2777), and thus the robust latent basis growth curve model was selected as the best fit model.

Table 2 Comparisons of models using DIC, EBIC, & EAIC

Model	$\overline{D(\Theta)}$	$D(\overline{\Theta})$	p_D	DIC	p	EBIC	EAIC
No growth	3965	3522	443	4408	4	3545	3530
Linear	3374	2816	558	3932	7	2856	2830
Latent basis	3327	2757	571	3898	10	2814	2777

The parameter estimates for the robust latent basis growth curve model are provided in Table 3. The estimated degrees of freedom for t distributed measurement errors is 2.919. The small degrees of freedom indicates that modeling the errors as normally distributed may not be sufficient. For the purpose of comparison, we also fit the normal latent basis growth curve model to the data using SAS PROC NLMIXED. Comparing the parameter estimates obtained using the normal distribution with those using the t distribution, the fixed-effects parameter estimates for the initial level and rate of growth are very close. However, the estimate for σ_S^2, the variability of slope, from the robust latent basis growth curve model is substantially different from that from the normal latent basis growth curve model. Especially, if one conducts a z-test for the significance of the individual differences of rates of growth on σ_S^2, it is significant for the robust model but insignificant for the normal model. Furthermore, the standard errors for the robust model are all smaller than the counterparts of the normal model. Based on our simulation studies followed immediately, the results from the robust model should be trusted.

Simulation Studies

We have shown how to apply the robust growth curve models in real data analysis. In this section, we first evaluate the performance of the robust growth curve models using the t distribution through three simulation studies and then evaluate the performance of deviance information criteria for model selection in the fourth simulation study. In the first simulation, we investigate whether the model parameters can be recovered in the robust growth curve models when the data are simulated from the robust models. In the second simulation, we investigate whether the robust growth curve models can be used to analyze normal data that are generated from the normal growth curve models. In the third simulation, we evaluate the performance of the robust growth curve models when the data are contaminated normal with outliers. In summary, we want to investigate whether the robust growth curve models can deal with heavy-tailed data, normal data, and normal data with outliers. In these three simulations, the focus is on a robust latent basis growth curve model. In the fourth simulation study, we investigate whether the use of DIC, EBIC, and EAIC can select correct growth curve models.

We first discuss the design of the first three simulation studies. The population parameters of the robust latent basis growth curve model in the simulation are given by

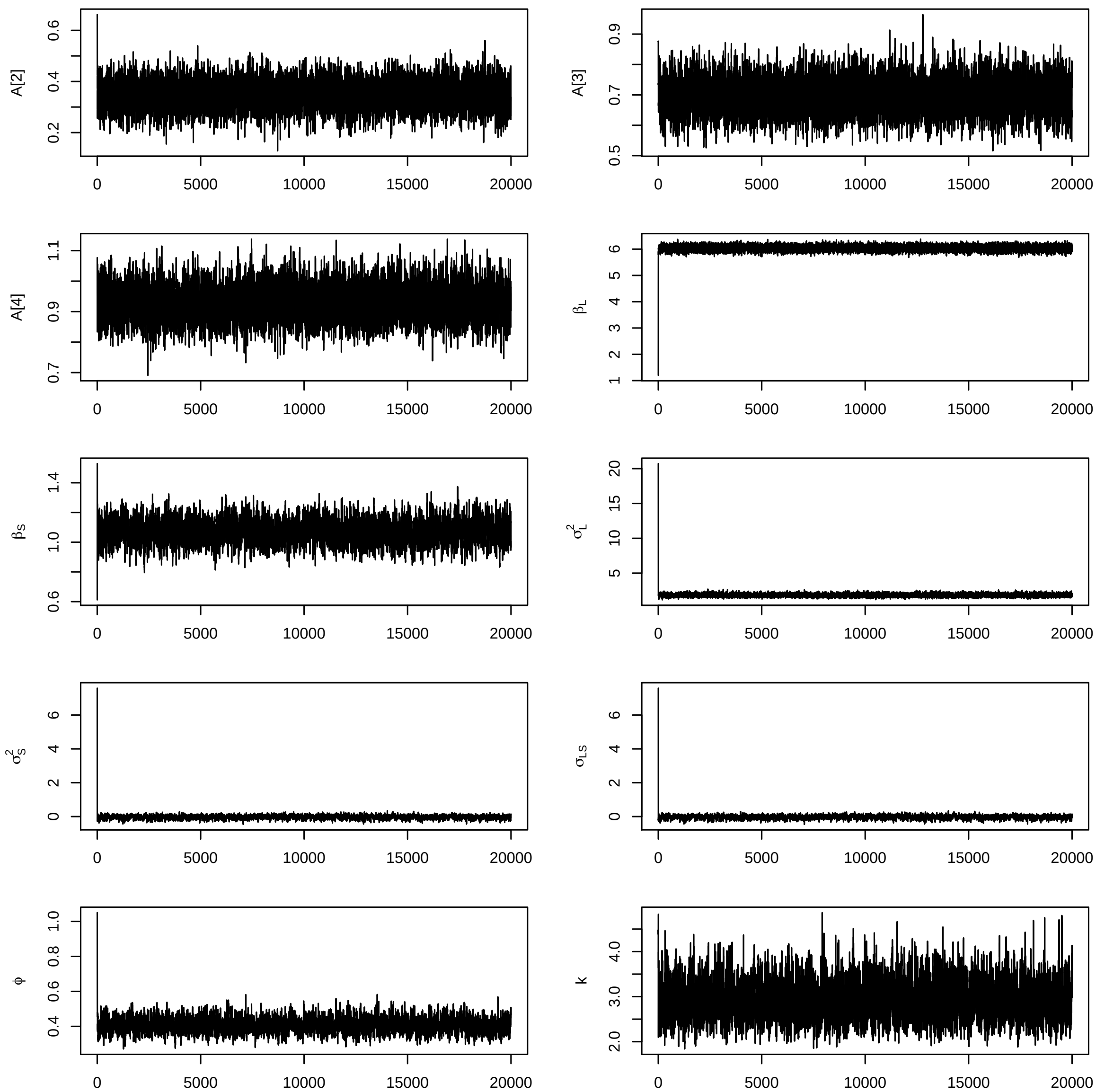

Figure 2 History plots of robust latent basis growth curve model parameters. The plots indicate that the Markov chains converged rapidly.

Table 3 Parameter estimates for the PIAT math data using robust and normal latent basis growth curve models

	Parameter	Estimate	sd	Est/sd	MC_error/sd	Geweke
Robust model	A[2]	0.348	0.049	7.115	1.033%	-0.691
	A[3]	0.695	0.050	14.001	0.994%	-0.391
	A[4]	0.929	0.051	18.176	1.109%	-0.545
	β_L	6.036	0.090	67.126	0.832%	-0.330
	β_S	1.071	0.071	15.029	1.462%	0.158
	σ_L^2	1.822	0.188	9.691	0.929%	0.469
	σ_S^2	0.314	0.096	3.259	2.329%	0.445
	σ_{LS}	-0.030	0.096	-0.310	1.709%	0.823
	σ_e^2	0.406	0.039	10.313	1.818%	-0.847
	k	2.919	0.381	7.663	1.429%	-0.988
Normal model	A[2]	0.222	0.079	2.811	-	-
	A[3]	0.69	0.073	9.4	-	-
	A[4]	1.009	0.084	11.973	-	-
	β_L	6.043	0.098	61.436	-	-
	β_S	1.009	0.091	11.129	-	-
	σ_L^2	1.817	0.206	8.84	-	-
	σ_S^2	0.251	0.145	1.73	-	-
	σ_{LS}	-0.032	0.131	-0.244	-	-
	σ_e^2	1.007	0.05	20.007	-	-

Note. sd: posterior standard deviation; Est/sd is the ratio of parameter estimate and its standard deviation; MC_error/sd is the ratio of the Monte Carlo error and standard deviation; Geweke: Geweke statistic.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1/(T-1) \\ 1 & 2/(T-1) \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}, \boldsymbol{\eta}_i = \begin{bmatrix} L_i \\ S_i \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_L \\ \beta_S \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \mathbf{\Psi} = \begin{bmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & .5 \end{bmatrix}. \quad (19)$$

A total of T occasions of data are generated for each individual. Based on $\mathbf{\Lambda}$, the simulated data have a linear trajectory and thus the deviation from linear can be easily observed. Furthermore, to simulate heavy-tailed data, we generate $\mathbf{e}_i$ from the multivariate t distribution $MT_T(\mathbf{0}, \mathbf{I}_{T\times T}\sigma_e^2, k)$ with $k = 3$ and $\sigma_e^2 = 0.5$. To simulate normal data, we generate $\mathbf{e}_i$ from a multivariate normal distribution $MN_T(\mathbf{0}, \mathbf{I}_{T\times T}\sigma_e^2)$ with $\sigma_e^2 = 1.5$. In the third simulation, to generate data with outliers, we randomly select 1% samples from the normal data and replace them with random values about 3 standard deviations away from the mean. For convenience, we call the three types of data t data, normal data, and outlier data, respectively. Note that the population parameters in the simulation are similar to those from the empirical data analysis in the previous section except for the basis coefficients. We vary the sample size in the simulation with $N = 100, 200, 300, 400,$ and 500 and the number of measurement occasions with $T = 4$ and $T = 5$.

For each simulation study, a total of 1000 data sets are generated and analyzed. Let θ denote a parameter and also its population value in the simulation, and let $\hat{\theta}_r, r = 1, \ldots, 1000$ denote its estimates from the rth simulation replication. Furthermore, let $\hat{s}_r$ denote the posterior standard deviation of $\hat{\theta}_r$, and let $\hat{l}_r$ and $\hat{u}_r$ denote the lower and upper limits of the 95% credible interval for θ constructed based on the posterior standard deviation and normal assumption in the rth replication, respectively.[1] For each simulation study, six statistics will be reported.

The first statistic is the parameter estimate that is calculated as the average of parameter estimates of

[1]The posterior standard deviation is analogical to the frequentist standard error and the posterior credible interval, also called credible interval or Bayesian confidence interval, is analogical to the frequentist confidence interval. For the credible interval, $[\hat{l}_r, \hat{u}_r] = [\hat{\theta}_r + \Phi^{-1}(0.025) \times \hat{s}_r, \hat{\theta}_r + \Phi^{-1}(0.975) \times \hat{s}_r]$ where Φ is the normal distribution function.

1000 simulation replications

$$\text{Estimate} = \frac{1}{1000}\sum_{r=1}^{1000}\hat{\theta}_r. \tag{20}$$

The second one is the relative bias

$$\text{Bias} = \begin{cases} 100 \times \left[\frac{\text{Estimate}}{\theta} - 1\right] & \theta \neq 0 \\ 100 \times \text{Estimate} & \theta = 0 \end{cases}. \tag{21}$$

Note that the relative bias is rescaled by multiplying 100. Smaller relative bias indicates that the point estimate is less biased and thus more accurate. The third statistic is the empirical standard deviation (ESD),

$$\text{ESD} = \frac{1}{999}\sum_{r=1}^{1000}(\hat{\theta}_r - \text{Estimate})^2. \tag{22}$$

The fourth one is the average posterior standard deviation (ASD),

$$\text{ASD} = \frac{1}{1000}\sum_{r=1}^{1000}\hat{s}_r. \tag{23}$$

If the standard deviation are precisely estimated, ASD should be very close to ESD. The fifth statistic is the coverage probability (CVG) of the 95% credible interval of each parameter. The CVG is calculated as

$$\text{CVG} = \frac{\#(\hat{l}_r < \theta < \hat{u}_r)}{1000} \tag{24}$$

where $\#(\hat{l}_r < \theta < \hat{u}_r)$ is the total number of replications with credible intervals covering the true parameter θ. Good 95% credible intervals should give coverage probabilities close to 0.95. The sixth one is statistical power or Type I error that is calculated by

$$\text{power/Type I error} = \frac{\#(\hat{l}_r > 0) + \#(\hat{u}_r < 0)}{1000} \tag{25}$$

where $\#(\hat{l}_i > 0)$ is the total number of replications with the lower limits of credible intervals larger than 0 and $\#(\hat{u}_r < 0)$ is the total number of replications with the upper limits smaller than 0. If $\theta = 0$, it is Type I error and otherwise statistical power.

For the fourth simulation, we focus on whether we can select the correct model from the no growth model, the linear model and the latent basis model based on DIC, EBIC, and EAIC. In the first condition of this simulation study, we generate 1000 sets of data from a robust latent basis growth curve model with $A = [0, .35, .7, .9, 1]$. In the second condition, we generate 1000 sets of data from a robust linear growth curve model. Then, under each condition, we fit the no growth, linear growth, and latent basis growth curve models to the data and calculate their DIC, EBIC, and EAIC. The model with the smallest DIC, EBIC, and EAIC is retained as the best fit model for the analysis of each generated data set.

Simulation study 1: Analysis of t data

In this simulation study, we evaluate the performance of the robust latent basis growth curve model in analyzing t data. The results for this simulation are given in Table 4.[2] First, the relative bias for parameter

[2]For the sake of space, only results for $T = 5$ and $N = 100, 200, 300$, and 400 are reported here. Results for other conditions are similar and available on request.

estimates is smaller than or around 5% except for k or when the sample size is small (e.g., $N = 100$). Although the degrees of freedom cannot be estimated accurately with a small sample size ($N < 300$), the growth curve related parameter estimates are accurate even with a small sample size 100. Second, the average posterior standard deviation (ASD) is close to the empirical standard deviation (ESD) indicating that the standard deviation estimates are also accurate. Third, the coverage probability for the 95% credible intervals is close to the nominal level 95% with almost all falling in the interval of [0.936, 0.964]. Furthermore, with the increase of sample sizes, both parameter estimate biases and posterior standard deviations become smaller.

Table 4 Analysis of t data using the robust Bayesian method

		True	Estimate	Bias(%)	ESD	ASD	CVG	Power/Type I
100	$A[2]$	0.25	0.239	-4.335	0.105	0.107	0.961	0.642
	$A[3]$	0.5	0.501	0.143	0.095	0.101	0.968	0.991
	$A[4]$	0.75	0.763	1.758	0.109	0.107	0.958	0.999
	β_L	6	6.017	0.286	0.166	0.170	0.956	1.000
	β_S	1	0.965	-3.539	0.137	0.142	0.948	1.000
	σ_L^2	2	2.124	6.178	0.361	0.387	0.968	1.000
	σ_S^2	0.5	0.526	5.199	0.244	0.259	0.927	0.446
	σ_{LS}	0	-0.014	-1.401	0.208	0.222	0.966	0.034
	σ_e^2	0.5	0.528	5.533	0.079	0.078	0.943	1.000
	k	3	3.563	18.774	1.244	0.903	0.982	0.990
200	$A[2]$	0.25	0.245	-1.902	0.069	0.071	0.954	0.901
	$A[3]$	0.5	0.501	0.257	0.066	0.068	0.957	1.000
	$A[4]$	0.75	0.757	0.899	0.071	0.071	0.956	1.000
	β_L	6	6.006	0.101	0.115	0.118	0.967	1.000
	β_S	1	0.984	-1.555	0.097	0.099	0.950	1.000
	σ_L^2	2	2.059	2.963	0.251	0.262	0.960	1.000
	σ_S^2	0.5	0.505	0.907	0.183	0.183	0.924	0.844
	σ_{LS}	0	-0.003	-0.295	0.152	0.155	0.952	0.048
	σ_e^2	0.5	0.513	2.636	0.055	0.054	0.949	1.000
	k	3	3.218	7.280	0.544	0.494	0.972	1.000
300	$A[2]$	0.25	0.245	-1.802	0.058	0.057	0.952	0.979
	$A[3]$	0.5	0.501	0.289	0.053	0.054	0.950	1.000
	$A[4]$	0.75	0.754	0.516	0.056	0.057	0.953	1.000
	β_L	6	6.003	0.053	0.095	0.096	0.952	1.000
	β_S	1	0.990	-1.034	0.081	0.080	0.948	1.000
	σ_L^2	2	2.035	1.733	0.205	0.211	0.957	1.000
	σ_S^2	0.5	0.498	-0.326	0.151	0.148	0.933	0.961
	σ_{LS}	0	0.001	0.135	0.125	0.126	0.949	0.051
	σ_e^2	0.5	0.511	2.204	0.044	0.044	0.954	1.000
	k	3	3.154	5.138	0.420	0.387	0.960	1.000
400	$A[2]$	0.25	0.247	-1.215	0.049	0.049	0.954	0.998
	$A[3]$	0.5	0.501	0.130	0.047	0.047	0.950	1.000
	$A[4]$	0.75	0.753	0.389	0.048	0.049	0.954	1.000
	β_L	6	6.001	0.018	0.082	0.083	0.953	1.000
	β_S	1	0.993	-0.731	0.069	0.069	0.954	1.000
	σ_L^2	2	2.031	1.548	0.180	0.182	0.951	1.000
	σ_S^2	0.5	0.503	0.615	0.129	0.128	0.940	0.992
	σ_{LS}	0	-0.003	-0.287	0.109	0.109	0.951	0.049
	σ_e^2	0.5	0.507	1.416	0.038	0.038	0.952	1.000
	k	3	3.116	3.856	0.343	0.328	0.954	1.000

For the purpose of comparison, we also analyzed the t data with the normal latent basis growth curve model and summarized the results in Table 5. The results show that the relative biases of parameter estimates are large even with a sample size of 400. Especially, σ_S^2 is over-estimated more than 40% and σ_{LS} is under-

Table 5 Analysis of t-data using normal model

		True	Estimate	Bias(%)	ESD	ASD	CVG	Power/Type I
100	$A[2]$	0.25	0.314	25.621	0.778	0.228	0.793	0.535
	$A[3]$	0.5	0.598	19.692	1.024	0.302	0.817	0.877
	$A[4]$	0.75	0.877	16.872	1.105	0.380	0.776	0.955
	β_L	6	6.048	0.808	0.208	0.183	0.910	1.000
	β_S	1	0.902	-9.837	0.278	0.181	0.847	0.944
	σ_L^2	2	2.170	8.495	1.379	0.436	0.865	0.986
	σ_S^2	0.5	0.927	85.416	2.855	0.481	0.625	0.511
	σ_{LS}	0	-0.254	-25.405	1.527	0.344	0.725	0.275
	σ_e^2	-	1.344	-	0.765	0.114	-	-
200	$A[2]$	0.25	0.262	4.753	0.511	0.148	0.808	0.671
	$A[3]$	0.5	0.542	8.500	0.395	0.118	0.815	0.959
	$A[4]$	0.75	0.788	5.123	0.423	0.130	0.776	0.984
	β_L	6	6.031	0.511	0.154	0.130	0.914	1.000
	β_S	1	0.938	-6.205	0.216	0.130	0.874	0.974
	σ_L^2	2	2.102	5.123	0.790	0.303	0.830	0.993
	σ_S^2	0.5	0.763	52.679	1.680	0.342	0.609	0.586
	σ_{LS}	0	-0.149	-14.932	0.934	0.243	0.696	0.304
	σ_e^2	-	1.382	-	0.453	0.082	-	-
300	$A[2]$	0.25	0.255	1.860	0.224	0.086	0.817	0.792
	$A[3]$	0.5	0.511	2.210	0.262	0.084	0.814	0.981
	$A[4]$	0.75	0.767	2.270	0.256	0.086	0.807	0.997
	β_L	6	6.019	0.314	0.129	0.106	0.903	1.000
	β_S	1	0.959	-4.088	0.171	0.107	0.878	0.993
	σ_L^2	2	2.073	3.645	0.603	0.245	0.826	0.996
	σ_S^2	0.5	0.711	42.218	1.284	0.278	0.592	0.638
	σ_{LS}	0	-0.112	-11.166	0.703	0.197	0.692	0.308
	σ_e^2	-	1.397	-	0.404	0.067	-	-
400	$A[2]$	0.25	0.208	-16.845	0.970	0.213	0.841	0.870
	$A[3]$	0.5	0.500	-0.099	0.777	0.172	0.818	0.985
	$A[4]$	0.75	0.740	-1.332	0.653	0.136	0.814	0.992
	β_L	6	6.017	0.278	0.117	0.092	0.900	1.000
	β_S	1	0.964	-3.583	0.157	0.092	0.882	0.993
	σ_L^2	2	2.075	3.773	0.577	0.213	0.821	0.997
	σ_S^2	0.5	0.720	43.929	1.212	0.242	0.596	0.701
	σ_{LS}	0	-0.115	-11.536	0.687	0.172	0.699	0.301
	σ_e^2	-	1.410	-	0.407	0.059	-	-

estimated more than 10% even for $N = 400$. Furthermore, these biases do not seem to monotonously decrease with the increase of sample sizes. Comparing empirical standard deviation (ESD) and the average posterior standard deviation (ASD), the posterior standard deviation are uniformly under-estimated more than 100% for most parameters. The standard deviations for the normal model are also consistently larger than those from the robust model. Furthermore, the coverage probabilities are consistently smaller than the nominal level 95%. Comparing the power in Table 4 and that in Table 5, the power is consistently smaller if the t data are analyzed as normal data. At the same time, the Type I error for σ_{LS} is over-estimated. Thus, using the normal model to analyze the t data, one may incorrectly conclude that there is no individual differences in the slope, and the slope and the initial level are negatively correlated in the growth curve model.

Overall, the robust latent growth curve model performs very well for the analysis of t data. The normal latent basis growth curve model , on the other hand, does not seem to work for t data.

Simulation study 2: Analysis of normal data

In this simulation study, we investigate whether the robust growth curve models can be applied to analyze normal data. For comparison, the generated normal data are analyzed by both the robust and normal latent basis growth curve models. The results from the robust model are given in Table 6 and the results for the normal model are summarized in Table 7.

Overall, it seems that both the robust and normal latent basis growth curve models can estimate model parameters very well except when the sample size is as small as $N = 100$. With $N = 100$, the robust model has much larger biases than the normal model. For example, the bias for σ_S^2 is about 22% for robust model but is only about -6% for the normal model. When the sample size is 200 or bigger, the parameter biases are smaller than 5% for both robust and normal models except for A[2] in the robust model; ESD and ASD are approximately equal; and the coverage probabilities are close to the nominal level 95%. On average, the normal model has smaller biases, smaller standard deviation estimates and larger power. Overall, the robust model obtains comparable results as the normal model for normal data analysis. Although the use of robust model for normal data reduces efficiency, the reduction in efficiency is not big. Practically, if the estimated degrees of freedom are larger than 30, one would switch to the normal growth curve model. Thus, the reduction in the efficiency may be avoided.

Simulation study 3: Analysis of normal data with outliers

In this simulation study, we investigate whether the robust growth curve models can be applied to analyze data with outliers. The results for data analysis using the robust model and the normal model are given in Table 8 and Table 9, respectively.

When the sample size is 100, neither the robust model nor the normal model seems to work well although the robust model still outperforms the normal model. For example, the bias for σ_S^2 is about 30% for the robust model and about 174% for the regular model. When the sample size increases to 200, the parameter estimates for the robust model are greatly improved but the biases for the normal model are still very large. For the robust model, only the bias for σ_S^2 is big and the posterior standard deviation estimates and coverage probabilities perform well in general. On the other hand, both posterior standard deviations and coverage probabilities are greatly under-estimated for the normal model.

The following discussion focuses on $N = 400$ when the biases for both models seem to be small for all parameters except for σ_S^2. For the robust model, other than the slightly over-estimated σ_S^2, parameter estimates, posterior standard deviation estimates, and coverage probabilities appear very good for all model parameters. However, for the normal model, the parameter estimates for both σ_S^2 and σ_{LS} are biased and the coverage probabilities for most parameters are either over-estimated or under-estimated. The power for the normal model is also smaller than the robust model. Furthermore, the Type I error for σ_{LS} is correct for the robust model but over-estimated for the normal model. Overall, the robust model performs better than the normal model to analyze outlier data in this simulation study.

In summary, the robust growth curve model outperform the regular model when data are t distributed or contaminated with outliers. For the normal data, although the robust model is not as efficient as the normal model, the reduction in efficiency is generally small.

Simulation study 4: Model selection using DIC, EBIC, and EAIC

In order to evaluate the performance of DIC, EBIC and EAIC in model selection, we first simulated data from a robust latent basis growth curve model and then fitted robust no growth, linear, and latent basis models to the data. The number of replications that a model is selected based on the smallest DIC, EBIC, or EAIC is given in Table 10.[3] Overall, with larger sample sizes and longer measurement occasions, the

[3]In all simulations, the robust no growth curve model has never been selected and thus is excluded in the table for the sake of space.

Table 6 Analysis of normal data using the robust Bayesian method

		True	Estimate	Bias(%)	ESD	ASD	CVG	Power/Type I
	$A[2]$	0.25	0.221	-11.679	0.189	0.202	0.968	0.328
	$A[3]$	0.5	0.507	1.482	0.191	0.190	0.971	0.828
	$A[4]$	0.75	0.783	4.427	0.185	0.204	0.972	0.971
	β_L	6	6.037	0.611	0.184	0.191	0.949	1.000
	β_S	1	0.922	-7.820	0.201	0.193	0.920	0.995
	σ_L^2	2	2.179	8.971	0.411	0.445	0.970	1.000
100	σ_S^2	0.5	0.610	21.963	0.347	0.397	0.971	0.107
	σ_{LS}	0	-0.086	-8.628	0.278	0.307	0.964	0.036
	σ_e^2	1.5	1.446	-3.591	0.114	0.125	0.931	1.000
	k	∞	60.623	-	6.534	23.316	-	-
	$A[2]$	0.25	0.237	-5.209	0.115	0.115	0.961	0.577
	$A[3]$	0.5	0.504	0.814	0.114	0.110	0.949	0.982
	$A[4]$	0.75	0.766	2.070	0.115	0.115	0.952	1.000
	β_L	6	6.015	0.245	0.129	0.133	0.955	1.000
	β_S	1	0.965	-3.505	0.135	0.133	0.938	1.000
	σ_L^2	2	2.085	4.259	0.293	0.302	0.957	1.000
200	σ_S^2	0.5	0.525	4.919	0.267	0.285	0.939	0.295
	σ_{LS}	0	-0.032	-3.244	0.203	0.217	0.962	0.038
	σ_e^2	1.5	1.454	-3.041	0.080	0.089	0.933	1.000
	k	∞	64.444	-	6.515	21.584	-	-
	$A[2]$	0.25	0.241	-3.435	0.091	0.091	0.960	0.757
	$A[3]$	0.5	0.504	0.745	0.087	0.087	0.955	1.000
	$A[4]$	0.75	0.759	1.158	0.093	0.091	0.960	1.000
	β_L	6	6.011	0.175	0.105	0.108	0.956	1.000
	β_S	1	0.976	-2.395	0.108	0.108	0.931	1.000
	σ_L^2	2	2.053	2.630	0.248	0.243	0.952	1.000
300	σ_S^2	0.5	0.498	-0.446	0.230	0.241	0.959	0.465
	σ_{LS}	0	-0.013	-1.327	0.168	0.180	0.959	0.041
	σ_e^2	1.5	1.458	-2.785	0.067	0.073	0.918	1.000
	k	∞	67.247	-	6.431	20.262	-	-
	$A[2]$	0.25	0.245	-1.930	0.078	0.077	0.964	0.852
	$A[3]$	0.5	0.504	0.773	0.074	0.074	0.959	1.000
	$A[4]$	0.75	0.755	0.698	0.078	0.077	0.950	1.000
	β_L	6	6.007	0.110	0.090	0.093	0.963	1.000
	β_S	1	0.984	-1.563	0.093	0.093	0.946	1.000
	σ_L^2	2	2.031	1.549	0.210	0.209	0.945	1.000
400	σ_S^2	0.5	0.486	-2.818	0.214	0.215	0.948	0.618
	σ_{LS}	0	-0.004	-0.377	0.154	0.158	0.954	0.046
	σ_e^2	1.5	1.461	-2.581	0.058	0.063	0.909	1.000
	k	∞	69.771	-	6.081	19.229	-	-

Table 7 Analysis of normal data using normal model

		True	Estimate	Bias(%)	ESD	ASD	CVG	Power/Type I
100	$A[2]$	0.25	0.227	-9.349	0.183	0.164	0.956	0.361
	$A[3]$	0.5	0.498	-0.318	0.167	0.155	0.959	0.874
	$A[4]$	0.75	0.756	0.790	0.174	0.162	0.948	0.991
	β_L	6	6.009	0.152	0.185	0.185	0.946	1.000
	β_S	1	0.978	-2.248	0.197	0.185	0.933	0.996
	σ_L^2	2	1.979	-1.060	0.417	0.416	0.941	1.000
	σ_S^2	0.5	0.470	-6.037	0.511	0.460	0.912	0.156
	σ_{LS}	0	0.013	1.282	0.350	0.329	0.932	0.068
	σ_e^2	1.5	1.485	-0.967	0.129	0.124	0.933	1.000
200	$A[2]$	0.25	0.238	-4.932	0.110	0.109	0.958	0.606
	$A[3]$	0.5	0.497	-0.512	0.106	0.104	0.959	0.990
	$A[4]$	0.75	0.746	-0.502	0.111	0.109	0.949	1.000
	β_L	6	6.002	0.038	0.132	0.131	0.949	1.000
	β_S	1	0.990	-0.988	0.135	0.131	0.935	1.000
	σ_L^2	2	1.996	-0.223	0.294	0.295	0.946	1.000
	σ_S^2	0.5	0.488	-2.383	0.335	0.326	0.941	0.289
	σ_{LS}	0	0.010	0.996	0.236	0.233	0.946	0.054
	σ_e^2	1.5	1.489	-0.714	0.084	0.087	0.952	1.000
300	$A[2]$	0.25	0.241	-3.539	0.090	0.088	0.946	0.747
	$A[3]$	0.5	0.499	-0.242	0.084	0.084	0.957	0.999
	$A[4]$	0.75	0.751	0.107	0.088	0.088	0.956	1.000
	β_L	6	6.003	0.044	0.106	0.107	0.952	1.000
	β_S	1	0.992	-0.830	0.110	0.107	0.941	1.000
	σ_L^2	2	1.999	-0.065	0.236	0.241	0.946	1.000
	σ_S^2	0.5	0.489	-2.271	0.272	0.265	0.937	0.448
	σ_{LS}	0	0.005	0.536	0.191	0.190	0.950	0.050
	σ_e^2	1.5	1.494	-0.387	0.070	0.071	0.944	1.000
400	$A[2]$	0.25	0.244	-2.213	0.077	0.076	0.940	0.869
	$A[3]$	0.5	0.500	-0.059	0.072	0.073	0.944	1.000
	$A[4]$	0.75	0.750	-0.040	0.075	0.076	0.954	1.000
	β_L	6	6.002	0.038	0.091	0.093	0.952	1.000
	β_S	1	0.994	-0.648	0.096	0.093	0.945	1.000
	σ_L^2	2	2.004	0.184	0.204	0.209	0.947	1.000
	σ_S^2	0.5	0.489	-2.215	0.235	0.230	0.937	0.571
	σ_{LS}	0	0.004	0.422	0.165	0.165	0.954	0.046
	σ_e^2	1.5	1.496	-0.257	0.061	0.061	0.944	1.000

correct robust latent basis growth curve model is more likely to be selected. In this condition, DIC seems to perform better than EAIC which, in turn, performs better than EBIC. For example, when $N = 300$ and $T = 5$, DIC can correctly select the robust latent basis model 995 times out of 1,000. However, EAIC and EBIC only correctly select the model 97.8% and 90.4% of the time, respectively.

Then we simulated data from a linear growth curve model to investigate the performance of DIC, EBIC, and EAIC. The number of times that a model is selected based on the smallest DIC, EBIC, or EAIC is given in Table 11. Overall, it seems that EBIC performs better than DIC and EAIC in this condition especially when $T = 5$. Furthermore, all three fit indices seem to have the tendency to prefer the more complex models. The observation is consistent with the existing literature (Spiegelhalter et al., 2002; Zhang & Nesselroade, 2007).

Table 8 Analysis of outlier data using the robust Bayesian method

		True	Estimate	Bias(%)	ESE	ASE	CVG	Power/Type I
	$A[2]$	0.25	0.206	-17.738	0.255	0.267	0.966	0.264
	$A[3]$	0.5	0.51	2.068	0.241	0.255	0.973	0.712
	$A[4]$	0.75	0.806	7.423	0.279	0.275	0.969	0.915
	β_L	6	6.046	0.767	0.199	0.203	0.943	1
	β_S	1	0.905	-9.476	0.229	0.214	0.914	0.973
	σ_L^2	2	2.198	9.922	0.447	0.485	0.978	1
100	σ_S^2	0.5	0.64	27.969	0.369	0.441	0.975	0.067
	σ_{LS}	0	-0.101	-10.058	0.297	0.341	0.968	0.032
	σ_e^2	-	1.439	-	0.156	0.185	-	-
	k	-	4.761	-	1.671	1.243	-	-
	$A[2]$	0.25	0.233	-6.691	0.128	0.129	0.958	0.494
	$A[3]$	0.5	0.505	0.959	0.13	0.123	0.956	0.966
	$A[4]$	0.75	0.768	2.379	0.133	0.129	0.953	1
	β_L	6	6.02	0.326	0.136	0.141	0.954	1
	β_S	1	0.959	-4.073	0.15	0.146	0.933	1
	σ_L^2	2	2.094	4.69	0.313	0.327	0.955	1
200	σ_S^2	0.5	0.552	10.394	0.28	0.315	0.956	0.238
	σ_{LS}	0	-0.044	-4.389	0.216	0.241	0.959	0.041
	σ_e^2	-	1.428	-	0.104	0.13	-	-
	k	-	4.357	-	0.634	0.699	-	-
	$A[2]$	0.25	0.238	-4.661	0.1	0.101	0.961	0.67
	$A[3]$	0.5	0.503	0.619	0.097	0.096	0.954	0.997
	$A[4]$	0.75	0.76	1.312	0.105	0.101	0.95	1
	β_L	6	6.014	0.233	0.109	0.115	0.955	1
	β_S	1	0.974	-2.594	0.12	0.119	0.945	1
	σ_L^2	2	2.063	3.136	0.266	0.264	0.955	1
300	σ_S^2	0.5	0.532	6.443	0.253	0.267	0.935	0.437
	σ_{LS}	0	-0.028	-2.786	0.186	0.199	0.957	0.043
	σ_e^2	-	1.425	-	0.086	0.107	-	-
	k	-	4.259	-	0.452	0.547	-	-
	$A[2]$	0.25	0.244	-2.349	0.085	0.085	0.961	0.796
	$A[3]$	0.5	0.503	0.659	0.084	0.082	0.957	0.999
	$A[4]$	0.75	0.756	0.83	0.087	0.085	0.95	1
	β_L	6	6.009	0.142	0.093	0.099	0.969	1
	β_S	1	0.984	-1.65	0.102	0.102	0.956	1
	σ_L^2	2	2.042	2.084	0.226	0.229	0.957	1
400	σ_S^2	0.5	0.53	6.027	0.229	0.238	0.936	0.597
	σ_{LS}	0	-0.024	-2.44	0.166	0.176	0.957	0.043
	σ_e^2	-	1.422	-	0.074	0.093	-	-
	k	-	4.226	-	0.322	0.467	-	-

Table 9 Analysis of outlier data using normal model

		True	Estimate	Bias(%)	ESE	ASE	CVG	Power/Type I
100	$A[2]$	0.25	2.089	735.563	18.043	6.968	0.798	0.410
	$A[3]$	0.5	1.160	131.977	8.283	1.542	0.803	0.713
	$A[4]$	0.75	0.980	30.632	7.102	1.153	0.795	0.862
	β_L	6	6.183	3.057	0.242	0.217	0.825	1.000
	β_S	1	0.812	-18.769	0.381	0.251	0.791	0.768
	σ_L^2	2	2.288	14.379	1.167	0.610	0.816	0.918
	σ_S^2	0.5	1.368	173.533	1.622	0.645	0.476	0.609
	σ_{LS}	0	-0.458	-45.800	1.237	0.549	0.756	0.244
	σ_e^2	-	2.697	-	0.631	0.217	-	-
200	$A[2]$	0.25	0.493	97.161	6.261	0.191	0.884	0.482
	$A[3]$	0.5	0.565	13.098	0.295	0.176	0.879	0.915
	$A[4]$	0.75	0.785	4.689	0.622	0.187	0.864	0.990
	β_L	6	6.127	2.118	0.155	0.157	0.893	1.000
	β_S	1	0.936	-6.362	0.220	0.184	0.916	0.977
	σ_L^2	2	2.116	5.788	0.644	0.425	0.867	0.993
	σ_S^2	0.5	0.880	75.962	0.968	0.467	0.529	0.605
	σ_{LS}	0	-0.224	-22.382	0.718	0.398	0.795	0.205
	σ_e^2	-	2.871	-	0.418	0.163	-	-
300	$A[2]$	0.25	0.269	7.791	0.132	0.130	0.947	0.538
	$A[3]$	0.5	0.525	4.968	0.128	0.123	0.938	0.988
	$A[4]$	0.75	0.772	2.878	0.153	0.129	0.923	0.999
	β_L	6	6.120	1.993	0.114	0.128	0.874	1.000
	β_S	1	0.972	-2.760	0.142	0.148	0.958	1.000
	σ_L^2	2	2.044	2.225	0.418	0.334	0.892	1.000
	σ_S^2	0.5	0.681	36.106	0.690	0.400	0.625	0.523
	σ_{LS}	0	-0.124	-12.360	0.471	0.319	0.845	0.155
	σ_e^2	-	2.924	-	0.312	0.136	-	-
400	$A[2]$	0.25	0.260	3.966	0.098	0.112	0.972	0.656
	$A[3]$	0.5	0.515	2.940	0.091	0.106	0.979	0.999
	$A[4]$	0.75	0.762	1.540	0.108	0.110	0.958	1.000
	β_L	6	6.120	2.003	0.097	0.111	0.835	1.000
	β_S	1	0.982	-1.832	0.110	0.127	0.974	1.000
	σ_L^2	2	2.021	1.047	0.308	0.284	0.932	1.000
	σ_S^2	0.5	0.565	12.935	0.514	0.361	0.736	0.447
	σ_{LS}	0	-0.075	-7.452	0.331	0.272	0.888	0.112
	σ_e^2	-	2.948	-	0.213	0.119	-	-

Discussion

Although growth curve modeling has become popular in analyzing change and individual differences in change, research on robust growth curve models is still rare. We proposed to model heavy-tailed data or outliers through the t distribution for growth curve analysis. The derived robust growth curve models can be estimated through Bayesian methods utilizing data augmentation and Gibbs sampling algorithms. Selection among different robust growth curve models can be conducted using DIC, EBIC and EAIC. The analysis of mathematical development data showed that the robust latent basis growth curve model best fit the growth trajectory and led to different conclusions in terms of individual differences in growth from the conclusions drawn from the corresponding normal growth curve model. Both the boxplots of data and the estimated small number of degrees of freedom suggested that the robust growth curve model is necessary for the analysis of this data set. The simulation studies further confirmed that the robust growth curve models significantly outperformed the normal growth curve models for both heavy-tailed t data and normal data with outliers but only lost slight efficiency when data were normally distributed.

Table 10 Model selected based on DIC, EBIC and EAIC. The robust latent basis growth curve model is the correct model.

			100	200	300	400	500
T=5	DIC	Linear	75	13	5	2	0
		Latent basis	925	987	995	998	1000
	EBIC	Linear	342	168	96	50	25
		Latent basis	658	832	904	950	975
	EAIC	Linear	135	50	22	7	4
		Latent basis	865	950	978	993	996
T=4	DIC	Linear	154	116	80	52	41
		Latent basis	846	884	920	948	959
	EBIC	Linear	481	432	334	273	245
		Latent basis	519	568	666	727	755
	EAIC	Linear	245	157	122	99	82
		Latent basis	745	843	878	901	928

Note. Linear: robust linear growth curve model. Latent basis: robust latent basis growth curve model. DIC: deviance information criterion. EBIC: extended Bayesian information criterion. EAIC: extended Akaike information criterion.

Table 11 Model selected based on DIC, EBIC and EAIC. The robust linear growth curve model is the correct model.

			100	200	300	400	500
T=5	DIC	Linear	678	669	678	673	665
		Latent basis	322	311	322	327	335
	EBIC	Linear	889	884	892	909	910
		Latent basis	111	116	108	91	90
	EAIC	Linear	722	699	708	702	697
		Latent basis	278	301	292	298	303
T=4	DIC	Linear	587	581	579	610	590
		Latent basis	413	419	421	390	410
	EBIC	Linear	816	810	801	821	820
		Latent basis	184	190	199	179	180
	EAIC	Linear	694	651	633	648	634
		Latent basis	306	349	367	352	366

Note. Linear: robust linear growth curve model. Latent basis: robust latent basis growth curve model. DIC: deviance information criterion. EBIC: extended Bayesian information criterion. EAIC: extended Akaike information criterion.

The finding that the heavy-tailed data affect random-effects parameters but not fixed-effects parameters has important implication in growth curve analysis. Unlike latent variable analysis where the estimated variances for latent variables are of no direct interests, one of the main purposes of growth curve analysis is to investigate the individual differences in change or growth based on the testing of variance parameters for random change. Thus, without dealing with heavy-tailed data, the conclusion from a normal growth curve analysis can be misleading.

The robust growth curve models discussed in this study are parametric and can be explicitly formed in probabilistic settings. In general, the robust models and methods based on the t distribution are easier to understand, estimate, and interpret than M-estimators (Huber, 1981). The use of the t distribution has also computational simplicity compared to the choice of other distributions such as the contaminated normal distribution and the slash distribution. It is possible to develop semi-parametric or nonparametric robust growth curve models in the future.

We have estimated the degrees of freedom k as an unknown parameter in the robust growth curve model. Alternatively, it can be fixed *a priori*. k determines how much one plans to downweight the outliers. If one would downweight more of the outlier, a smaller k can be used. Lange et al. (1989) suggested fixing k for small data sets and estimating k for large data sets. Venables & Ripley (1999, p121) suggested that

$k = 5$ is often a good choice. One advantage to estimate k is that its magnitude provides information on whether a robust model is necessary. From our experience, if $k > 30$, it is safe to use a normal model.

Implementation and extensions

To help the adoption of robust growth curve models, online software has been developed using web programming languages to implement the Bayesian estimation procedure for all models discussed in this study. The software can be used at any place with Internet connection throught the URL http://webstats.psychstat.org/semrgcm/. To use the software, one only needs to upload the data set to be analyzed. The output includes the summary statistics and Anscombe-Glynn kurtosis test, the spaghetti (longitudinal) plot of data, the boxplot of data, Geweke statistic, history plot and autocorrelation plot for convergence diagnostics, and Bayesian parameter estimates. Users can specify prior distributions and starting values of model parameters and control the burn-in period and length and thinning of Markov chains. The software and its manual are freely available at (the link is removed for review purpose). A step-by-step illustration of the software is given in Appendix B.

In order to clearly show the differences between the robust and normal growth curve models, we have assumed that the residuals follow a multivariate t distribution and the random coefficients have a multivariate normal distribution. Admittedly, a general robust growth curve model should allow both residuals and random coefficients to follow t distributions. Therefore, our online software allows the use of any combination of normal and t distributions for residuals and random coefficients. Several different strategies can be used to determine the choice of t distributions fore residual and random coefficients. One strategy is fit growth curves to individuals and then check the distributions of residuals and random coefficients for long-tailness and outliers. Another strategy is to check the estimated degrees of freedom of t distributions. A small degrees of freedom indicates the necessity of the use of t distribution. Fit indices DIC, EBIC, and EAIC can also be used to assist the decision of the choice of t distribution by comparing competing models.

In growth curve analysis, covariates can be used to explain the individual differences in growth parameters such as initial level and change (McArdle & Nesselroade, 2002). Thus, the software also allows the inclusion of any number of covariates. Furthermore, in longitudinal data analysis, missing data are almost inevitable and are a challenge for growth curve modeling. The software deals with missing data in the dependent variable and covariates by sampling them from their posterior distributions. As mentioned earlier, with a small sample size, the degrees of freedom cannot be estimated accurately. Therefore, the software allows the supply of a fixed value as the degrees of freedom.

Limitations

Like normal distribution, t distribution is symmetric and can be sensitive to skewed data. Practically, we found that if the skewness in the data was caused by outliers, the robust growth curve models could still perform very well. However, if the skewness was because of skewed distributions such as lognormal distribution and Gamma distribution, the robust growth curve models could break down although they still worked relatively better than the normal growth curve models. For data with systematical skewness, one may conduct a transformation to symmetry (Hoaglin et al., 1983) before applying the robust models. Another possible solution is to use a skew-t distribution (Azzalini & Genton, 2008) to accommodate to skewness in the data. However, the performance of skew-t distribution in robust growth curve modeling needs further evaluation.

For robust growth curve model selection, we have used DIC, EAIC, and EBIC. However, caution should be used when applying those fit indices for model selection. As shown in our simulation, all three fit indices seem to prefer more complex models. Although DIC is very widely used, it has received many criticisms since it was proposed (Spiegelhalter et al., 2002). Especially, for models involving random-effects as in our robust growth curve models, the computation of DIC is still not very clear (Celeux et al., 2006).

Furthermore, as pointed out by Torre & Douglas (2008), the use of posterior mean in constructing EAIC and EBIC also needs further justification. Thus, the development of better and sophisticated fit indices should be carried out in the future.

References

Anscombe, F., & Glynn, W. (1983). Distribution of kurtosis statistic for normal statistics. *Biometrika*, *70*(1), 227–234.

Azzalini, A., & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, *76*(1), 106–129.

Celeux, G., Forbers, F., Robert, C., & Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651–674.

Congdon, P. (2003). *Applied Bayesian modelling*. New York: John Wiley & Sons, Inc.

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

Gill, P. S. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, *19*, 975–987.

Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). *Robust statistics: The approach based on influence functions*. New York: John Wiley & Sons, Inc.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. New York: John Wiley & Sons.

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

Lachos, V. H., Dey, D. K., & Cancho, V. G. (2009). Robust linear mixed models with skew-normal independent distributions from a bayesian perspective. *Journal of Statistical Planning and Inference*, *139*, 4098–4110.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling uisng the t distribution. *Journal of the Americal Statistical Association*, *84*(408), 881–896.

Lee, S. Y. (2007). *Structual equation modeling: A bayesian approach*. New York: John Wiley & Sons, Inc.

Lee, S. Y., & Shi, J. Q. (2000). Bayesian analysis of structural equation model with fixed covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*, 411–430.

Lee, S. Y., & Song, X. Y. (2008). On bayesian estimation and model comparison of an integrated structural equation model. *Computational Statistics and Data Analysis*, *52*, 4814–4827.

Lee, S. Y., & Xia, Y. M. (2006). Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data. *Psychometrika*, *71*, 541–565.

Lee, S. Y., & Xia, Y. M. (2008). A robust bayesian approach for structural equation models with missing data. *Psychometrika*, *73*, 343–364.

Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *37*(1), 23–38.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, *91*(435), 1219–1227.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. New York: John Wiley & Sons, Inc.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561–614). New York: Plenum Press.

McArdle, J. J., & Nesselroade, J. R. (2002). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology, volume two: Research methods in psychology* (pp. 447–480). New York: Pergamon Press.

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

Pan, J.-X., & Fang, K.-T. (2002). *Growth curve models and statistical diagnostics*. New York: Springer.

Pendergast, J. F., & Broffitt, J. D. (1985). Robust estimation in growth curve models. *Communications in Statistics: Theory and Methods*, *14*, 1919–1939.

Pinheiro, J. C., Liu, C., & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, *10*(2), 249–276.

Press, S. J. (1972). *Applied multivariate analysis*. New York: Holt, Rinehart, and Winston, Inc.

Rachman-Moore, D., & Wolfe, R. G. (1984). Robust analysis of a nonlinear model for multilevel educational survey data. *Journal of Educational Statistics*, *9*(4), 277–293.

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

Shoham, S. (2002). Robust clustering by deterministic agglomeration em of mixtures of multivariate t-distributions. *Pattern Recognition*, *35*, 1127–1142.

Silvapulle, M. J. (1992). On m-methods in growth curve analysis with asymmetric errors. *Journal of Statistical Planning and Inference*, *32*(3), 303–309.

Singer, J. M., & Sen, P. K. (1986). M-methods in growth curve analysis. *Journal of Statistical Planning and Inference*, *13*, 251–261.

Song, P., Zhang, P., & Qu, A. (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distribution. *Statistica Sinica*, *17*, 929–943.

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

Sutradhar, B. C., & Ali, M. M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics - Theory and Methods*, *15*(2), 429–450.

Torre, J. de la, & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*, 595–624.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, *33*, 1–67.

Venables, W. N., & Ripley, B. D. (1999). *Modern applied statistics with s-plus*. Springer.

Wang, H.-X., Zhang, Q.-B., Luo, B., & Wei, S. (2004). Robust mixture modelling using multivariate t-distribution with missing information. *Pattern Recognition Letters*, *25*, 701–710.

Yau, K. K. W., & Kuk, A. Y. C. (2002). Robust estimation in generalized linear mixed models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64*(1), 101–107.

Yeap, B. Y., & Davidian, M. (2001). Robust two-stage estimation in hierarchical nonlinear models. *Biometrics*, *57*(1), 266–272.

Yuan, K.-H., & Bentler, P. M. (1998). Structural equation modeling with robust covariances. *Sociological Methodology*, *28*, 363–396.

Yuan, K.-H., & Bentler, P. M. (2002). On normal theory based inference for multilevel models with distributional violations. *Pschometrika*, *67*, 539–561.

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

Yuan, K.-H., Chan, W., & Bentler, P. M. (2000). Robust transformation with applications to structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *53*, 31–50.

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

Zhang, Z., & Nesselroade, J. R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research*, *42*(4), 729-756.

Zhong, X., & Yuan, K.-H. (2010). Weights. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1617–1620). Thousand Oaks, CA: Sage.

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

Appendix A
Priors and posteriors of robust growth curve models

In this appendix, we first present the prior and posterior distributions of robust growth curve models and then outline the Gibbs sampling algorithm. In the following, we use MN, IG, IW, and U to denote the density function for multivariate normal distribution, inverse Gamma distribution, inverse Wishart distribution, and uniform distribution, respectively.

Priors

For $\boldsymbol{\beta}$, a multivariate normal prior is used

$$p(\boldsymbol{\beta}) = MN_q(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) = (2\pi)^{-q/2}|\boldsymbol{\Sigma}_0|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right]. \tag{26}$$

For $\boldsymbol{\Psi}$, the inverse Wishart prior is used

$$p(\boldsymbol{\Psi}) = IW(m_0, V_0) = \frac{|V_0|^{m_0/2}|\boldsymbol{\Psi}|^{-\frac{m_0+q+1}{2}}\exp\left[-\mathrm{tr}(V_0\boldsymbol{\Psi}^{-1})/2\right]}{2^{m_0 q/2}\Gamma(m_0/2)}. \tag{27}$$

For $\boldsymbol{\Phi}$, the inverse Wishart prior is used

$$p(\boldsymbol{\Phi}) = IW(n_0, W_0) = \frac{|W_0|^{n_0/2}|\boldsymbol{\Phi}|^{-\frac{n_0+T+1}{2}}\exp\left[-\mathrm{tr}(W_0\boldsymbol{\Phi}^{-1})/2\right]}{2^{n_0 T/2}\Gamma(n_0/2)}. \tag{28}$$

If $\boldsymbol{\Phi} = \mathbf{I}_{T\times T}\sigma_e^2$, an inverse Gamma prior for σ_e^2 is used

$$p(\sigma_e^2) = IG(c_0, d_0) = \frac{d_0^{c_0}}{\Gamma(c_0)}(\sigma_e^2)^{-(c_0+1)}\exp\left(-\frac{d_0}{\sigma_e^2}\right). \tag{29}$$

For k, we can use a uniform prior,

$$p(k) = U[a,b] = \frac{1}{b-a}. \tag{30}$$

Some models may involve unknown parameters in $\mathbf{\Lambda}$, for example, the $A[j], j = 2, \ldots, T-1$ in the latent basis growth curve model. For those parameters, a multivariate normal prior can be used so that

$$\boldsymbol{\lambda} = MN(\boldsymbol{\lambda}_0, \mathbf{\Sigma}_\lambda) \tag{31}$$

where $\boldsymbol{\lambda}$ is a vector formed by the unknown elements in $\mathbf{\Lambda}$.

In these priors, $\boldsymbol{\beta}_0$, $\mathbf{\Sigma}_0$, m_0, V_0, n_0, W_0 (or c_0 and d_0), a, b, $\boldsymbol{\lambda}_0$, and $\mathbf{\Sigma}_\lambda$ are pre-determined hyper-parameters. Prior distributions other than those illustrated here can also be used.

Posteriors

Given the above priors, the following conditional posterior distributions can be obtained. The conditional posterior distribution for $\boldsymbol{\beta}$ is a multivariate normal distribution

$$p(\boldsymbol{\beta}|\mathbf{\Psi}, \boldsymbol{\eta}_i, i = 1, \ldots, N) = MN_q(\boldsymbol{\beta}_1, \mathbf{\Sigma}_1) \tag{32}$$

where

$$\boldsymbol{\beta}_1 = \mathbf{\Sigma}_1(\mathbf{\Psi}^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i + \mathbf{\Sigma}_0^{-1}\boldsymbol{\beta}_0) \tag{33}$$

and

$$\mathbf{\Sigma}_1 = (N\mathbf{\Psi}^{-1} + \mathbf{\Sigma}_0^{-1})^{-1}. \tag{34}$$

The conditional posterior distribution for $\mathbf{\Psi}$ is an inverse Wishart distribution

$$p(\mathbf{\Psi}|\boldsymbol{\beta}, \boldsymbol{\eta}_i, i = 1, \ldots, N) = IW(m_1, V_1) \tag{35}$$

where

$$m_1 = m_0 + N \tag{36}$$

and

$$V_1 = V_0 + \sum_{i=1}^{N}(\boldsymbol{\eta}_i - \boldsymbol{\beta})(\boldsymbol{\eta}_i - \boldsymbol{\beta})'. \tag{37}$$

The conditional posterior distribution for $\mathbf{\Phi}$ is an inverse Wishart distribution

$$p(\mathbf{\Phi}|\mathbf{\Lambda}, \eta_i, \mathbf{y}_i, w_i, i = 1, \ldots, N) \quad = \quad IW(n_1, W_1)$$

where

$$n_1 = n_0 + N \tag{38}$$

and

$$W_1 = W_0 + \sum_{i=1}^{n} w_i(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)(\mathbf{y}_i - \mathbf{\Lambda}\boldsymbol{\eta}_i)'. \tag{39}$$

If $\mathbf{\Phi} = \mathbf{I}_{T\times T}\sigma_e^2$, the conditional posterior distribution for σ_e^2 is an inverse Gamma distribution

$$p(\sigma_e^2|\mathbf{\Lambda}, \eta_i, \mathbf{y}_i, w_i, i = 1, \ldots, N) = IG(c_1, d_1) \tag{40}$$

where

$$c_1 = c_0 + \frac{NT}{2} \tag{41}$$

and

$$d_1 = d_0 + \sum_{i=1}^{n} w_i(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)/2. \tag{42}$$

The conditional posterior distribution for $\boldsymbol{\eta}_i$ is a multivariate normal distribution

$$p(\boldsymbol{\eta}_i|\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \boldsymbol{\Lambda}, w_i, \mathbf{y}_i) = MN_q(\boldsymbol{\mu}_{\boldsymbol{\eta}_i}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_i}) \tag{43}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\eta}_i} = \boldsymbol{\Sigma}_{\boldsymbol{\eta}_i}\left[\boldsymbol{\Psi}^{-1}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\left(\frac{\boldsymbol{\Phi}}{w_i}\right)^{-1}\mathbf{y}_i\right] \tag{44}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}_i} = \left[\boldsymbol{\Psi}^{-1} + \boldsymbol{\Lambda}'\left(\frac{\boldsymbol{\Phi}}{w_i}\right)^{-1}\boldsymbol{\Lambda}\right]^{-1}. \tag{45}$$

The conditional distribution for w_i is a Gamma distribution

$$p(w_i|k, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \mathbf{y}_i, \boldsymbol{\eta}_i) = G(\frac{k+T}{2}, k^*) \tag{46}$$

where

$$k^* = \left\{\frac{2}{k} + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Phi}^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'\right]\right\}^{-1}. \tag{47}$$

The conditional posterior distribution of k does not have a standard form. The kernel for the conditional posterior distribution of k is

$$p(k|w_i, i = 1, \ldots, N) \propto U(a, b)(\prod_{i=1}^{N} w_i)^{k/2-1} \exp\left(-\frac{\sum_{i=1}^{N} w_i}{k/2}\right)\left[\Gamma(k/2)^{-N}\right](k/2)^{-Nk/2}. \tag{48}$$

For certain models such as the robust latent basis growth curve model, the conditional posterior distribution for the unknown parameters $\boldsymbol{\lambda}$ in $\boldsymbol{\Lambda}$ is

$$p(\boldsymbol{\lambda}|w_i, \boldsymbol{\eta}_i, \boldsymbol{\Phi}, i = 1, \ldots, N) \propto \exp\left[-(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)'\boldsymbol{\Sigma}_{\lambda}^{-1}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) - \sum_{i=1}^{N}\frac{w_i}{2}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)'\boldsymbol{\Phi}^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\right]. \tag{49}$$

This distribution may not have a standard form for robust growth curve models.

Gibbs sampling algorithm

Given the above conditional posterior distributions, the following Gibbs sampling algorithm can be implemented.

1. Start with initial values $\boldsymbol{\beta}^{(0)}, \boldsymbol{\Psi}^{(0)}, \boldsymbol{\Phi}^{(0)}, k^{(0)}, \boldsymbol{\eta}_i^{(0)}, w_i^{(0)}, \boldsymbol{\lambda}^{(0)}$.

2. Assume at the jth iteration, we have $\boldsymbol{\beta}^{(j)}, \boldsymbol{\Psi}^{(j)}, \boldsymbol{\Phi}^{(j)}, k^{(j)}, \boldsymbol{\eta}_i^{(j)}, w_i^{(j)}\boldsymbol{\lambda}^{(j)}$.

3. At the $(j+1)$th iteration,

   (a) Sample $\boldsymbol{\beta}^{(j+1)}$ from $p(\boldsymbol{\beta}|\boldsymbol{\Psi}^{(j)}, \boldsymbol{\eta}_i^{(j)}, i = 1, \ldots, N)$;

   (b) Sample $\boldsymbol{\Psi}^{(j+1)}$ from $p(\boldsymbol{\Psi}|\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\eta}_i^{(j)}, i = 1, \ldots, N)$;

(c) Sample $\mathbf{\Phi}^{(j+1)}$ from $p(\mathbf{\Phi}|\mathbf{\Lambda}(\boldsymbol{\lambda}^{(j)}), \eta_i^{(j)}, \mathbf{y}_i, w_i^{(j)}, i = 1, \ldots, n)$ or sample $\sigma_e^{2(j+1)}$ from $p(\sigma_e^2|\mathbf{\Lambda}(\boldsymbol{\lambda}^{(j)}), \eta_i^{(j)}, \mathbf{y}_i, w_i^{(j)}, i = 1, \ldots, n)$;

(d) Sample $k^{(j+1)}$ from $p(k|w_i^{(j)}, i = 1, \ldots, n)$;

(e) Sample $\boldsymbol{\eta}_i^{(j+1)}, i = 1, \ldots, N$ from $p(\boldsymbol{\eta}_i|\mathbf{\Phi}^{(j+1)}, \mathbf{\Psi}^{(j+1)}, \boldsymbol{\beta}^{(j+1)}, \mathbf{\Lambda}^{(j)}(\boldsymbol{\lambda}^{(j)}), w_i^{(j)}, \mathbf{y}_i)$;

(f) Sample $w_i^{(j+1)}, i = 1, \ldots, N$ from $p(w_i|k,^{(j+1)} \mathbf{\Phi}^{(j+1)}, \mathbf{\Lambda}^{(j)}(\boldsymbol{\lambda}^{(j)}), \mathbf{y}_i, \boldsymbol{\eta}_i^{(j+1)})$

(g) Sample $\boldsymbol{\lambda}^{(j+1)}$ from $p(\boldsymbol{\lambda}|w_i^{(j+1)}, \boldsymbol{\eta}_i^{(j+1)}, \mathbf{\Phi}^{(j+1)}, i = 1, \ldots, N)$.

In the algorithm, $\mathbf{\Lambda}(\boldsymbol{\lambda})$ denote the factor loading matrix after plugging the samples for the unknown parameters $\boldsymbol{\lambda}$. The Gibbs sampling algorithm can be programmed using C++, R or other programming languages. For parameters with standard forms, they can be sampled from directly. For parameters without standard forms, the Metropolis-Hastings algorithm (Robert & Casella, 2004) can be applied in which the prior distributions can be used as the proposal distributions.

Appendix B
Illustration of the software

We use the robust latent basis growth curve model as an example to illustrate the use of the software. All figures used here are screens captured directly from an Internet browser. To use the software, go to this web address – http://webstats.psychstat.org/semrgcm/ – in a web browser. On can select the model to be used and then follow the on-screen direction as shown in this example. In the first step (Figure B1), basic information related to the data and model can be provided such as the sample size, the number of measurement occasions, the number of covariates in the model, and the missing data indicator. By default, there is no covariate in the model and a missing datum is indicated by `99999`. Also by default, the degrees of freedom will be estimated. If one would like to fix the degrees of freedom, one can supply it as a non-zero positive number. A data file can be specified by clicking on the `Choose File` button to set the path to the data file on users' local computer. Data should be in free format and organized according to the order of occasion 1, occasion 2, ..., covariate 1, covariate 2, ... By clicking on the `Next` button, we move on to the next screen.

In the second step (Figure B2), the prior for each model parameter can be specified. The hyperparameters for each prior distribution can be input. For example, the default prior for the intercept β_L is a normal distribution with mean 0 and variance 10^6, representing a non-informative prior. Informative priors can also be supplied. If, based on previous research, one believes that β_L has an average of 6 and its variation is about 1, one can input 6 in the field of `Mean` and 1 in the field of `Variance`.

In the third step (Figure B3), the starting values for model parameters can be supplied. In the fourth step (Figure B4), the control parameters for MCMC can be specified. The default `Burn-in` period, the number of iterations to be discarded, is 1,000. This number should be increased for complex models or analysis involving missing data. The `Length of Markov Chain` is the number of iterations used for Bayesian parameter estimation. If `Thinning` is larger than 1, the iterations will only be saved every m (=`Thinning`) iterations. DIC, EAIC and EBIC for the model will be calculated if the option `DIC` is checked. Leaving `DIC` unchecked can marginally speed up the analysis. Furthermore, if one checks the option `Random effects`, the random effects $\boldsymbol{\eta}$ will be estimated and the predicted individual growth trajectories will be plotted.

The results for this analysis are given on the next screen and only the header of the output is presented in Figure B5 for the sake of space. By clicking on the links in the header, one can go to each section of the output. The results include the summary statistics as shown in Table 1, the spaghetti plot or the longitudinal plot, the boxplot as shown in Figure 1, the Bayesian parameter estimates as in Table 3, the DIC, EAIC and EBIC as in Table 10, and the history plot of Markov chain for each model parameter as in Figure 2. In

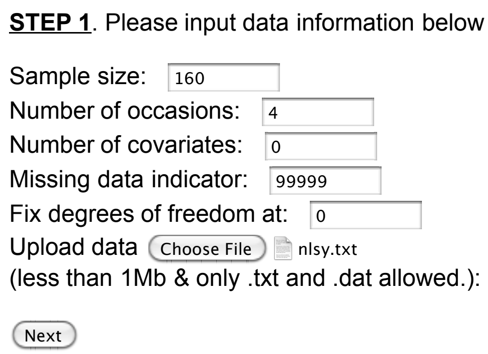

Figure B1 Step 1: Basic data and model information

addition, the autocorrelation plot for each Markov chain is also provided. If the `Random effects` option is checked, the output also includes the Bayesian estimates and credible interval of random effects and the plot of the predicted growth trajectories.

STEP 2. Specify priors for model parameters

Normal Prior $N(\mu, \sigma^2)$ for Intercept β_L :

- Mean 0
- Variance 1000000

Normal Prior $N(\mu, \sigma^2)$ for Slope β_S:

- Mean 0
- Variance 1000000

Normal Prior $N(\mu, \sigma^2)$ for basis coefficient $A[2]$:

- Mean 0
- Variance 1000000

Normal Prior $N(\mu, \sigma^2)$ for basis coefficient $A[3]$:

- Mean 0
- Variance 1000000

Inverse Wishart Prior $IW(m_0, V_0)$ for Covariance matrix Ψ :

- Scale matrix 1 0 0 1
- Degrees of freedom 2

Inverse Gamma Prior $IG(\alpha, \beta)$ for Error Variance (Scale parameter) σ_e^2:

- Shape 0.001
- Scale 0.001

Uniform Piror $U[a, b]$ for Degrees of Freedom k:

- Minimum 0
- Maximum 100

Next

Figure B2 Step 2: Priors

STEP 3. Starting values for model parameters

Mean of Intercept β_L: 4.65

Mean of Slope β_S: 1.51

Basis coefficients $A[2]$: 0

Basis coefficients $A[3]$: 0

Covariance Matrix of Intercept and Slope Ψ: 1 0 0 1

Error Variance (Scale parameter) σ_e^2: 1

Degrees of Freedom k: 5

Next

Figure B3 Step 3: Starting values

STEP 4. Control MCMC

Burn-in: 1000

Length of Markov Chain: 10000

Thinning: 1

DIC?

Random effects η?

Next

Figure B4 Step 4: Controlling MCMC

Results

Summary statistics | Spaghetti plot | Boxplot | Bayesian estimates | History plot | Autocorrelation plot | Go to bottom

The program started to run at 11:17:50 on Dec 06, 2010 .

Figure B5 Step 5: Output (only the header is presented here)