

This article was downloaded by: [University of Notre Dame]

On: 30 April 2013, At: 05:25

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Applied Statistics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cjas20>

Bayesian growth curve models with the generalized error distribution

Zhiyong Zhang^a

^a Department of Psychology , University of Notre Dame , 118B Haggar Hall, Notre Dame , IN , 46556 , USA

Published online: 30 Apr 2013.

To cite this article: Zhiyong Zhang (2013): Bayesian growth curve models with the generalized error distribution, Journal of Applied Statistics, DOI:10.1080/02664763.2013.796348

To link to this article: <http://dx.doi.org/10.1080/02664763.2013.796348>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Bayesian growth curve models with the generalized error distribution

Zhiyong Zhang*

Department of Psychology, University of Notre Dame, 118B Hagggar Hall, Notre Dame, IN 46556, USA

(Received 8 May 2012; accepted 12 April 2013)

To deal with the longitudinal data with both leptokurtic and platykurtic errors, we extend growth curve models using the generalized error distribution (GED) model. The Metropolis–Hastings algorithm is used to estimate the GED model parameters in the Bayesian framework. The application of the GED model is illustrated through the analysis of mathematical development data. Results show that the GED model can correctly identify the deviation from normal of the error distributions.

Keywords: growth curve modeling; generalized error distribution; Bayesian estimation; longitudinal data analysis; robust statistics

1. Introduction

Growth curve models have been widely used in analyzing developmental processes through longitudinal data collection in social and behavioral research [11,14,15]. The path diagram for a typical growth curve model is portrayed in Figure 1. The observed variables are drawn as squares, the unobserved or latent variables are drawn as circles, and the constant is represented by a triangle. The squares labeled y_{i1} through y_{iT} are the observed data at time 1 through T , respectively, for the i th individual. L_i in the circle is the latent intercept of the i th individual, β_L the mean of the latent intercept across all the participants, and σ_L^2 the variability around the intercept representing inter-individual differences in the latent intercept. S_i in the circle corresponds to the slope, β_S is the mean of the slope across all the participants, and σ_S^2 is the variance, i.e. the individual differences around the mean slope. S_i can often be interpreted as the rate of change/growth. The covariance σ_{LS} between the intercept and slope is represented by the double headed arrow between the latent intercept and slope variables. The circles labeled e_{i1} through e_{iT} are random errors added at each measurement occasion with the error variance σ_e^2 .

*Email: ZhiyongZhang@nd.edu

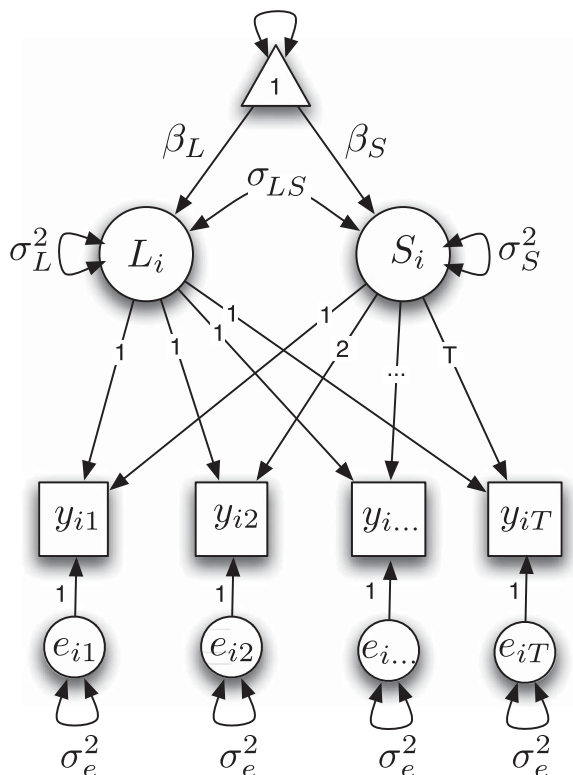


Figure 1. Path diagram for a growth curve model.

In mathematical notation, the growth curve model in Figure 1 can be expressed as

$$\begin{aligned}
 y_{it} &= L_i + S_i \alpha_t + e_{it}, \\
 L_i &= \beta_L + v_{1i}, \\
 S_i &= \beta_S + v_{2i},
 \end{aligned} \tag{1}$$

where v_{1i} is the deviation of L_i from β_L and v_{2i} is the deviation of S_i from β_S . Given the notations in the path diagram, $\text{Var}(v_{1i}) = \sigma_L^2$, $\text{Var}(v_{2i}) = \sigma_S^2$, and $\text{Cov}(v_{1i}, v_{2i}) = \sigma_{LS}$. Different growth curve shapes can be produced by adjusting the loadings $\alpha_t, t = 1, \dots, T$. For example, if $\alpha_t = t - 1$, the model becomes a linear growth curve model where L_i represents the initial level and S_i is the rate of growth. If $\alpha_t = (t - 1)/(T - 1)$, the model is still a linear growth curve model while S_i denotes the total change from time 1 to time T . One can also obtain the latent basis growth curve model by letting $\alpha_1 = 0$ (sometimes, also $\alpha_T = 1$) and freely estimating the other coefficients.

For growth curve analysis, it is typically assumed that e_{it} is normally distributed with mean 0 and variance σ_e^2 . However, in reality, data are more likely to be non-normally distributed than normally distributed, especially for social and behavioral research. For example, Micceri [17] examined 440 data sets on achievement and psychometric measures and found all to be significantly non-normal at the α 0.01 significance level. The violation of the normal assumption may result in unreliable parameter estimates, incorrect standard errors and confidence intervals, and misleading statistical tests and inference [13,26,30]. To deal with non-normal data, robust growth curve analysis has been proposed. For example, Pendergast and Broffitt [19] discussed two robust estimators and their asymptotic properties for growth curve models. Singer and Sen [22] proposed an M-method

to obtain the parameter estimates for growth curve models by transforming them into standard multivariate linear models. Silvapulle [21] further extended the M-method to allow asymmetric measurement errors for growth curve analysis. Recently, Zhang *et al.* [29] proposed to use the t -distribution to model heavy-tailed data and contaminated normal data with outliers for growth curve analysis and argued that the parametric form of the robust growth curve modeling has many advantages.

However, the previous non-normal data analysis methods have focused on leptokurtic data with positive excess kurtosis that is larger than normal data, for example, heavy-tailed data or outliers. Yuan *et al.* [27] showed that both positive and negative excess kurtosis influenced statistical inference. Furthermore, in longitudinal data analysis, it is likely to show both positive and negative excess kurtosis. For example, at the beginning of a longitudinal study, the participants may show more similarity and therefore data may demonstrate negative excess kurtosis in the early phase of data collection. However, more diversity may develop over time, therefore, data could show positive excess kurtosis in the late phase of data collection.

In this study, we propose to model the error e_{it} in growth curve models using the generalized error distribution (GED) to account for positive, negative, and zero excess kurtosis. In the following, we first introduce the GED. Then, we present the extension of growth curve models with the GED. To estimate the models, the Bayesian method is proposed. After that, we illustrate the application of growth curve models with the GED through the analysis of a set of mathematical development data.

2. Generalized error distribution

The GED, also called the exponential power distribution or generalized normal distribution, can model the error distribution more flexibly than the normal distribution [2,18]. The GED comes with different forms. In this study, the form by Box and Tiao [2, p. 157] is adopted. If Y follows a GED, its density function $\text{GED}(\mu, \sigma, \gamma)$ can be expressed as

$$f(y) = \omega(\gamma)\sigma^{-1} \exp \left[-c(\gamma) \left| \frac{y - \mu}{\sigma} \right|^{2/(1+\gamma)} \right],$$

where

$$\omega(\gamma) = \frac{\{\Gamma[3(1 + \gamma)/2]\}^{1/2}}{(1 + \gamma)\{\Gamma[(1 + \gamma)/2]\}^{3/2}} \quad (2)$$

and

$$c(\gamma) = \left\{ \frac{\Gamma[3(1 + \gamma)/2]}{\Gamma[(1 + \gamma)/2]} \right\}^{1/(1+\gamma)}. \quad (3)$$

Here y ranges from $-\infty$ to $+\infty$, μ is a location parameter, and σ is a scale parameter. The advantage of this specification is that the scale parameter σ is also the standard deviation (SD) of the population. γ is a shape parameter that is related to the kurtosis of the distribution that characterizes the ‘non-normality’ of the population. The mean and variance of the GED are

$$\begin{aligned} E(y) &= \mu, \\ \text{Var}(y) &= \sigma^2. \end{aligned}$$

The GED is symmetric and therefore its skewness is 0. Its kurtosis is

$$g_2 = \frac{\Gamma[5(1 + \gamma)/2]\Gamma[(1 + \gamma)/2]}{\Gamma[3(1 + \gamma)/2]^2}.$$

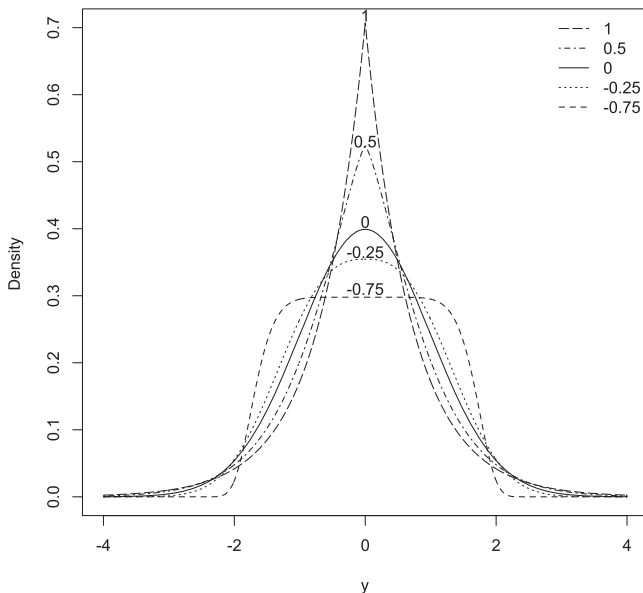


Figure 2. The GED with $\mu = 0$, $\sigma = 1$, and different values of γ . Each value of γ is placed on the top of the density plot. The corresponding kurtosis for $\gamma = -0.75, -0.25, 0, 0.5, 1$ is 1.92, 2.55, 3, 4.22, 6.

From the GED, we can derive many other distributions. For example, if $\gamma = 0$, the GED becomes the normal distribution. If $\gamma = 1$, it becomes the double exponential distribution. Furthermore, when γ approaches -1 , it becomes a uniform distribution.

To better investigate the influence of the shape parameter γ , we plot the density function of the GED with different γ values in Figure 2. The five distributions in the figure share the same location parameter $\mu = 0$ and scale parameter $\sigma = 1$. γ takes one of the five values $-0.75, -0.25, 0, 0.5$, and 1 . When $\gamma = 0$, the distribution becomes a normal distribution. When $\gamma > 0$, the GED has a fatter tail than the normal distribution and when $\gamma < 0$, it has a thinner tail than the normal distribution. Therefore, the GED can deal with data with both fat and thin tails. For real data, γ can be estimated as an unknown parameter [2].

3. Growth curve models with the GED

To extend the growth curve models with normal error for dealing with data with positive or negative excess kurtosis, we propose to replace the normal distribution for the error term e_{it} with the GED. Specifically,

$$e_{it} \sim \text{GED}(0, \sigma_e, \gamma_t). \quad (4)$$

Note that the error is assumed to have a mean 0 and variance σ_e^2 as for the regular growth curve models. For generalization, the shape parameter γ_t is allowed to vary across time so that the error can have different kurtosis at different time. However, one can fix the distribution to be the same across time if necessary. As in the normal growth curve models, it is still assumed that

$$\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} \sim \text{BN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix} \right], \quad (5)$$

where BN represents a bivariate normal distribution. Equations (1), (4), and (5) together define a growth curve model with the GED. In this study, we refer to the model with the normally

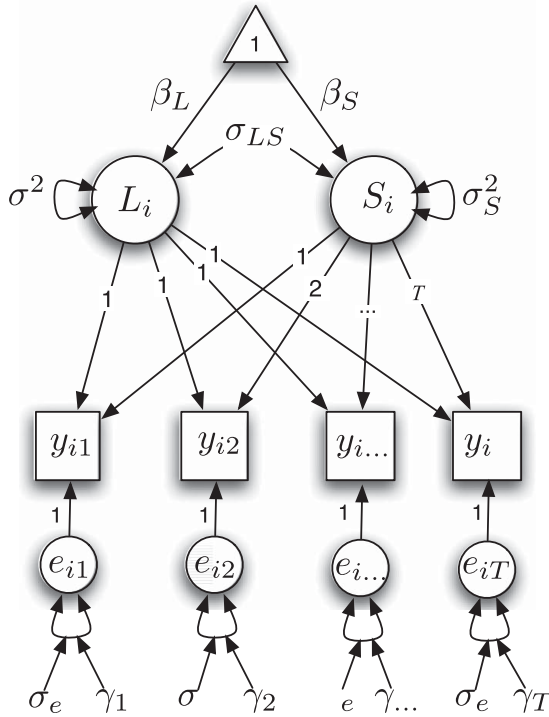


Figure 3. Path diagram for a GED growth curve model.

distributed errors as the normal growth curve model or the normal model and refer to the model with the errors following the GED as the GED growth curve model or the GED model.

As for the normal growth curve models, we can also portray a path diagram for the GED growth curve models. Figure 3 displays such a path diagram. In the diagram, the error distribution is influenced by two parameters, the shape parameter γ_t and the scale parameter σ_e .

3.1 Model estimation

To estimate the parameters in a GED growth curve model, Bayesian method can be employed. The parameters in a typical GED growth curve model include $\beta_L, \beta_S, \sigma_L^2, \sigma_{LS}, \sigma_S^2, \sigma_e$, and γ_t . We refer to γ_t as shape parameters and the rest of the parameters as the growth curve parameters. For convenience, let $\mathbf{b}_i = (L_i, S_i)'$, $\mathbf{x}_t = (1, \alpha_t)$, $\boldsymbol{\beta} = (\beta_L, \beta_S)'$, $\mathbf{v}_i = (v_{1i}, v_{2i})'$, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix}.$$

With such notations, the growth curve model can be written as

$$y_{it} = \mathbf{x}_t \mathbf{b}_i + e_{it},$$

$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{v}_i.$$

Because of the involvement of the random effects \mathbf{b}_i , the data augmentation algorithm [24] is used to augment the observed data y_{it} and the random effects. Therefore, the joint posterior

distribution of the random effects \mathbf{b}_i and model parameters is

$$\begin{aligned}
 & p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t, \mathbf{b}_i | y_{it}, x_t, i = 1, \dots, N, t = 1, \dots, T) \\
 &= K p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t) \prod_{i=1}^N \left\{ p(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}) \prod_{t=1}^T [p(y_{it} | \mathbf{b}_i, x_t, \sigma_e, \gamma_t)] \right\} \\
 &= K p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t) \prod_{i=1}^N \left\{ \frac{1}{2\pi |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{b}_i - \boldsymbol{\beta}) \right] \right\} \\
 &\quad \times \prod_{i=1}^N \prod_{t=1}^T \left\{ \omega(\gamma_t) \sigma_e^{-1} \exp \left[-c(\gamma_t) \left| \frac{y_{it} - \mathbf{x}_t \mathbf{b}_i}{\sigma_e} \right|^{2/(1+\gamma_t)} \right] \right\} \\
 &= K p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t) (2\pi)^{-N/2} |\boldsymbol{\Sigma}|^{-N/2} \exp \left[-\frac{1}{2} \sum_{i=1}^N (\mathbf{b}_i - \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{b}_i - \boldsymbol{\beta}) \right] \\
 &\quad \times \sigma_e^{-NT} \prod_{t=1}^T [\omega(\gamma_t)]^N \exp \left[-\sum_{t=1}^T c(\gamma_t) \sum_{i=1}^N \left| \frac{y_{it} - \mathbf{x}_t \mathbf{b}_i}{\sigma_e} \right|^{2/(1+\gamma_t)} \right], \tag{6}
 \end{aligned}$$

where $\omega(\gamma)$ and $c(\gamma)$ are defined in Equations (2) and (3) and K is a normalization constant. $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t)$ is the joint prior distribution for model parameters.

3.1.1 Priors

To obtain the posterior distribution, we need to specify priors for model parameters. In this study, the independent priors are used for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, σ_e , and γ_t so that $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma_e, \gamma_t) = p(\boldsymbol{\beta})p(\boldsymbol{\Sigma})p(\sigma_e) \prod_{t=1}^T p(\gamma_t)$. For $\boldsymbol{\beta}$, a bivariate normal prior $\text{BN}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$ is used

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

For $\boldsymbol{\Sigma}$, an inverse Wishart prior distribution $\text{IW}(m_0, V_0)$ is used

$$p(\boldsymbol{\Sigma}) = \frac{|V_0|^{m_0/2} |\boldsymbol{\Sigma}|^{-(m_0+q+1)/2} \exp[-\text{tr}(V_0 \boldsymbol{\Sigma}^{-1})/2]}{2^{m_0 q/2} \Gamma_q(m_0/2)},$$

where $q = 2$ for the current study. For σ_e , an inverse Gamma distribution $\text{IG}(c_0, d_0)$ is used

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

For γ_t , a uniform distribution $\text{U}(a_0, b_0)$ is used

$$p(\gamma_t) = \frac{1}{b_0 - a_0}.$$

3.1.2 Posterior

Incorporating the priors above to Equation (6), the resulting posterior is readily available but in a very complex format. To evaluate the posterior, we use the Metropolis–Hastings (M–H) algorithm within Gibbs sampling [4,5,16,20]. Gibbs sampling can generate random numbers for a complex joint distribution by iteratively sampling from a full set of conditional distributions.

When a conditional distribution does not have a known form such as normal distribution, the M–H algorithm can then be used. The M–H algorithm first generates a random number from a distribution with relatively simple form, often called a proposal distribution. Then, based on certain criterion, the generated random number is kept with certain probability. After convergence, the generated random number can be viewed as a sample from the joint distribution.

To implement the M–H algorithm within Gibbs sampling for the GED growth curve models, we first divide the model parameters into different blocks and obtain the conditional distribution for each block. Then, we sample each block of parameter successively. Specifically, the full set of parameters is divided into the following blocks – β , Σ , σ_e , γ_t , and \mathbf{b}_i , $t = 1, \dots, T$; $i = 1, \dots, N$. Specifically, the following algorithm can be used.

- (1) Initial values. Start with $\mathbf{b}_i^{(0)}$, $\Sigma^{(0)}$, $\sigma_e^{(0)}$, $\gamma_t^{(0)}$, $i = 1, \dots, N$, $t = 1, \dots, T$. The starting values can be generated in the following way:
 - (a) Fit the regression model $y_{it} = L_i + S_i\alpha_t + e_{it}$ to each individual i to obtain \hat{L}_i , \hat{S}_i , and \hat{e}_{it} .
 - (b) Set $\mathbf{b}_i^{(0)} = (\hat{L}_i, \hat{S}_i)'$ and $\Sigma^{(0)} = (1/N) \sum_{i=1}^N (\mathbf{b}_i^{(0)} - \overline{\mathbf{b}_i^{(0)}})' (\mathbf{b}_i^{(0)} - \overline{\mathbf{b}_i^{(0)}})$ with $\overline{\mathbf{b}_i^{(0)}} = (1/N) \sum_{i=1}^N \mathbf{b}_i^{(0)}$.
 - (c) Set $\sigma_e^{(0)} = \sqrt{(1/NT) \sum_{t=1}^T \sum_{i=1}^N (\hat{e}_{it} - (1/NT) \sum_{t=1}^T \sum_{i=1}^N \hat{e}_{it})^2}$.
 - (d) Set $\gamma_t^{(0)} = 0$ starting as if e_{it} were normally distributed.
- (2) Generate the random number $\beta^{(1)}$ from the bivariate normal distribution $\text{BN}(\beta_1, \Sigma_1)$ with

$$\beta_1 = \Sigma_1 \left[(\Sigma^{(0)})^{-1} \sum_{i=1}^N \mathbf{b}_i^{(0)} + \Sigma_0^{-1} \beta_0 \right]$$

and

$$\Sigma_1 = [N(\Sigma^{(0)})^{-1} + \Sigma_0^{-1}]^{-1}.$$

- (3) Generate the random number $\Sigma^{(1)}$ from the inverse Wishart distribution $\text{IW}(m_0, V_0)$ with

$$m_1 = m_0 + N$$

and

$$V_1 = V_0 + \sum_{i=1}^N (\mathbf{b}_i^{(0)} - \beta^{(1)}) (\mathbf{b}_i^{(0)} - \beta^{(1)})'.$$

- (4) Generate the random number $\mathbf{b}_i^{(1)}$, $i = 1, \dots, N$ from the distribution

$$p(\mathbf{b}_i) = K_1 \exp \left[-\frac{1}{2} (\mathbf{b}_i - \beta^{(1)})' (\Sigma^{(1)})^{-1} (\mathbf{b}_i - \beta^{(1)}) - \sum_{t=1}^T c(\gamma_t^{(0)}) \left| \frac{y_{it} - \mathbf{x}_t \mathbf{b}_i}{\sigma_e^{(0)}} \right|^{2/(1+\gamma_t^{(0)})} \right]$$

with K_1 denoting a normalization constant. The distribution of \mathbf{b}_i is not a known one. To generate the random number from it, the M–H algorithm can be used.

- (a) Generate \mathbf{b}_i^* from the bivariate normal proposal distribution $q(\mathbf{b}_i) = \text{BN}(\mathbf{b}_i^{(0)}, \Sigma^{(0)})$.
- (b) Compute the probability

$$P = \min \left(1, \frac{p(\mathbf{b}_i^*)}{p(\mathbf{b}_i^{(0)})} \frac{q(\mathbf{b}_i^{(0)})}{q(\mathbf{b}_i^*)} \right).$$

(c) Take

$$\mathbf{b}_i^{(1)} = \begin{cases} \mathbf{b}^* & \text{with probability } P, \\ \mathbf{b}_i^{(0)} & \text{with probability } 1 - P. \end{cases}$$

(5) Generate the random number $\sigma_e^{(1)}$ from the distribution

$$p(\sigma_e) = K_2 \sigma_e^{-NT-c_0-1} \exp \left[-\frac{d_0}{\sigma_e} - \sum_{t=1}^T c(\gamma_t^{(0)}) \sum_{i=1}^N \left| \frac{y_{it} - \mathbf{x}_t \mathbf{b}_i^{(1)}}{\sigma_e} \right|^{2/(1+\gamma_t^{(0)})} \right]$$

with K_2 denoting a normalization constant. σ_e can also be generated through the M–H algorithm. The inverse Gamma $\text{IG}(c_0^*, d_0^*)$ distribution can be used as the proposal distribution with

$$c_0^* = \frac{\hat{\mu}^2}{\hat{\sigma}^2} + 2, \\ d_0^* = (c_0^* - 1) * \hat{\mu},$$

where $\hat{\mu}$ and $\hat{\sigma}$ are estimates of σ_t with normally distributed errors.

(6) Generate the random number $\gamma_t^{(1)}$ from the distribution

$$p(\gamma_t) = K_3 [\omega(\gamma_t)]^N \exp \left[-c(\gamma_t) \sum_{i=1}^N \left| \frac{y_{it} - \mathbf{x}_t \mathbf{b}_i}{\sigma_e^{(1)}} \right|^{2/(1+\gamma_t)} \right]$$

with K_3 denoting a normalization constant. To generate γ_t , we use the uniform distribution as the proposal distribution to implement the M–H algorithm.

(7) Repeat Steps 2–7 to get a Markov chain for each model parameter.

3.2 Convergence diagnostics

There are two practical issues in applying Markov chain Monte Carlo (MCMC) methods. The first is to diagnose the convergence of Markov chains. The second is to decide the length of Markov chains. Brooks and Roberts [3] and Cowles and Carlin [7] discussed many different methods for testing convergence. In this study, we focus on visually inspecting the trace plot of Markov chains and the Geweke test [10]. The Geweke test calculates a statistic that is normally distributed. For a converged Markov chain, the absolute value of the Geweke statistic should be smaller than 1.96 at 0.05 α level. By nature, Markov chains always have autocorrelation. For two Markov chains, the one with higher autocorrelation provides less information about the posterior distribution than the one with smaller autocorrelation. In other words, a longer Markov chain is needed to accurately describe a posterior if the autocorrelation is higher. To characterize the information in a Markov chain, we use the statistic called effective sample size (ESS). The ESS is the equivalent sample size as if there were no autocorrelation. For two Markov chains with the same length, the one with a bigger ESS provides more information. A practical rule of thumb is to get an ESS at least 400.

Figure 4 presents trace plots for four Markov chains based on which convergence can be judged. Figure 4(a) displays a ‘well-behaved’ Markov chain. The Markov chain fluctuates around its center with similar variation. The Geweke statistics is 0.17, also indicating the convergence of the Markov chain. Furthermore, the ESS is 4668, close to the total number of iterations, 5000. This indicates that the autocorrelation for this Markov chain is low.

The Markov chain in Figure 4(b) seems to converge well based on the trace plot and the Geweke statistic (−1.82). However, the ESS is only 122 indicating high autocorrelation. Therefore, a longer

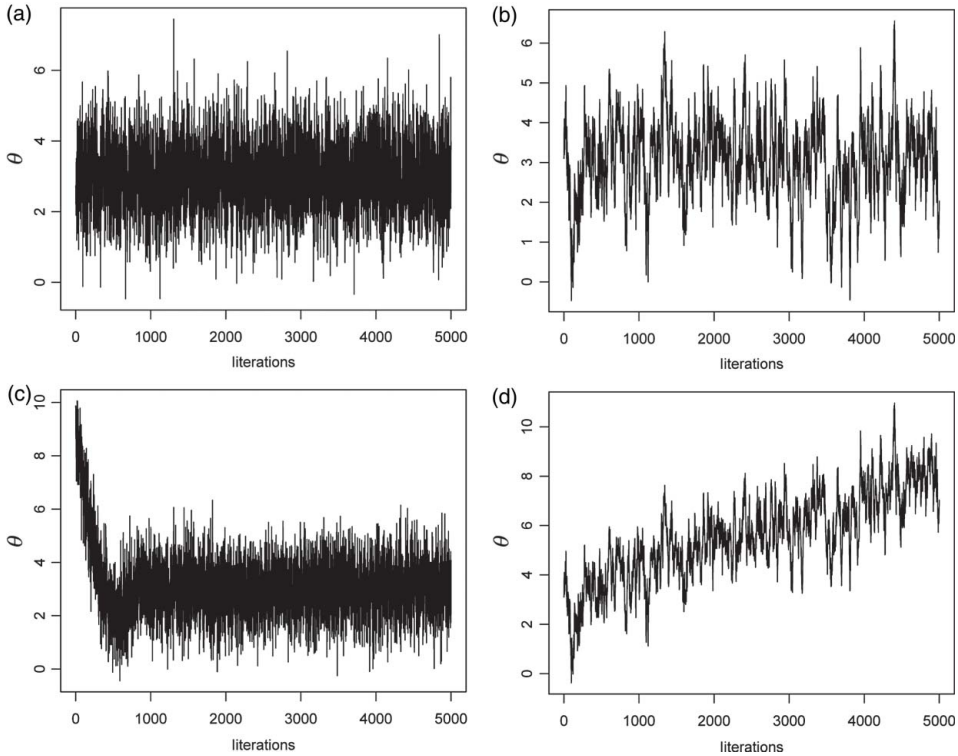


Figure 4. Illustration of the convergence of Markov chains: (a) convergence with low autocorrelation, Geweke = 0.17 and ESS = 4668. (b) Convergence with high autocorrelation, Geweke = -1.82 and ESS = 122. (c) Convergence with low autocorrelation after 1000 iterations, Geweke = 6.39 for the whole Markov chain, Geweke = 0.58 and ESS = 3512 for the kept iterations. (d) Non-convergence, Geweke = -16.36 .

Markov chain is needed if one wants to make inference of the posterior from which the Markov chain is generated.

There appear two parts in the Markov chain in Figure 4(c). There is a decline trend in the first part before around the 1000th iteration. The second part seems very stable similar to the Markov chain in Figure 4(a). In this situation, the first part of the Markov chain is often discarded as burn-in. The Geweke statistic for the whole Markov chain is 6.39 indicating non-convergence of the chain. However, after discarding the burn-in period of the 1000 iterations, the Geweke statistic becomes 0.58. The ESS is 3512 indicating low autocorrelation for the kept part of the Markov chain. The kept Markov chain can be used for statistical inference.

Figure 4(d) portrays a bad-behaved Markov chain. First, there appears a growth trend. Second, it seems that the autocorrelation is high. Furthermore, the Geweke statistic is -16.38 , also indicating non-convergence. A Markov chain like this one cannot be used for further inference.

3.3 Evaluation of posterior

For a converged Markov chain with sufficient ESS, we can construct posterior mean, posterior SD, and highest posterior density (HPD) credible interval (CI) for each model parameter for inference. Let θ represent a single parameter in the vector of $\theta = (\beta, \Sigma, \sigma_e, \gamma_t)$. The generated Markov chain after burn-in period for θ is $\theta_i, i = 1, \dots, n$, where n is the number of iterations, also the length

of the kept Markov chain. Then, a point estimate for θ can be constructed by the sample mean of the Markov chain

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_i.$$

The SD of θ is given by

$$\text{SD}(\theta) = \frac{1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2.$$

CI's for θ can also be constructed based on the Markov chain. The most widely used CI's include the equal-tail CI and the HPD CI [9]. A $100(1 - \alpha)\%$ equal-tail CI is $[\theta^{\alpha/2}, \theta^{(1-\alpha)/2}]$, where the lower and upper bounds are the $100\alpha/2$ th and $100((1 - \alpha)/2)$ th percentiles of the Markov chain, respectively. The HPD CI is a CI that covers $100(1 - \alpha)\%$ region of the density formed by the Markov chain but at the same time has the smallest interval width. For symmetrical posteriors, the equal-tail CI is the same as the HPD CI. For non-symmetrical posteriors, the HPD CI's have smaller width than the equal-tail CI.

3.4 Model comparisons

For the analysis of empirical data, deviance information criterion (DIC) [23] can be used to compare and select models, for example, to compare the GED growth curve model and the normal growth curve model. 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

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D = D(\bar{\boldsymbol{\theta}}) + 2p_D,$$

where $\boldsymbol{\theta}$ represents a vector of all the unknown parameters in the model, $\overline{D(\boldsymbol{\theta})}$ is the posterior mean of $-2(\text{LogLikelihood function})$, and $D(\bar{\boldsymbol{\theta}})$ is the value of $-2(\text{LogLikelihood function})$ 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 ($D(\bar{\boldsymbol{\theta}})$). In other words,

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}).$$

The model with the minimum DIC will make the best short-term predictions and thus indicates the best model among evaluated ones [23].

4. Illustration

In this section, we use an example to illustrate the application of the growth curve models with the GED. The data used here are from a sample of 563 participants. Each participant was measured on their mathematical ability at four times. The sample means at each time are 4.62, 7.45, 9.00, and 11.95, and the longitudinal plot of the data suggests a linear growth pattern.

A linear growth curve model in Figure 5(a) with the normal error distribution is first fitted to the data to diagnose the error distribution. To estimate the model, the Bayesian estimation method discussed by Zhang *et al.* [28] is used. The linear growth curve model has six parameters – β_L ,

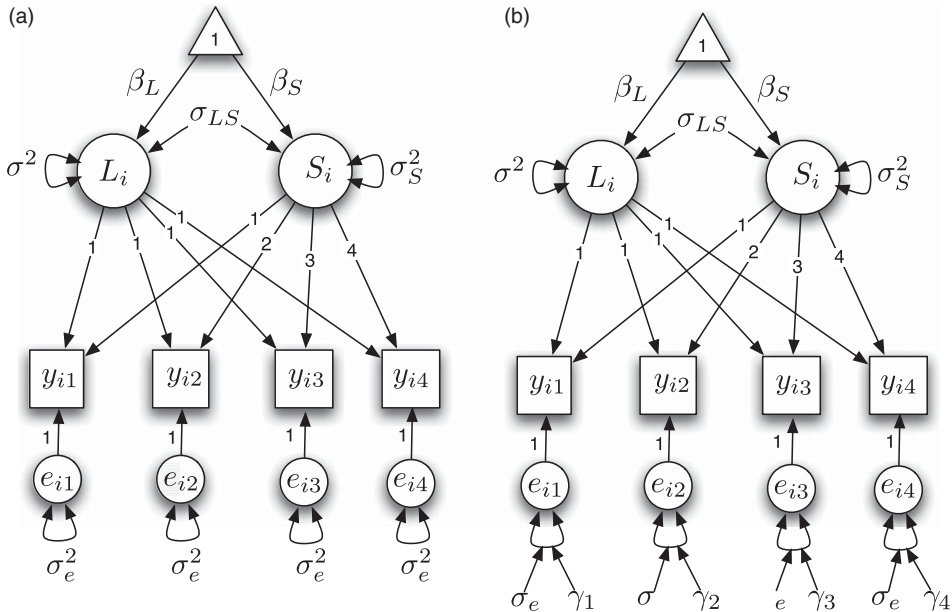


Figure 5. Linear growth curve models fitted to the four times of mathematical data: (a) normal model and (b) GED model.

$\beta_S, \sigma_L^2, \sigma_{LS}, \sigma_S^2,$ and σ_e^2 . To reduce the influence of priors, the following uninformative priors are used:

$$\begin{pmatrix} \beta_L \\ \beta_S \end{pmatrix} \sim \text{BN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0E+3 & 0 \\ 0 & 1.0E+3 \end{pmatrix} \right],$$

$$\begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix} \sim \text{IW} \left[2, \begin{pmatrix} 1.0 & 0 \\ 0 & 1.0 \end{pmatrix} \right],$$

$$\sigma_e^2 \sim \text{IG}(0.001, 0.001),$$

following Ref. [6]. The SAS and OpenBUGS scripts for estimating the model are given in Appendix 1.

Trace plots for the model parameters after a burn-in period of 5000 iterations are presented in Figure 6(a). The Markov chains for all parameters appeared to converge well. Furthermore, all Markov chains passed the Geweke test (see Table 1). The ESSs ranged from 749 to 7091. Therefore, the Markov chains can be used to make further inference.

For each parameter in the normal growth curve model, its estimate, SD, and CIs are given in Table 1. In order to check whether the normal error assumption is met or not, we also estimated the error e_{it} . Then, the D’Agostino test for skewness and the Anscombe–Glynn test for kurtosis are used to test the normality of e_{it} at each time [1,8]. The test results are given in Table 2. Based on the D’Agostino test, the distribution of e_{it} seemed to be symmetric. However, from the Anscombe–Glynn test, the kurtosis of the error at time 2 is significantly smaller (kurtosis = 2.65, p -value = 0.048) than 3 and the kurtosis at time 4 is significant larger (kurtosis = 3.79, p -value = 0.003) than 3. Overall, we conclude that the errors at time 1 and time 3 follow a normal distribution but the errors at the second and fourth waves do not. Because the errors show both leptokurtic and platykurtic, we allow the error distributions to vary by estimating the shape parameters at each time.

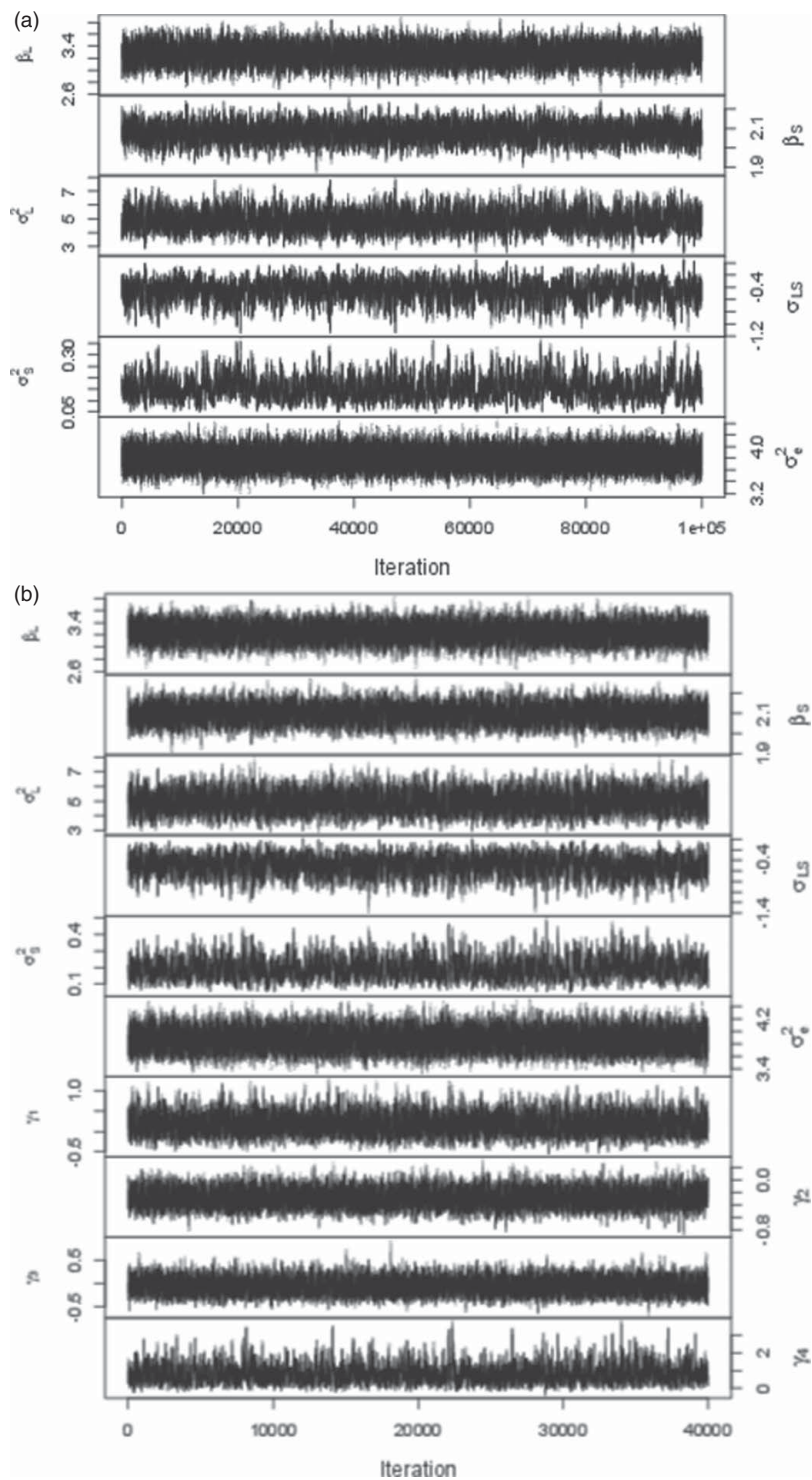


Figure 6. Trace plots of generated Markov chains: (a) normal model and (b) GED model.

Table 1. Parameter estimates for the normal growth curve model.

	Estimates	SD	Est/SD	Equal-tail CI		HPD CI		Geweke		ESS
				2.5%	97.5%	2.5%	97.5%	Statistic	p-Value	
β_L	2.367	0.149	15.921	2.076	2.663	2.072	2.658	0.079	0.937	7091
β_S	2.355	0.035	66.722	2.286	2.424	2.287	2.425	-0.239	0.811	3259
σ_L^2	8.883	0.761	11.675	7.461	10.439	7.450	10.417	-1.349	0.177	2812
σ_{LS}	-0.599	0.150	-3.983	-0.909	-0.322	-0.896	-0.312	1.188	0.235	1129
σ_S^2	0.218	0.043	5.106	0.139	0.306	0.135	0.301	-1.059	0.290	749
σ_e^2	2.411	0.100	24.182	2.222	2.614	2.217	2.608	1.295	0.195	3041

Note: Est/SD is the ratio of parameter estimate to its SD.

Table 2. Tests of kurtosis and skewness of errors from the normal growth curve model.

Time	D'Agostino skewness test			Anscombe-Glynn kurtosis test		
	Skewness	z	p-Value	Kurtosis	z	p-Value
1	-0.08	-0.49	0.625	3.19	1.02	0.309
2	0.18	1.18	0.237	2.65	-1.97	0.048
3	-0.18	-1.15	0.249	3.34	1.60	0.112
4	0.18	1.18	0.237	3.79	2.99	0.003

Note: For the D'Agostino skewness test, the null hypothesis is skewness = 0 and for the Anscombe-Glynn test, the null hypothesis is kurtosis = 3.

Table 3. Results from the GED growth curve model.

	Estimates	SD	Est/SD	Equal-tail CI		HPD CI		Geweke		ESS
				2.5%	97.5%	2.5%	97.5%	Statistic	p-Value	
β_L	2.354	0.148	15.934	2.063	2.645	2.064	2.646	-0.291	0.771	9791
β_S	2.364	0.035	66.966	2.295	2.434	2.295	2.433	0.011	0.992	3964
σ_L^2	8.911	0.762	11.702	7.494	10.479	7.434	10.412	0.715	0.475	3164
σ_{LS}	-0.654	0.156	-4.198	-0.972	-0.363	-0.962	-0.354	-0.770	0.441	1160
σ_S^2	0.242	0.047	5.106	0.154	0.339	0.150	0.335	0.950	0.342	672
σ_e^2	2.444	0.110	22.237	2.236	2.666	2.232	2.660	-1.011	0.312	2394
γ_1	0.599	0.490	1.224	-0.238	1.712	-0.289	1.658	-0.264	0.792	448
γ_2	-0.196	0.161	-1.220	-0.491	0.137	-0.500	0.126	1.802	0.072	1831
γ_3	0.091	0.164	0.555	-0.216	0.435	-0.238	0.406	0.937	0.349	2313
γ_4	0.649	0.317	2.048	0.104	1.360	0.078	1.321	-1.384	0.167	861

In order to deal with the violation of normality assumption, we fit a linear growth curve model with the GED to the data. The path diagram for the model is given in Figure 5(b). The SAS and OpenBUGS scripts for the model estimation are provided in Appendix 2.

Trace plots for the GED model parameters are presented in Figure 6(b). The Markov chains for all parameters appeared to converge well. Furthermore, all Markov chains passed the Geweke test (see Table 3). The ESSs ranged from 448 to 9791.

Table 4. DIC for the normal and GED growth curve models.

	Normal	GED
\bar{D}	8370	8266
\hat{D}	7675	7526
p_D	695	739
DIC	9065	9006

Parameter estimates for the GED model are given in Table 3. Based on the estimates of γ_t , the errors at time 1, 3, and 4 have kurtosis larger than normal (positive γ) while the error at time 2 has kurtosis smaller than normal (negative γ). Furthermore, based on the CIs, γ_4 is statistically significant indicating that the error distribution at time 4 cannot be treated as normal. Note the estimates for γ are consistent with the kurtosis calculated from the estimates of e_{it} . The parameters for β_L , β_S , σ_L^2 , σ_{LS} , and σ_S^2 can be compared for the normal model and GED model. Overall, the estimates are similar. However, the ratios of parameter estimates to their SDs are consistently smaller for the GED model. Therefore, the GED model can be considered as more efficient in general.

To compare the normal model and the GED model, we obtained their DICs (see Table 4). Based on p_D , the GED model is more complex than the normal model as expected. The GED model, however, has smaller DIC. Therefore, the GED model fits the data better than the normal model. This is not surprising, either, because the GED model can correctly model the error distributions that are not normal.

5. Discussion

Longitudinal study design has gradually dominated the research design in social and behavioral research. With the collection of longitudinal data, techniques for longitudinal data analysis also become more and more important. Growth curve models are one of the most widely used techniques for longitudinal data analysis. However, typical growth curve models assume that the error distributions are normal although in practice, social and behavioral data are often non-normal [17]. In order to deal with data with both leptokurtic and platykurtic errors, we proposed the growth curve models with the GED. Comparing to the normal growth curve model, the GED model can gain insight on the error distributions based on the estimation of the shape parameters of the GED. The normal growth curve model can be viewed as a special case of GED model where the shape parameters are 0.

In this study, we have focused on the discussion of the linear growth curve model with the GED error. However, the method developed can be readily applied to other types of growth curve models such as quadratic and exponential growth curve models [14]. Furthermore, because of the focus on the social and behavioral research, we have mainly studied growth curve models. However, the techniques used for growth curve models can be equally applied to mixed-effects models and multilevel models.

The M-H algorithm outlined in this study can be implemented in a variety of softwares. We have used both commercial software SAS and free open-source software OpenBUGS [12] to carry out our data analysis. Given the large user base of SAS in research institutes and the free availability of OpenBUGS, we foresee the future adoption and application of the developed GED growth curve models.

The study is not without limitation. First, the GED is symmetric. Therefore, for asymmetric data, one has to first transform the data to be symmetric, for example, using the method developed

by Tukey [25]. Second, our analysis has focused on the use of non-informative priors to reduce the influence of priors. Interested readers may look into the influence of the priors on the data analysis. Third, the current study only discussed model comparison but not model assessment. Future work should be carried out to evaluate the model fit for data analysis. Fourth, through practical data analysis, we showed that the GED model worked better than the normal model. In the future, simulation studies can be conducted to compare the two models systematically.

References

- [1] F.J. Anscombe and W.J. Glynn, *Distribution of kurtosis statistic for normal statistics*, *Biometrika* 70 (1983), pp. 227–234.
- [2] G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Wiley, New York, 1973.
- [3] S.P. Brooks and G.O. Roberts, *Assessing convergence of Markov chain Monte Carlo algorithms*, *Stat. Comput.* 8 (1998), pp. 319–335.
- [4] G. Casella and E.I. George, *Explaining the Gibbs sampler*, *Am. Stat.* 46 (1992), pp. 167–174.
- [5] S. Chib and E. Greenberg, *Understanding the Metropolis–Hastings algorithm*, *Am. Stat.* 49 (1995), pp. 327–335.
- [6] P. Congdon, *Applied Bayesian Modelling*, John Wiley & Sons, New York, 2003.
- [7] M.K. Cowles and B.P. Carlin, *Markov chain Monte Carlo convergence diagnostics: A comparative review*, *J. Am. Stat. Assoc.* 91 (1996), pp. 883–904.
- [8] R.B. D’Agostino, *Transformation to normality of the null distribution of g_1* , *Biometrika* 57 (1970), pp. 679–681.
- [9] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, London, 2003.
- [10] J. Geweke, *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*, in *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., Clarendon Press, Oxford, 1992, pp. 169–193.
- [11] N.M. Laird and J.H. Ware, *Random-effects models for longitudinal data*, *Biometrics* 38 (1982), pp. 963–974.
- [12] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best, *The BUGS project: Evolution, critique and future directions (with discussion)*, *Stat. Med.* 28 (2009), pp. 3049–3082.
- [13] R.A. Maronna, R.D. Martin, and V.J. Yohai, *Robust Statistics: Theory and Methods*, John Wiley & Sons, Inc., New York, 2006.
- [14] J.J. McArdle and J.R. Nesselrode, *Growth curve analysis in contemporary psychological research*, in *Comprehensive Handbook of Psychology: Research Methods in Psychology*, J. Schinka and W. Velicer, eds., Vol. 2, Wiley, New York, 2003, pp. 447–480.
- [15] W. Meredith and J. Tisak, *Latent curve analysis*, *Psychometrika* 55 (1990), pp. 107–122.
- [16] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equations of state calculations by fast computing machines*, *J. Chem. Phys.* 21 (1953), pp. 1087–1092.
- [17] T. Micceri, *The unicorn, the normal curve, and other improbable creatures*, *Psychol. Bull.* 105 (1989), pp. 156–166.
- [18] S. Nadarajah, *A generalized normal distribution*, *J. Appl. Stat.* 32 (2005), pp. 685–694.
- [19] J.F. Pendergast and J.D. Broffitt, *Robust estimation in growth curve models*, *Commun. Stat.: Theory Methods* 14 (1985), pp. 1919–1939.
- [20] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.
- [21] M.J. Silvapulle, *On m -methods in growth curve analysis with asymmetric errors*, *J. Stat. Plan. Inference* 32 (1992), pp. 303–309.
- [22] J.M. Singer and P.K. Sen, *M -methods in growth curve analysis*, *J. Stat. Plan. Inference* 13 (1986), pp. 251–261.
- [23] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A.v.d. Linde, *Bayesian measures of model complexity and fit*, *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* 64 (2002), pp. 583–639.
- [24] M.A. Tanner and W.H. Wong, *The calculation of posterior distributions by data augmentation*, *J. Am. Stat. Assoc.* 82 (1987), pp. 528–540.
- [25] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.
- [26] K.H. Yuan, P.M. Bentler, and W. Chan, *Structural equation modeling with heavy tailed distributions*, *Psychometrika* 69 (2004), pp. 421–436.
- [27] K.H. Yuan, P.M. Bentler, and W. Zhang, *The effect of skewness and kurtosis on mean and covariance structure analysis*, *Sociol. Methods Res.* 34 (2005), pp. 240–258. Available at <http://dx.doi.org/10.1177/0049124105280200>
- [28] Z. Zhang, F. Hamagami, L. Wang, K.J. Grimm, and J.R. Nesselrode, *Bayesian analysis of longitudinal data using growth curve models*, *Int. J. Behav. Dev.* 31 (2007), pp. 374–383.
- [29] Z. Zhang, K. Lai, Z. Lu, and X. Tong, *Bayesian inference and application of robust growth curve models using Student’s t distribution*, *Struct. Equ. Model.* 20 (2012), pp. 47–78.
- [30] J. Zu and K.H. Yuan, *Local influence and robust procedures for mediation analysis*, *Multivariate Behav. Res.* 45 (2010), pp. 1–44.

Appendix 1. SAS and OpenBUGS scripts for the normal growth curve model

A1.1 SAS scripts

```

PROC MCMC DATA=gem NMC=100000 NBI=5000 THIN=5
  OUTPOST=histnorm DIC SEED=17 INIT=RANDOM DIAG=(ESS
  GEWEKE(F1=.2 F2=.5)) STATISTICS(ALPHA=0.05)=(
  SUMMARY INTERVAL);
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
  TADPANEL GEWEKE ESS DIC POSTINTERVALS;
ARRAY b[2] L S;
ARRAY beta[2];
ARRAY Sigma_b[2,2];
ARRAY beta0[2] (0 0);
ARRAY sigma0[2,2] (1000 0 0 1000);
ARRAY V[2,2] (1 0 0 1);
PARMS beta {3 2};
PARMS sigma_b {9 0 0 0.2};
PARMS var_e 1;
PRIOR beta ~ MVN(beta0, sigma0);
PRIOR Sigma_b ~ IWISH(2, V);
PRIOR var_e ~ IGAMMA(0.001, SCALE=0.001);
RANDOM b ~ MVN(beta, Sigma_b) SUBJECT=id;
mu = L + S * time;
MODEL y ~ NORMAL(mu, var=var_e);
RUN;

```

A1.2 OpenBUGS scripts

```

model{
  for (i in 1:N){
    b[i,1:2]~dmnorm(beta[1:2], Inv_Sigma[1:2,1:2])
    for (t in 1:T){
      y[i, t] ~ dnorm(muY[i,t], Inv_sig_e2)
      muY[i,t]<-b[i,1]+b[i,2]*t
    }
  }
  Inv_sig_e2 ~ dgamma(.001, .001)
  sig_e2 <- 1/Inv_sig_e2
  for (i in 1:2){
    beta[i] ~ dnorm(0, .001)
  }
  Inv_Sigma[1:2,1:2]~dwish(R[1:2,1:2], 2)
  Sigma[1:2,1:2]<-inverse(Inv_Sigma[1:2,1:2])
}

```

Appendix 2. SAS and OpenBUGS scripts for the GED growth curve model

A2.1 SAS scripts

```

PROC MCMC DATA=gem NMC=300000 THIN=10 OUTPOST=
  histged DIC SEED=17 INIT=RANDOM DIAG=(ESS GEWEKE(
  F1=.2 F2=.5)) STATISTICS(ALPHA=0.05)=(SUMMARY
  INTERVAL) MONITOR=(_PARMS_ var_e);
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
  TADPANEL GEWEKE ESS DIC POSTINTERVALS;
ARRAY b[2] L S;
ARRAY beta[2];
ARRAY sigma_b[2,2];
ARRAY beta0[2] (0 0);

```

```

ARRAY sigma0[2,2] (1000 0 0 1000);
ARRAY V[2,2] (1 0 0 1);
PARMS beta {3 2};
PARMS sigma_b {5 0 0 0.2};
PARMS sigma_e 1;
PARMS g1 0;
PARMS g2 0;
PARMS g3 0;
PARMS g4 0;
PRIOR beta ~ MVN(beta0, sigma0);
PRIOR sigma_b ~ IWISH(2, V);
PRIOR sigma_e ~ IGAMMA(0.001, SCALE=0.001);
PRIOR g: ~ UNIFORM(-1,5);
RANDOM b ~ MVN(beta, sigma_b) SUBJECT=id;
mu = L + S * time;
g = g1*(time=1)+ g2*(time=2)+g3*(time=3)+g4*(time=4);
p = 1+g;
ll = -LOG(sigma_e) + .5*LGAMMA(1.5*p)-1.5*LGAMMA(p/2)-log(p) - (
  abs(y-mu)/sigma_e)**(2/p) * (GAMMA(1.5*p)/GAMMA(.5*p))**(1/p);
var_e = sigma_e*sigma_e;
MODEL y~GENERAL(ll);
RUN;

```

A2.2 OpenBUGS scripts

```

model{
  for (t in 1:T){
    p[t]<-1+g[t]
  }
  for (i in 1:N){
    b[i,1:2]~dmnorm(beta[1:2], Inv_Sigma[1:2,1:2])
    for (t in 1:T){
      muY[i,t]<-b[i,1]+b[i,2]*t
      dummy[i,t,1]<-0
      dummy[i,t,1] ~ dloglik(logLike[i,t])
      logLike[i,t] <- -log(sig_e) + .5*loggam(1.5*p[t]) - 1.5*
        loggam(.5*p[t]) - log(p[t]) - pow(abs(y[i,t]-muY[i,t])/
          sig_e,2/p[t]) * pow(exp(loggam(1.5*(p[t])))-loggam(.5*(p
            [t]))), 1/p[t] )
    }
  }
  sig_e <- 1/Inv_sig_e
  Inv_sig_e~ dgamma(.001, .001)
  sig_e2 <- sig_e*sig_e
  for (i in 1:2){
    beta[i] ~ dnorm(0, .001)
  }
  for (t in 1:T){
    g[t] ~dunif(-1,5)
  }
  Inv_Sigma[1:2,1:2]~dwish(R[1:2,1:2], 2)
  Sigma[1:2,1:2]<-inverse(Inv_Sigma[1:2,1:2])
}

```