

Modeling error distributions of growth curve models through Bayesian methods

Zhiyong Zhang¹

Published online: 28 May 2015
© Psychonomic Society, Inc. 2015

Abstract Growth curve models are widely used in social and behavioral sciences. However, typical growth curve models often assume that the errors are normally distributed although non-normal data may be even more common than normal data. In order to avoid possible statistical inference problems in blindly assuming normality, a general Bayesian framework is proposed to flexibly model normal and non-normal data through the explicit specification of the error distributions. A simulation study shows when the distribution of the error is correctly specified, one can avoid the loss in the efficiency of standard error estimates. A real example on the analysis of mathematical ability growth data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 is used to show the application of the proposed methods. Instructions and code on how to conduct growth curve analysis with both normal and non-normal error distributions using the the MCMC procedure of SAS are provided.

Keywords Growth curve models · Bayesian estimation · Non-normal data · *t*-distribution · Exponential power distribution · Skew normal distribution · SAS PROC MCMC

Introduction

In the past half century, growth curve models (GCMs) have evolved from fitting a single curve of one individual to

fitting multilevel or mixed-effects models and from linear to nonlinear models (e.g., McArdle, 2000; McArdle & Nesselroade, 2002; Preacher et al., 2008; Tucker, 1958). The application of GCMs in social and behavioral research has grown rapidly since Meredith and Tisak (1990) showed that such models can be fitted as a restricted common factor model in the structural equation modeling framework (see also, McArdle (1986) and McArdle and Epstein (1987)). Nowadays, GCMs are widely used in analyzing longitudinal data to study change (e.g., Laird & Ware, 1982; McArdle & Nesselroade, 2003; Meredith & Tisak, 1990; Zhang et al., 2007). For a more comprehensive discussion of GCMs, see (McArdle & Nesselroade, 2002).

To estimate GCMs, the maximum likelihood estimation (MLE) method is commonly used (e.g., Demidenko, 2004; Laird & Ware, 1982). MLE is available in both commercial software, such as SAS PROC MIXED, and free software such as the R package lme4. MLE typically assumes that the errors of GCMs are normally distributed. However, in reality, data are more likely to be non-normally than normally distributed (Micceri, 1989). The violation of the normality assumption 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).

Recently, Bayesian methods have become very popular in the social and behavioral sciences largely advanced by the work of Lee and colleagues (e.g., Lee, 2007; Song & Lee, 2012). Nowadays, Bayesian methods have also received significant attention as useful tools for estimating a variety of models including GCMs, especially complex GCMs that can be difficult or impossible to estimate in the current MLE framework or in MLE based software (e.g., Chow et al., 2011; Muthén & Asparouhov, 2012; Song et al., 2009, 2011; Wang & McArdle, 2008; Zhang

✉ Zhiyong Zhang
ZhiyongZhang@nd.edu

¹ Department of Psychology, University of Notre Dame, 118 Haggard Hall, Notre Dame, IN 46556, USA

et al., 2007). For example, Song et al. (2009) proposed a novel latent curve model for studying the dynamic change of multivariate manifest and latent variables and their linear and interaction relationships. Wang and McArdle (2008) applied the Bayesian method to estimate a GCM with unknown change points. Zhang et al. (2007) discussed how to integrate information in growth curve modeling through informative priors. There are also Bayesian studies to analyze non-normal longitudinal data using GCMs. For example, Wang et al. (2008) developed Tobit GCMs to deal with limited response data. Zhang et al. (2013) extended GCMs to allow the errors to follow a t -distribution.

The Bayesian estimation for GCMs has been generally implemented in the specialized software WinBUGS. Although WinBUGS is flexible to estimate a wide range of models, it is unfamiliar to a lot of researchers and can be difficult to use. Since the version 9.2, the SAS institute released a Markov chain Monte Carlo (MCMC) procedure to implement a general Bayesian estimation procedure through the adaptive block random-walk Metropolis algorithm with normal proposal distributions (SAS, 2009). The MCMC procedure has the potential to become popular because of many important advantages, a few given here. First, SAS already has a large user base who can understand and learn the use of the MCMC procedure relatively easily. Second, the data manipulation procedure available in the SAS data step can be used with the MCMC procedure seamlessly. Third, specifically for the growth curve analysis, the use of the MCMC procedure is similar to the use of the existing NLMIXED procedure. Fourth, it is easy to define new prior and data distributions in the MCMC procedure. Currently, the use of the MCMC procedure is still rare, not to say its application in growth curve modeling. One exception is Zhang (2013) that provides SAS code for the MCMC procedure for growth curve modeling but does not give any instruction on its use.

Therefore, this study serves two main purposes. First, we will present a general framework to flexibly model the error distributions of GCMs. Second, we will demonstrate how to estimate the models using the MCMC procedure with instructions. In the following, we will first discuss how to model different types of error distributions. Specifically, we will focus on how to use Bayesian methods to estimate GCMs with the normal, t , exponential power, and skew-normal errors. Then, we will demonstrate how to practically apply the method through the analysis of data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). Finally, we will evaluate the influence of distribution misspecification through a simulation study.

Growth curve models with different types of error distributions

A typical GCM can be written as a two-level model. The first level involves fitting a growth curve with random coefficients for each individual and the second level models the random coefficients. Specifically, the first level of a GCM can be written as,

$$y_{it} = f(\text{time}_{it}, \mathbf{b}_i) + e_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N,$$

where y_{it} is the observed response for individual i at trial or occasion t , time_{it} denotes the measurement time of the t th measurement occasion on the i th individual, $f(\text{time}_{it}, \mathbf{b}_i)$ is a function of time representing the growth curve with individual coefficients $\mathbf{b}_i = (b_{i1}, b_{i2}, \dots, b_{ip})'$, a $p \times 1$ vector of random coefficients, and e_{it} is the residual or error.

The function $f(\text{time}_{it}, \mathbf{b}_i)$ can take either a linear or a nonlinear form of the time variable. The function f determines the growth trajectory representing how an individual changes over time. For example, if $f(\text{time}_{it}, \mathbf{b}_i) = b_{i0} + b_{i1}t$ with $\text{time}_{it} = t$ and $\mathbf{b}_i = (b_{i0}, b_{i1})'$, it represents a linear growth trajectory. If $f(\text{time}_{it}, \mathbf{b}_i) = b_{i0} + b_{i1}t + b_{i2}t^2$ with $\text{time}_{it} = t$ and $\mathbf{b}_i = (b_{i0}, b_{i1}, b_{i2})'$, it represents a quadratic growth trajectory.

The second level of a GCM can be expressed as,

$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{u}_i$$

where $\boldsymbol{\beta}$ is a vector of fixed-effects that can be seen as the average of the random effects \mathbf{b}_i without covariates in the second level, and \mathbf{u}_i represents, typically, the multivariate normal errors with $E(\mathbf{u}_i) = \mathbf{0}_{p \times 1}$ and $Cov(\mathbf{u}_i) = \mathbf{D}_{p \times p}$. The \mathbf{u}_i is also called the random effects from a mixed-effects model perspective.

For growth curve analysis, it is typically assumed that e_{it} is normally distributed with mean 0 and variance σ_e^2 . However, the normal distribution often lacks the flexibility to deal with complexity of real data. For example, the normal distribution is symmetric but data in practice can be skewed. Furthermore, it has been shown that the normal distribution is not robust to outliers or data with heavy tails (Yuan and Zhang, 2012b). If a normal distribution is mis-specified for a non-normal distribution, the parameter estimator will not be as efficient.

In this study, we propose a general framework to model the distribution of e_{it} explicitly and flexibly. The e_{it} can be assumed to have the general form $p(e_{it}|\boldsymbol{\zeta})$ where its exact form can be decided from data. The distribution parameter $\boldsymbol{\zeta}$ can typically be estimated within a GCM. To be consistent with the growth curve analysis literature, the mean of e_{it} is assumed to be 0. To determine the distribution of e_{it} ,

individual growth curve analysis can be utilized; to estimate the model parameters including β , \mathbf{D} , σ_e^2 , and ζ , Bayesian estimation methods can be applied.

Individual growth curve analysis

To explore the distribution of e_{it} , the method based on the individual growth curve analysis discussed by Tong and Zhang (2012) can be applied. Using the method, an individual growth curve

$$y_{it} = f(\text{time}_{it}, \mathbf{b}_i) + e_{it}, \quad t = 1, \dots, T$$

is first fitted to data from each individual through the least squares method. The individual coefficients \mathbf{b}_i and the error e_{it} can be estimated and retained. Specifically, we have

$$\hat{e}_{it} = y_{it} - f(\text{time}_{it}, \hat{\mathbf{b}}_i).$$

With the estimated errors, one can first conduct a D'Agostino skewness test (D'Agostino, 1970) on the errors. The D'Agostino skewness test evaluates whether a distribution is symmetric or not. If the test rejects the symmetry of the distribution, a skew distribution is desired. One then conducts an Anscombe test (Anscombe & Glynn, 1983) on the kurtosis. The Anscombe test evaluates whether the kurtosis of the error is different from the normal kurtosis. If the test rejects the normal kurtosis, a distribution with excess kurtosis is then needed.

In addition to the summary statistics on skewness and kurtosis, more detailed distributional information can be observed through the histogram or density of \hat{e}_{it} . Based on the information, one can decide on one or more candidate error distributions. If more than one candidate distribution can be used, the Kolmogorov–Smirnov test (e.g., Chakravarti et al., 1967) can also be used to select the best one. One can also try out all candidate distributions for comparison. To ease the procedure, an online program has been developed to assist the selection of error distribution. The URL to the program is <http://psychstat.org/gcmdiag>.

Two issues need to be emphasized here. First, the individual analysis method assumes that the growth function is already known a priori. The following strategy can be used to assist the choice of the growth function. If a growth function is supported theoretically or empirically, it should be used. Otherwise, exploratory techniques can be used to assist the selection of a growth function. For example, both the individual and mean growth trajectories from the data can be plotted. Based on the plot, one can decide whether a trajectory is linear or nonlinear. With the linear trajectory, a linear function can be used. For the nonlinear trajectory, it might also suggest what kind of nonlinear form fits best.

There are situations where several nonlinear forms can be applied. In the situations, one might try all the nonlinear functions or select the one with the simplest form. Note that in studying the trajectory, both the mean and individual trajectories should be considered because the mean trajectory is often influenced by the non-normality of data. Second, even when the growth function is known, the population error distribution is often impossible to know. The individual analysis method aims to best approach the population distribution through the empirical data. In addition, one would prefer a distribution that is less affected by potential misspecifications.

Bayesian estimation

General introduction on Bayesian analysis can be found in both introductory books (e.g., Kruschke, 2011) and specialized ones (e.g., Lee, 2007; Song & Lee, 2012). For a GCM with the error distribution $p(e_{it}|\zeta)$, the joint posterior distribution of the random effects \mathbf{b}_i and model parameters based on the data augmentation algorithm (Tanner & Wong, 1987) is

$$\begin{aligned} & p(\beta, \mathbf{D}, \zeta, \mathbf{b}_i | y_{it}, \text{time}_{it}, i = 1, \dots, N, t = 1, \dots, T) \\ & \propto p(\beta, \mathbf{D}, \zeta) \prod_{i=1}^N \left\{ p(\mathbf{b}_i | \beta, \mathbf{D}) \prod_{t=1}^T [p(y_{it} | f(\text{time}_{it}, \mathbf{b}_i), \zeta)] \right\} \\ & = p(\beta, \mathbf{D}, \zeta) \prod_{i=1}^N \left\{ \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{b}_i - \beta)' \Sigma^{-1} (\mathbf{b}_i - \beta) \right] \right\} \\ & \quad \times \prod_{i=1}^N \prod_{t=1}^T [p(y_{it} | f(\text{time}_{it}, \mathbf{b}_i), \zeta)] \end{aligned} \quad (1)$$

where $p(\beta, \mathbf{D}, \zeta)$ is the joint prior distribution for model parameters.

In this study, the uninformative independent priors are used for β , \mathbf{D} , and ζ such that $p(\beta, \mathbf{D}, \zeta) = p(\beta)p(\mathbf{D})p(\zeta)$ to minimize the influence of priors. For β , a bivariate normal prior $MN(\beta_0, \Sigma_0)$ is used with $\beta_0 = (0, 0)'$ and $\Sigma_0 = 1000\mathbf{I}_2$. For \mathbf{D} , an inverse Wishart prior distribution $IW(m_0, \mathbf{V}_0)$ is used with $m_0 = 2$ and $\mathbf{V}_0 = \mathbf{I}_2$. Here, \mathbf{I}_2 denotes a 2 by 2 identity matrix. These priors are chosen so that the density is relatively flat and are often viewed as uninformative in the Bayesian literature (e.g., Congdon, 2003; Gelman et al., 2003). The prior distribution for ζ depends on the form of the error distribution.

Incorporating the priors above to Eq. 1, the resulting posterior is readily available but usually in a very complex form. To evaluate the posterior, the Markov chain Monte Carlo (MCMC) methods can be used to generate a sequence of random numbers conditionally on consecutive draws, formally known as Markov chains, from the posterior distribution, and to construct the parameter estimates using the average and standard deviation of the generated numbers. In the SAS MCMC procedure, it uses the block

Metropolis method that divides the model parameters into different blocks and samples each block of parameter successively. Specifically, we divide the parameters and the random effects into the following blocks - β , \mathbf{D} , ζ , and \mathbf{b}_i , $i = 1, \dots, N$. From the joint posterior distribution, we can obtain the conditional posterior distribution for each block given the rest of parameters. Within each block, the conditional posterior can be obtained and sampled. Details on such algorithms can be found in the literature (e.g., Robert & Casella, 2004).

There are two practical issues in applying MCMC methods. The first is to diagnose the convergence of Markov chains. The convergence ensures that the empirical average from a Markov chain converges to the mean of the posterior distribution. Brooks and Roberts (1998) and Cowles and Carlin (1996) discussed many different methods for testing convergence. For example, one can use the Geweke test (Geweke, 1992) and/or visually inspect the trace plot of Markov chains. The Geweke test calculates a statistic that is normally distributed. For a converged Markov chain, the absolute value of a Geweke statistic should be smaller than 1.96 at the 0.05 alpha level. Visual inspection focuses on whether there is any unstable pattern in the plot of the generated Markov chains. Often times, the initial part of a Markov chain does not converge to the posterior and is often discarded as the burn-in period. The second is to decide the length of Markov chains. By nature, Markov chains always have autocorrelation. For two Markov chains, the one with greater 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 its autocorrelation is greater. To characterize the information in a Markov chain, we use the statistic called effective sample size (ESS). The ESS is the equivalent sample size assuming no autocorrelation. For two Markov chains with the same length, the one with the greater ESS provides more information. A practical rule of thumb is to get an ESS at least 400 (Lunn et al., 2012).

For a converged Markov chain with sufficient effective sample size, we can construct the posterior mean, posterior standard deviation, and highest posterior density (HPD; Box & Tiao, 1973) credible interval for each model parameter for inference. Let θ represent a single parameter in the vector $\boldsymbol{\theta} = (\beta, \mathbf{D}, \zeta)'$. Suppose the converged Markov chain for θ is $\theta_i, i = 1, \dots, n$ where n is the number of iterations. Then, a point estimate of θ can be constructed by the sample mean of the Markov chain

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

The standard deviation, or the standard error in the frequentist view, of θ is given by

$$\widehat{s.d.}(\hat{\theta}) = \frac{1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2.$$

Credible intervals for θ can also be constructed based on the Markov chain. The most widely used credible intervals include the equal-tail credible interval and the HPD credible interval. A $100(1 - \alpha) \%$ equal-tail credible interval 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 credible interval is the interval 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 credible interval is the same as the HPD credible interval. For non-symmetrical posteriors, the HPD credible interval has smaller width than the equal-tail credible interval. Therefore, HPD is often used for asymmetrical posterior distributions.

Illustration of error distributions

Potentially, the error e_{it} can follow any distribution. The normal distribution is certainly most widely used. In addition to the normal distribution, we present three other distributions that can extend GCMs to non-normal data analysis. They include the t , exponential power, and skew normal distributions. It can be seen that the normal distribution is a special case of the three distributions.

Student's t -distribution

The t -distribution can be viewed as a generalization of the normal distribution. The t -distribution has a kurtosis greater than that of the normal distribution. If e follows a t -distribution $t(e|\zeta = (\mu, \phi, k)')$, its density function can be written as

$$p_t(e) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \left(\frac{1}{\pi k \phi} \right)^{1/2} \left[1 + \frac{(e - \mu)^2}{k \phi} \right]^{-\frac{k+1}{2}}$$

where μ is the location parameter, ϕ is the scale parameter, and k is the degrees of freedom. The t -distribution has longer tails and is more robust to outliers than the normal distribution (e.g., Lange et al., 1989; Zhang et al., 2013). Therefore, the t -distribution is often used in robust data analysis. As the degrees of freedom k goes to infinity, the t -distribution approaches the normal distribution. The mean of the t -distribution is

$$E(e) = \mu, k > 1,$$

and the variance is

$$Var(e) = \frac{k}{k-2}\phi, k > 2.$$

The t -distribution is symmetric with a skewness value 0. Its kurtosis is

$$\gamma_2 = \frac{3k-6}{k-4}, k > 4.$$

Note that when k goes to infinity, $\gamma_2 = 3$, the kurtosis of the normal distribution. If the error $e_{it} \sim t(e_{it}|0, \phi, k)$, then $y_{it} \sim t(y_{it} - f(time_{it}, \mathbf{b}_i)|0, \phi, k)$.

In Fig. 1, we plot the density of the t -distribution with the degrees of freedom $k = 1, 3, 6, 30, \infty$ and the same location parameter $\mu = 0$ and scale parameter $\phi = 1$. When $k = \infty$, its density is the same as the normal distribution. Note that with the increase of degrees of freedom, the t -distribution moves closer to the normal distribution. Also, it can be observed, on the tails, the t -distribution has greater density than the normal distribution and therefore allows data to be far away from the center. For real data, k can be estimated as an unknown parameter to quantify the deviation from normality.

Exponential power distribution

The t -distribution can model the error e_{it} with kurtosis larger, but not smaller, than 3. The exponential power distribution, also called the generalized error distribution or generalized normal distribution, on the other hand, can model the error with both positive and negative kurtosis (Box & Tiao, 1973; Nadarajah, 2005). The exponential

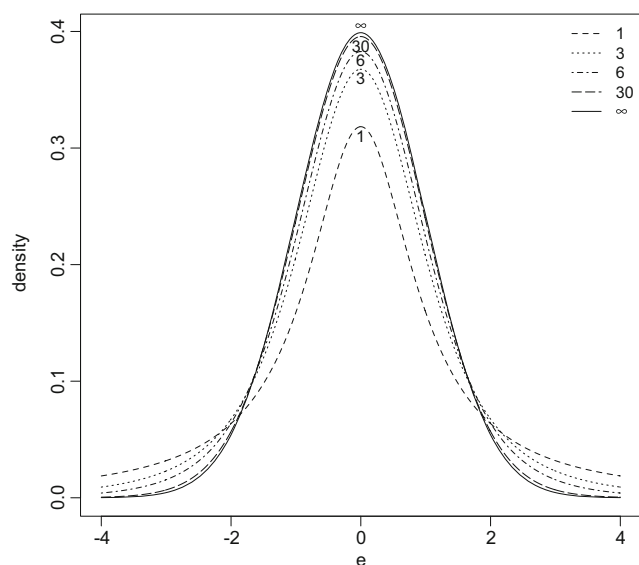


Fig. 1 The t -distribution with the location parameter $\mu = 0$ and scale parameter $\phi = 1$. The degrees of freedom is 1,3,5,30, and ∞ , respectively. The kurtosis for $k = 1$ or $k = 3$ does not exist, and the kurtosis is 9,3.23, and 3 for $k = 6, 30$, and ∞ , respectively

power distribution comes with different forms. In this study, the form by Box and Tiao (1973) is used. If e follows an exponential power distribution $ep(e|\xi = (\mu, \sigma, \gamma)')$, the density function can be expressed as

$$p_{ep}(e) = \omega(\gamma)\sigma^{-1} \exp \left[-c(\gamma) \left| \frac{e - \mu}{\sigma} \right|^{2/(1+\gamma)} \right] \tag{2}$$

where

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

and

$$c(\gamma) = \left\{ \frac{\Gamma[3(1 + \gamma)/2]}{\Gamma[(1 + \gamma)/2]} \right\}^{1/(1+\gamma)}. \tag{4}$$

Here μ is a location parameter, and σ is a scale parameter that is also the standard deviation of the population. γ is a shape parameter that is related to the kurtosis of the distribution and characterizes the “non-normality” of the population. The mean and variance of the exponential power distribution are

$$E(e) = \mu$$

$$Var(e) = \sigma^2.$$

The exponential power distribution is symmetric and therefore its skewness is 0. Its kurtosis is

$$\gamma_2 = \frac{\Gamma[5(1 + \gamma)/2]\Gamma[(1 + \gamma)/2]}{\Gamma[3(1 + \gamma)/2]^2}.$$

From the exponential power distribution, we can derive many other distributions. For example, if $\gamma = 0$, the exponential power distribution becomes a normal distribution. If $\gamma = 1$, it becomes a double exponential distribution. Furthermore, when γ approaches -1 , it approaches a uniform distribution. For growth curve analysis, the error at different times can take an exponential power distribution with different γ values to account for both platykurtic and leptokurtic errors simultaneously. If the error $e_{it} \sim ep(e_{it}|0, \sigma, \gamma)$, then $y_{it} \sim ep(y_{it} - f(time_{it}, \mathbf{b}_i)|0, \sigma, \gamma)$.

To better illustrate the influence of γ , we plot the density function of the exponential power distribution with different γ values in Fig. 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 $-.75, -.25, 0, .5, 1$. When $\gamma = 0$, the distribution becomes a normal distribution. When $\gamma > 0$, the exponential power distribution has a fatter tail than the normal distribution; when $\gamma < 0$, it has a thinner tail than the normal distribution. Therefore, the generalized error distribution can deal with data with both fat and thin tails. For real data, γ can be estimated as an unknown parameter (Box & Tiao, 1973).

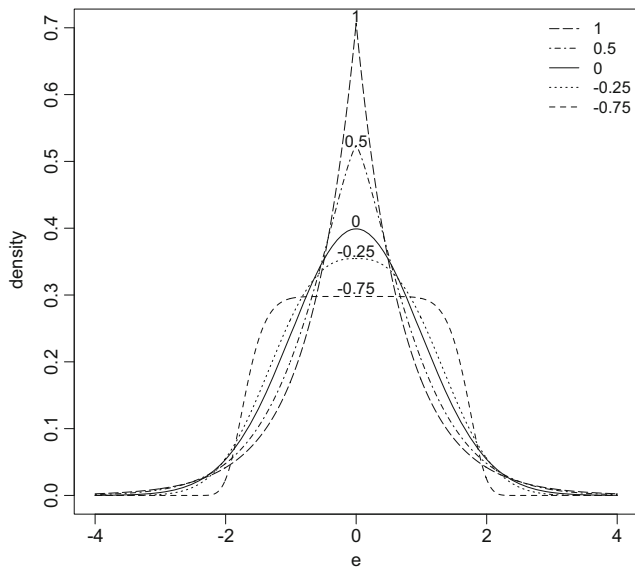


Fig. 2 The generalized error distribution with $\mu = 0$, $\phi = 1$ and different values of γ . Each value of γ is placed on the top of the density plot. The corresponding kurtosis for $\gamma = -0.75, -0.25, 0, .5, 1$ is 1.92, 2.55, 3, 4.22, 6, respectively

Skew normal distribution

Both the t -distribution and the exponential power distribution are symmetric and, therefore, may not be applied if the error distribution is asymmetric. A skew normal distribution captures potential departure from symmetry by allowing for skewness. A skew normal distribution can also be viewed as a generalization of the normal distribution (Azzalini, 1985). If e follows a skew normal distribution $sn(e|\xi = (\mu, \omega, \alpha)'$), its density function can be written as

$$p_{sn}(e) = \frac{2}{\omega} \phi\left(\frac{e - \mu}{\omega}\right) \Phi\left(\alpha \frac{e - \mu}{\omega}\right)$$

where μ is a location parameter, ω is a scale parameter, and α is a shape parameter that determines the skewness of the distribution. ϕ and Φ are the standard normal density function and cumulative standard normal distribution function, respectively. The mean and variance of the skew normal distribution are

$$E(e) = \mu + \omega \delta \sqrt{\frac{2}{\pi}}$$

and

$$Var(e) = \omega^2 \left(1 - \frac{2\delta^2}{\pi}\right)$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$. Furthermore, its skewness is

$$\gamma_1 = \frac{4 - \pi}{2} \frac{(\delta\sqrt{2/\pi})^3}{(1 - 2\delta^2/\pi)^{3/2}}$$

The kurtosis of the distribution is

$$\gamma_2 = 3 + 2(\pi - 3) \frac{(\delta\sqrt{2/\pi})^4}{(1 - 2\delta^2/\pi)^2}$$

and is greater than that of the normal distribution.

To show the influence of α on skewness, we plot the density function of the skew normal distribution with different α values in Fig. 3. For comparison, we maintained the same location parameter $\mu = 0$ and scale parameter $\omega = 1$. The shape parameter α takes value $-4, -2, 0, 2, 4$. From the figure, when $\alpha = 0$, the skew normal distribution becomes a normal distribution and its density is symmetric. If $\alpha > 0$, the distribution is right skewed and if $\alpha < 0$, the distribution is left skewed. For real data, α can be estimated as an unknown parameter. If the error $e_{it} \sim sn(e_{it}|\mu, \omega, \phi)$, then $y_{it} \sim ep(y_{it} - f(time_{it}, \mathbf{b}_i)|\mu, \omega, \phi)$. Because we often assume the mean of error is zero, we can reparameterize the model parameter μ so that $\mu = -\omega\delta\sqrt{2/\pi}$.

The error distributions discussed here can be plugged into the posterior in Eq. 1. The Bayesian method can then be used to obtain model parameter estimates. We now explain how to estimate GCMs with different error distributions using the MCMC procedure of SAS.

Model estimation using the MCMC procedure

GCMs with different error distributions can be estimated using the MCMC procedure of SAS. We illustrate the use of the MCMC procedure in estimating GCMs with the normal distribution, t -distribution, exponential power distribution,

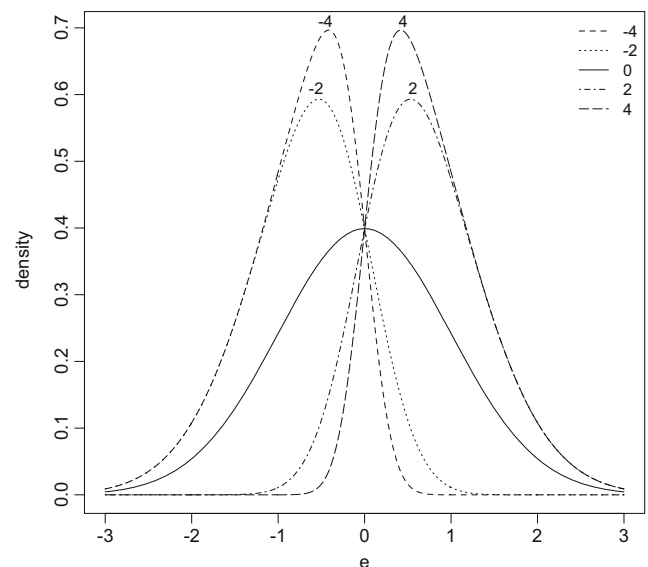


Fig. 3 The skew normal distribution with $\mu = 0$ and $\omega = 1$ and different values of α . The corresponding skewness for $\alpha = -4, -2, 0, 2, 4$ is $-0.78, -0.45, 0, .45, .78$, respectively

and skew normal distribution. The code can be found from our website at <http://psychstat.org/brm2015>.

Growth curve model with the normal error

The SAS code using the procedure MCMC for estimating the normal error model is given in Code Block 1. The code on the first line controls the overall data analysis. `DATA=ndata` specifies the data set to use. `NBI=5000` tells SAS to throw away the first 5,000 iterations as burn-in. `NMC=40000` tells SAS to generate Markov chains with another 40,000 iterations for analysis. `THIN=2` means that every other iteration will be saved for inference. Therefore, combining with NMC, a total of 20,000 iterations are saved. `SEED=17` specifies the random number generator seed. `INIT=RANDOM` is used to generate the random initial values for parameters that are not provided with initial values. `DIAG=(ESS GEWEKE)` requests the calculation of the effective sample size and the Geweke test statistic. `STATISTICS(ALPHA=0.05)=(SUMMARY INTERVAL)` requests the calculation of the summary and interval statistics for each parameter at the alpha level 0.05. `OUTPOST=histnorm` will save the generated Markov chains into a SAS data set named `histnorm`, which can be processed further if needed.

Line 2 uses the SAS output delivery system (ODS) to output nicely arranged tables and figures. `PARAMETERS` requests the output of the sampling method, prior distribution, and initial value information for each parameter. `POSTSUMMARIES` requests the output of the mean, standard deviation, and 25 %, 50 %, and 75 % percentiles, and `POSTINTERVALS` requests the output of the 95 % percentile and HPD credible intervals of model parameters. `GEWEKE` and `ESS` request the output of the Geweke test and the effective sample size. `TADPANEL` will generate the trace, autocorrelation and density plots in one figure for visually inspecting Markov chain convergence.

The keyword `ARRAY` specifies an array. For example, `b[2]` on Line 3 is an array with two elements and the first element is L and the second element is S . `Sigma_b[2,2]` on Line 5 specifies a 2 by 2 array. If the values of an array are known, they can be provided as fixed values. For example, `beta0[2]` on Line 6 has two elements with values 0 and $V[2,2]$ on Line 8 is an identity matrix.

The keyword `PARMS` specifies each block of parameters and their initial values. For example, on Line 9, `beta` is treated as one block of parameters with the initial values $\{5, 2\}$. Note that the initial values are enclosed by `{}` for vectors and arrays. The keyword `PRIOR` specifies the prior distribution for the model parameters. For example, a multivariate normal prior (MVN) is used for `beta` on Line 12. Note that `beta0` and `sigma0` are given in

the `ARRAY` statement. Furthermore, `Sigma_b` is given an inverse Wishart prior on Line 13 and `var_e` is given an inverse Gamma prior on Line 14.

The random coefficients `b` are augmented with model parameters in the estimation. The distribution of `b` is a multivariate (bivariate specifically) normal distribution and is specified as `RANDOM b ~ MVN(beta, Sigma_b)` `SUBJECT=id` on Line 15 where `id` is a variable that distinguishes each individual.

Since the error e_{it} follows a normal distribution, y_{it} also has a normal distribution with the mean $b_{i0} + b_{i1}t$ and the same variance. Therefore, the data are modeled using a normal distribution as shown in `MODEL y ~ NORMAL(mu, var=var_e)` on Line 17 with the mean is calculated by `mu = L + S * time` on Line 16.

Growth curve model with the t error

The SAS code for the linear GCM with the t error is given in Code Block 2. The code is very similar to that for the normal error except that the data `y` are modeled by the t -distribution `y ~ t(mu, var=var_e, df)` with the unknown degrees of freedom `df` as shown on Line 19. Furthermore, the `df` is given a uniform prior `UNIFORM(1, 500)` on Line 16.

Growth curve model with the exponential power error

The SAS code for the linear GCM with the exponential power error is given in Code Block 3. The code is similar to that for the normal error and the t error. In addition to the extra parameters on γ (`g1, g2, g3, g4` on Lines 12-15), a notable difference is that the data are given a `general` distribution with argument `ll` on Line 26. This is because there is no built-in exponential power distribution in SAS. However, a new distribution can be constructed flexibly. Comparing `ll` on Line 24 and the exponential power density function in Eq. 2, the use of a new distribution is clear and the new distribution is defined as the logarithm density function. Another difference is that `MONITOR=(_PARMS_ var_e)` is used on Line 1. `MONITOR` lists model parameters or newly created parameters with Markov chains to be saved. `_PARMS_` represents all parameters directly used in a model as specified by the keyword `PARMS`. New parameter can be calculated. For example, `var_e` is the variance calculated based on the standard deviation `sigma_e` on Line 25.

Growth curve model with the skew normal error

The SAS code for the linear GCM with the skew normal error is given in Code Block 4. The difference between

Code Block 1 SAS code for the GCM with the normal error

```

1 PROC MCMC DATA=ndata NBI=5000 NMC=40000 THIN=2 SEED=17 INIT=
  RANDOM DIAG=(ESS GEWEKE) STATISTICS(ALPHA=0.05)=(SUMMARY
  INTERVAL) OUTPOST=histnorm;
2 ODS SELECT PARAMETERS POSTSUMMARIES POSTINTERVALS GEWEKE ESS
  TADPANEL;
3 ARRAY b[2] L S;
4 ARRAY beta[2];
5 ARRAY Sigma_b[2,2];
6 ARRAY beta0[2] (0 0);
7 ARRAY sigma0[2,2] (1000 0 0 1000);
8 ARRAY V[2,2] (1 0 0 1);
9 PARMS beta {5 2};
10 PARMS sigma_b {1 0 0 1};
11 PARMS var_e 1;
12 PRIOR beta ~ MVN(beta0, sigma0);
13 PRIOR Sigma_b ~ IWISH(2, V);
14 PRIOR var_e ~ IGAMMA(0.001, SCALE=0.001);
15 RANDOM b ~ MVN(beta, Sigma_b) SUBJECT=id;
16 mu = L + S * time;
17 MODEL y ~ NORMAL(mu, var=var_e);
18 RUN;

```

the current code and previous one is that the error e , instead of the data y , is modeled directly as indicated by MODEL $e \sim \text{general}(11)$ on Line 23. The reason to model e is because the mean of the skew normal distribution is not “centered” around μ in the sense of the normal distribution. Directly modeling the data y will make the estimates of β not comparable with those from using the other distributions. Therefore, we “trick” the MCMC procedure by specifying a skew normal distribution with a mean $-\omega\delta\sqrt{2/\pi}$ as indicated by $x_i = -\text{sigma_e}*\alpha/\text{sqrt}(1+\alpha*\alpha)*\text{sqrt}(2/3.1415926)$ on Line 21.

An example

The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) focuses on children’s early school experiences from kindergarten to middle school. The ECLS-K began in the fall of 1998 with a nationally representative sample of 21,260 kindergartners. The ECLS-K is a longitudinal study with data collected in the fall and spring of kindergarten (1998-99), the fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007). Data on children’s home environment, home educational activities,

school achievement, school environment, classroom environment, classroom curriculum, and teacher qualifications are available.

Typical longitudinal studies do not have such a large sample size. Therefore, we randomly selected a sample of 563 (about 2.5 %) children with complete data on mathematical ability in the fall and spring of kindergarten and the 1st grade. Summary statistics for this data set are provided in Table 1 and the longitudinal plot for all data is presented in Fig. 4. Both the summary statistics and the longitudinal plot suggested a linear growth pattern.¹ Based on the skewness, the data appear to be symmetric. However, based on the kurtosis, only the data at time 1 seem to be normally distributed. The data distribution is platykurtic at time 2 and leptokurtic at time 4.

In order to diagnose the error distribution, we fitted a linear regression to the data from each individual. The residual errors from the regression analysis were pooled together for normality test. Specifically, the D’Agostino skewness test and Anscombe-Glynn kurtosis test were conducted (See

¹ A more rigorous method is to fit different models to the data to select the best fit one. However, it is not clear how to compare different models in the current framework. Instead, the no growth model, linear growth model, and the quadratic growth model were fitted to the data through MLE based on the normal error and it was found that the linear model fitted best overall.

Code Block 2 SAS code for the growth curve model with the t error

```

1 PROC MCMC DATA=ndata NMC=40000 NBI=5000 THIN=2 DIC SEED=17 INIT=
  RANDOM DIAG=(ESS GEWEKE) STATISTICS(ALPHA=0.05)=(SUMMARY
  INTERVAL) OUTPOST=histt;
2 ODS SELECT PARAMETERS POSTSUMMARIES POSTINTERVALS GEWEKE ESS
  TADPANEL;
3 ARRAY b[2] L S;
4 ARRAY beta[2];
5 ARRAY Sigma_b[2,2];
6 ARRAY beta0[2] (0 0);
7 ARRAY sigma0[2,2] (1000 0 0 1000);
8 ARRAY V[2,2] (1 0 0 1);
9 PARMS beta {5 2};
10 PARMS sigma_b {1 0 0 1};
11 PARMS var_e 1;
12 PARMS df 5;
13 PRIOR beta ~ MVN(beta0, sigma0);
14 PRIOR Sigma_b ~ IWISH(2, V);
15 PRIOR var_e ~ IGAMMA(0.001, SCALE=0.001);
16 PRIOR df~UNIFORM(1,500);
17 RANDOM b ~ MVN(beta, Sigma_b) SUBJECT=id;
18 mu = L + S * time;
19 MODEL y ~ t(mu, var=var_e, df);
20 RUN;

```

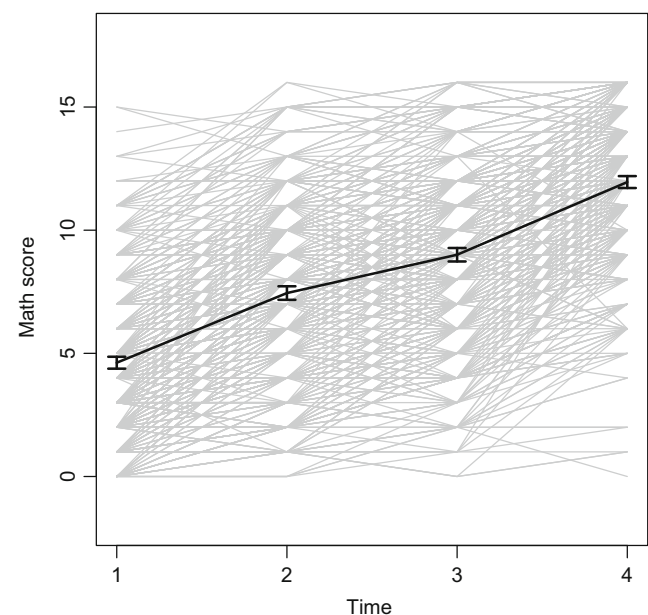
Table 1 Summary statistics of the example data

Time	Mean	Variance	Skewness	Kurtosis
1 (Fall, kindergarten)	4.62	8.58	0.69	3.22
2 (Spring, kindergarten)	7.45	11.11	0.04	2.58
3 (Fall, 1st year)	9.00	10.96	-0.22	2.74
4 (Spring, 1st year)	11.95	8.59	-1.00	4.27

Table 2; Anscombe & Glynn, 1983; D’Agostino, 1970). First, the skewness was -0.06 and the kurtosis was 3.77 for the overall error distribution. The D’Agostino test indicates that the error distribution was symmetric but the Anscombe-Glynn test showed that the kurtosis of the error distribution was greater than that of the normal distribution. Then, the error distribution was checked at each measurement time. The error distribution at each time seemed to be symmetric, too. However, the distribution at time 2 was platykurtic and at time 4 was leptokurtic. Therefore, the error distribution seems to be best modeled with the exponential power distribution at each time.

The linear GCM with the exponential power error distribution was fitted to the data. The trace plots for the model parameters indicate that the Markov chains for all parameters appeared to converge well after a burn-in period of

5,000 iterations. Furthermore, all Markov chains passed the Geweke test as shown in Table 3. The effective sample sizes ranged from 448 to 9791 for model parameters. Parameter

**Fig. 4** Longitudinal plot of the mathematical data of the ECLS-K sample

Code Block 3 SAS code for the growth curve model with the exponential power error

```

1 PROC MCMC DATA=ndata NMC=100000 NBI=5000 THIN=2 MONITOR=(_PARMS_
  var_e) DIC SEED=13 INIT=RANDOM DIAG=(ESS GEWEKE(F1=.2 F2=.5))
  STATISTICS(ALPHA=0.05)=(SUMMARY INTERVAL) OUTPOST=histged;
2 ODS SELECT PARAMETERS POSTSUMMARIES POSTINTERVALS GEWEKE ESS
  TADPANEL;
3 ARRAY b[2] L S;
4 ARRAY beta[2];
5 ARRAY Sigma_b[2,2];
6 ARRAY beta0[2] (0 0);
7 ARRAY sigma0[2,2] (1000 0 0 1000);
8 ARRAY V[2,2] (1 0 0 1);
9 PARMs beta {5 2};
10 PARMs sigma_b {1 0 0 1};
11 PARMs sigma_e 1;
12 PARMs g1 0;
13 PARMs g2 0;
14 PARMs g3 0;
15 PARMs g4 0;
16 PRIOR beta ~ MVN(beta0, sigma0);
17 PRIOR Sigma_b ~ IWISH(2, V);
18 PRIOR sigma_e ~ IGAMMA(0.01, SCALE=0.01);
19 PRIOR g:~UNIFORM(-.6,5);
20 RANDOM b ~ MVN(beta, Sigma_b) SUBJECT=id;
21 mu = L + S * time;
22 g = g1*(time=1)+ g2*(time=2)+g3*(time=3)+g4*(time=4);
23 p = 1+g;
24 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);
25 var_e = sigma_e*sigma_e;
26 MODEL y ~ general(ll);
27 RUN;

```

Table 2 Tests of skewness and kurtosis for errors from the individual growth curve analysis

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
Overall	-0.06	-0.78	0.435	3.77	5.41	0.000

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

Code Bock 4 SAS code for the growth curve model with the skew normal error

```

1 PROC MCMC DATA=ndata NMC=40000 NBI=5000 THIN=2 DIC SEED=17 INIT=
  RANDOM DIAG=(ESS GEWEKE) STATISTICS(ALPHA=0.05)=(SUMMARY
  INTERVAL) MONITOR=(_PARMS_ var_e) OUTPOST=histsn;
2 ODS SELECT PARAMETERS POSTSUMMARIES POSTINTERVALS GEWEKE ESS
  TADPANEL;
3 ARRAY b[2] L S;
4 ARRAY beta[2];
5 ARRAY Sigma_b[2,2];
6 ARRAY beta0[2] (0 0);
7 ARRAY sigma0[2,2] (1000 0 0 1000);
8 ARRAY V[2,2] (1 0 0 1);
9 PARMS beta {5 2};
10 PARMS sigma_b {1 0 0 1};
11 PARMS sigma_e 1;
12 PARMS alpha 0;
13 PRIOR beta ~ MVN(beta0, sigma0);
14 PRIOR Sigma_b ~ IWISH(2, V);
15 PRIOR sigma_e ~ IGAMMA(0.01, SCALE=0.01);
16 PRIOR alpha~UNIFORM(-.5,.5);
17 RANDOM b ~ MVN(beta, Sigma_b) SUBJECT=id;
18 mu = L + S * time;
19 var_e = sigma_e*sigma_e;
20 e = y - mu;
21 xi = -sigma_e*alpha/sqrt(1+alpha*alpha)*sqrt(2/3.1415926);
22 ll = log(2)-log(sigma_e)+logpdf('normal',(e-xi)/sigma_e,0,1)+
  logcdf('normal',alpha*(e-xi)/sigma_e,0,1);
23 MODEL e ~ general(ll);
24 RUN;

```

Table 3 Results from the linear growth curve model with the exponential power error distribution

	Estimates	SD	Est/SD	Equal-tail CI		HPD CI		Geweke		ESS
				2.5 %	97. %	2.5 %	97.%	statistic	p-value	
β_0	2.354	0.148	15.934	2.063	2.645	2.064	2.646	-0.291	0.771	9791
β_1	2.364	0.035	66.966	2.295	2.434	2.295	2.433	0.011	0.992	3964
σ_0^2	8.911	0.762	11.702	7.494	10.479	7.434	10.412	0.715	0.475	3164
σ_{01}	-0.654	0.156	-4.198	-0.972	-0.363	-0.962	-0.354	-0.770	0.441	1160
σ_1^2	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

estimates for the model are given in Table 3. Based on the estimates of γ_t , the errors at time 1, 3, and 4 had kurtoses greater than 3 (positive γ) while the errors at time 2 had kurtosis smaller than 3 (negative γ). Furthermore, based on the credible intervals, γ_4 is statistically significant indicating that the error distribution at time 4 cannot be treated as normal. Based on the parameter estimates, the average intercept was 2.354 with an average rate of growth 2.364 in the mathematical ability. There were also significant individual differences in both the intercept and rate of growth. Furthermore, children with a greater intercept would show a lower rate of growth based on the covariance estimate.

Influence of error distribution misspecification

We have demonstrated that we can model error distributions explicitly through Bayesian methods. If the errors are not normally distributed but the normal based method is used, the error distribution is misspecified. On the other hand, if the errors are normally distributed while a non-normal distribution is used, the error distribution is equally misspecified. In this section, we evaluate the influence of error distribution misspecification under the two situations through a simulation study. In the literature, the bootstrap method is often used to obtain the standard errors for model parameters when data are suspected of non-normality. Therefore, we also compare our methods using the non-normal distributions with the bootstrap method.

Data generation

Data are generated from the following linear GCM,

$$y_{it} = b_{i0} + b_{i1}t + e_{it}$$

$$b_{i0} = \beta_0 + v_{i0}$$

$$b_{i1} = \beta_1 + v_{i1}$$

where $\beta_0 = 5$ and $\beta_1 = 2$. The random effects follow a bivariate normal distribution with mean zeros and $Var(v_{i0}) = \sigma_0^2 = 2$, $Var(v_{i1}) = \sigma_1^2 = 1$, and $Var(v_{i0}, v_{i1}) = \sigma_{01} = 0.2$. To evaluate the influence of misspecifying normal error to non-normal error, e_{it} is set to follow a normal distribution with mean 0 and variance $\sigma_e^2 = 1$. To evaluate the influence of misspecifying non-normal error to normal error, e_{it} is set to follow a t -distribution with $\xi = (\mu, \phi, k) = (0, 1/3, 3)$, an exponential power distribution with $\xi = (\mu, \sigma, \gamma) = (0, 1, \gamma)$, and a skew normal distribution with $\xi = (\mu, \omega, \alpha) = c(-1.31, 1.64, 10)$, respectively. Note that the parameters of the skew normal distribution are chosen so that the mean of the distribution

²The population parameter values were chosen rather arbitrarily. But we do not expect that the choice of the values affects the simulation conclusions.

is 0. For the exponential power distribution, the parameter γ is set at different values at different measurement times. Three different sample sizes with $N = 100, 300, 500$ and two different measurement occasions with $T = 4, 5$ are investigated. The parameters $\gamma = (-0.8, -0.8, 0, 0.8, 0.8)$ for $T = 5$ and $\gamma = (-0.8, -0.8, 0.8, 0.8)$ for $T = 4$. Under each condition, 100 sets of data are generated and analyzed.

To estimate the model with the normal, t , exponential power, and skew normal distributions, the Bayesian method is used and implemented in the MCMC procedure of SAS as discussed earlier. For the bootstrap method, the same data are analyzed using Mplus and 1000 bootstrap samples are used. R code for data generation, running simulation, as well as Mplus script can be found online at <http://psychstat.org/brm2015>.

Misspecifying non-normal error as normal error

For the linear GCM with the normal error, t error, exponential power error, and skew normal error, the parameters $\beta_0, \beta_1, \sigma_0^2, \sigma_{01}$, and σ_1^2 are comparable and of the greatest interest to users. Figure 5 displays the scatter plots of parameter estimates and their standard errors for the 5 parameters when modeling the error distribution as the normal and as non-normal distributions with $N = 500$ and $T = 5$ across all 100 replications of data analysis. The scatterplots of the parameter estimates on the left panel of Fig. 5 clearly show that when the non-normal error distributions were misspecified as a normal distribution, there was not much difference in terms of parameter estimates. This is evidenced by the fact that the scatterplots nicely fall on the straight diagonal line.

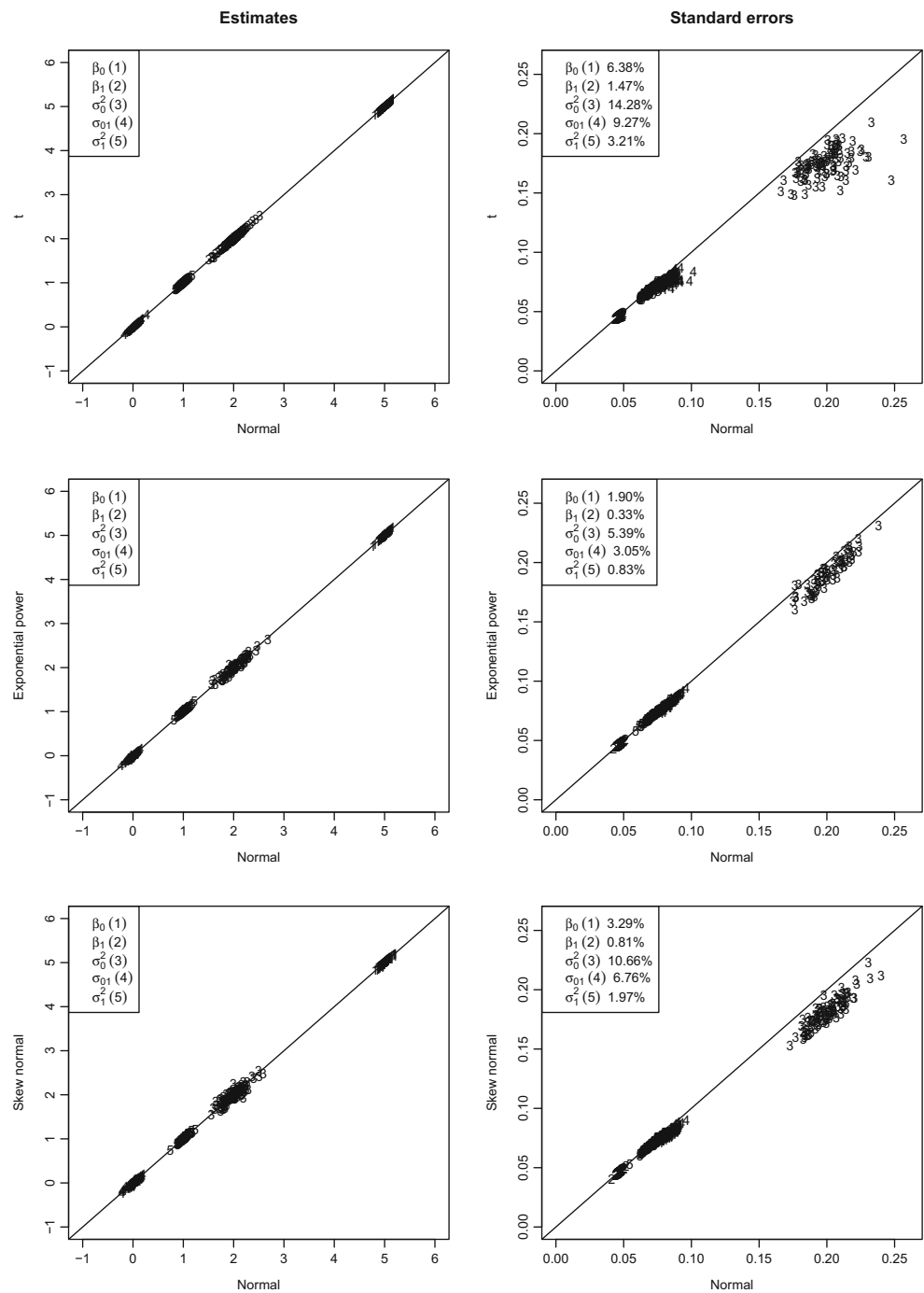
However, the right panel of Fig. 5 shows that the standard errors became greater if the non-normal error distributions were misspecified as normal distribution because in the scatterplots, the majority individual points fall under the straight line. This means that the standard errors were overestimated with the incorrect error distributions. To further quantify the overestimation, we calculate a statistic called reduced efficiency in standard error (*RESE*)³ defined as

$$RESE = 100 \times \left(\frac{1}{100} \sum_{r=1}^{100} \frac{\widehat{se}_r(\hat{\theta}_w)}{\widehat{se}_r(\hat{\theta}_c)} - 1 \right)$$

where $\widehat{se}_r(\hat{\theta}_w)$ is the standard error estimates of the parameter estimate $\hat{\theta}_w$ from the wrongly specified error distribution and $\widehat{se}_r(\hat{\theta}_c)$ is the corresponding one with the correctly specified error distribution for the r th replication of data. Note

³One can also define similar statistic based on mean squared error. The pattern of the results is similar and can be obtained from the author along with the parameter estimates, standard error estimates, HPD interval, Geweke statistic, and effective sample size for each replication.

Fig. 5 Parameter estimates and their standard errors when the t , exponential power, and skew normal distributions are, respectively, misspecified as normal distribution from top to bottom



that if $RESE > 0$, it indicates a loss of efficiency; otherwise, there is a gain in the efficiency.⁴ The $RMSE$ for each parameter is given in Fig. 5 for the condition $N = 500$ and

⁴Given that the standard error is generally inversely proportional to the square root of the sample size, $RESE = 5\%$ indicates an increase of about 10% sample size and $RESE = 10\%$ indicates an increase of about 20% sample size to achieve the same efficiency. Therefore, $RESE = 5\%$ should not be ignored and $RESE = 10\%$ can be considered as substantial.

$T = 5$ and all of them are greater than 0, indicating the loss of efficiency when the error distribution is misspecified.

Table 4 summarizes the $RESE$ when the non-normal error distributions are misspecified as the normal distribution with different sample sizes and measurement occasions. Although the $RESE$ varies, the overall patterns are the same. First, under all studies condition, there shows a reduced efficiency. Second, the random-effects parameters σ_0^2 and σ_{01}^2 are influenced the most by the misspecification

Table 4 RESE when the non-normal error distributions are misspecified as normal (in percentage)

Sample Size (<i>N</i>)		t			Exponential Power			Skew Normal		
		100	300	500	100	300	500	100	300	500
T=4	β_0	7.25	8.52	8.18	0.64	1.97	2.29	2.43	2.54	2.78
	β_1	2.04	2.95	2.89	0.61	0.59	0.53	0.94	1.04	1.11
	σ_0^2	14.47	20.87	19.54	8.45	6.59	8.02	11.29	10.49	11.97
	σ_{01}	10.07	14.68	14.24	4.12	3.99	4.96	7.37	7.49	8.22
	σ_1^2	4.39	6.50	6.23	0.88	1.26	1.55	2.55	2.78	3.02
T=5	β_0	5.61	7.77	6.38	0.89	1.86	1.90	2.79	3.39	3.29
	β_1	1.21	1.92	1.47	0.11	0.40	0.33	0.91	1.03	0.81
	σ_0^2	12.57	19.34	14.28	1.04	5.51	5.39	10.60	11.01	10.66
	σ_{01}	7.40	9.66	9.27	0.78	2.79	3.05	6.32	7.11	6.76
	σ_1^2	2.59	2.77	3.21	0.22	0.87	0.83	2.01	2.10	1.97

of the error distributions. Third, there is no clear pattern of the relationship between RESE and sample sizes as well as measurement occasions.

Table 5 summarizes the RESE to compare the bootstrap method and the Bayesian method with the correctly specified error distributions. If the RESE is negative, the bootstrap method is more efficient. Otherwise, the Bayesian method is more efficient. The results show that for β_0 , σ_0^2 , and σ_1^2 , the Bayesian method is consistently more efficient. For β_1 and σ_{01} , the results are mixed. Overall, it seems that the Bayesian method is more efficient than the bootstrap method.

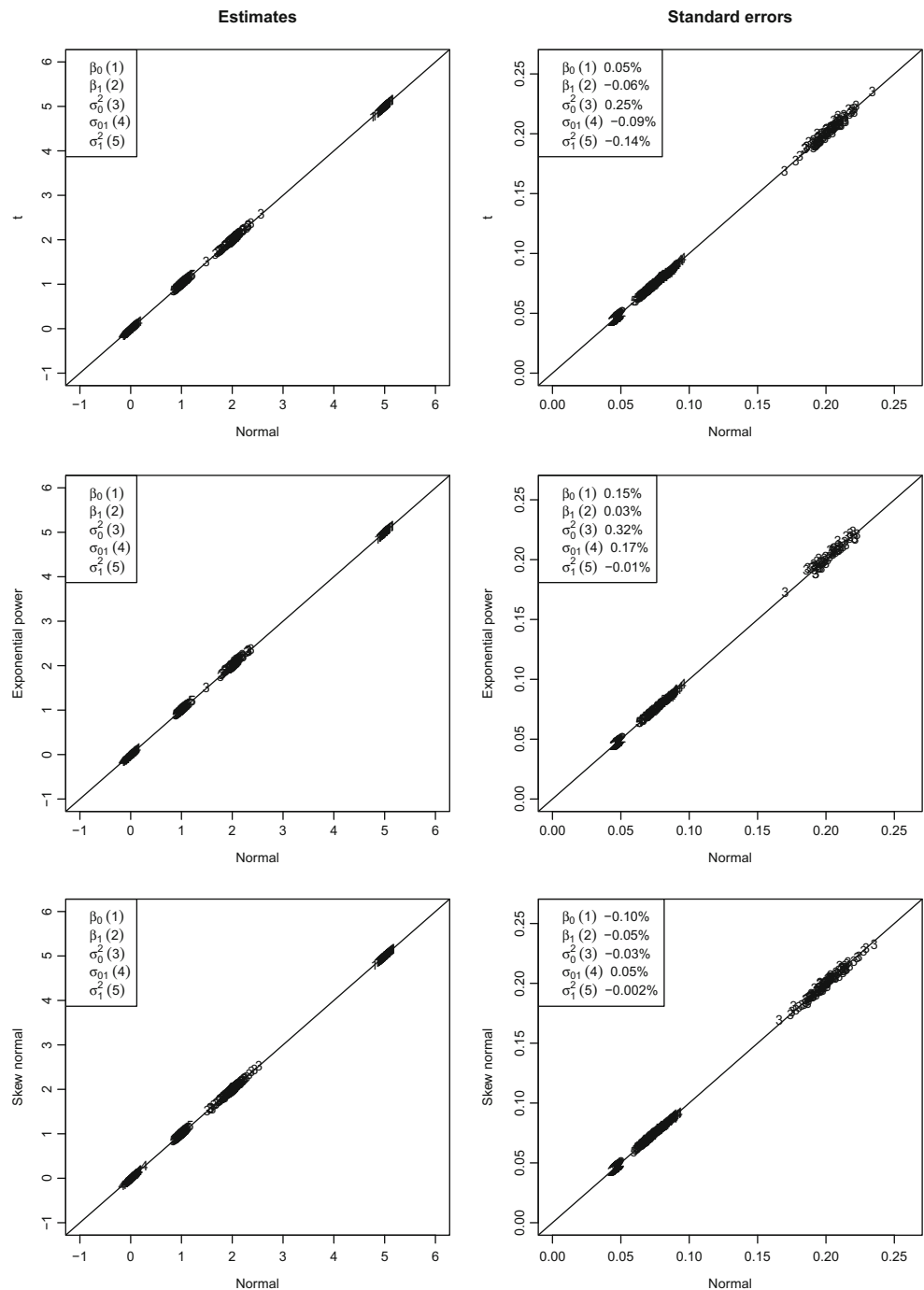
Misspecifying normal error as non-normal error

We have shown that misspecifying a non-normal error distribution to the normal distribution can reduce the efficiency of standard error estimates. Now, we evaluate the situation where the normal error distribution is misspecified as a non-normal distribution. The normal distribution can be viewed as a special case of the *t*-distribution, exponential power distribution, and skew normal distribution. If the errors are normally distributed but the three non-normal distributions are used in the model, the error distribution is over-specified, more precisely than misspecified. Therefore,

Table 5 RESE comparing the bootstrap standard errors with the Bayesian standard deviations using correctly specified error distributions (in percentage)

Sample Size (<i>N</i>)		t			Exponential Power			Skew Normal		
		100	300	500	100	300	500	100	300	500
T=4	β_0	11.20	10.56	10.22	5.33	5.07	3.62	5.60	4.62	5.90
	β_1	-0.14	1.24	1.90	0.13	1.98	0.30	-0.51	0.97	1.16
	σ_0^2	24.12	44.16	42.35	3.79	8.20	8.72	9.17	12.29	16.09
	σ_{01}	10.20	22.01	20.66	-0.72	4.22	3.75	-1.58	4.11	5.18
	σ_1^2	27.01	43.72	42.50	4.97	9.79	8.99	9.21	12.77	15.28
T=5	β_0	14.52	14.62	14.01	9.89	10.01	9.96	11.28	11.04	11.29
	β_1	-0.59	1.39	1.47	-1.50	1.06	0.35	-0.98	-0.63	0.38
	σ_0^2	28.53	43.93	42.56	12.28	20.66	21.08	21.43	25.37	26.94
	σ_{01}	-2.85	7.62	7.19	-1.73	-1.97	0.41	-2.26	-1.88	0.52
	σ_1^2	20.46	33.74	33.83	10.39	18.27	18.41	16.11	19.79	21.12

Fig. 6 Parameter estimates and their standard errors when the normal error distribution is misspecified as the t , exponential power, and skew normal distributions, respectively, from top to bottom



one would expect this kind of misspecification should not have a big influence on the parameter estimates and their standard errors.

Figure 6 shows scatterplots of the parameter estimates and their standard errors when modeling the error distribution as both the normal and non-normal distributions with $N = 500$ and $T = 5$ across all 100 replications of data analysis. Clearly, the parameter estimates and their standard

errors were very similar when the error was modeled as both the normal and non-normal distributions. Table 6 summarizes the *RESE* for all investigated conditions. Across the table, *RESE* is mostly smaller than 0.5 %. Compared to *RESE* in Table 4, the loss in efficiency in the standard error estimates is negligible when the normal error distribution was misspecified as any of the three non-normal distributions discussed here.

Table 6 RESE when the normal error distribution is misspecified as non-normal (in percentage)

Sample Size (<i>N</i>)		t			Exponential Power			Skew Normal		
		100	300	500	100	300	500	100	300	500
T=4	β_0	0.20	-0.12	-0.08	0.43	0.12	0.07	0.58	-0.04	-0.09
	β_1	-0.01	0.03	0.14	0.22	-0.01	0.10	0.29	-0.06	-0.07
	σ_0^2	0.17	-0.15	-0.05	-0.11	0.19	0.53	1.20	-0.11	0.14
	σ_{01}	0.10	-0.09	-0.09	0.35	0.05	0.30	0.60	0.16	0.14
	σ_1^2	-0.08	0.02	-0.11	0.23	0.12	0.27	0.33	0.16	0.03
T=5	β_0	0.09	0.05	0.05	0.39	0.16	0.15	0.20	0.05	-0.10
	β_1	-0.07	0.04	-0.06	0.01	-0.01	0.03	0.07	0.07	-0.05
	σ_0^2	0.14	0.00	0.25	0.38	0.31	0.32	0.62	0.01	-0.03
	σ_{01}	0.16	-0.06	-0.09	0.16	0.22	0.17	0.31	-0.03	0.05
	σ_1^2	0.07	0.13	-0.14	0.01	0.12	-0.01	0.02	-0.02	0.00

Summary

When the error distributions were misspecified, there was no noticeable influence on the parameter estimates of the GCMs. If the normal error distribution was misspecified as one of the three non-normal distributions discussed in the study, its influence on the standard error estimates was also negligible. However, if the non-normal error distributions were misspecified as normal distributions, the loss in the efficiency of the standard error estimates can be large. In addition, by correctly specifying the error distribution, one can obtain more efficient results than the bootstrap method.

Discussion

Growth curve models have been widely used in social and behavioral sciences. However, typical GCMs often assume that the errors are normally distributed. The assumption is further carried by both commercial and free software for growth curve analysis and makes it less flexible for non-normal data analysis although non-normal data may be even more common than normal data (Micceri, 1989). As already shown by the literature, blindly assuming normality may lead to severe statistical inference problems (Yuan et al., 2004; Yuan & Zhang, 2012a).

In this study, we proposed a general Bayesian framework to flexibly model normal and non-normal data through the explicit specification of the error distributions. Within the framework, the error distribution is first explored through exploratory data analysis techniques. Then, a statistical distribution that best matches the error distribution is used in growth curve modeling. Bayesian methods are then

used to estimate the model parameters. Although we have focused on the discussion of the *t*-distribution, exponential power distribution, and skew normal distribution, almost any distribution can be applied to the error in growth curve analysis.

Through a simulation study, we found that when the normal error distribution was misspecified as non-normal error distributions discussed in this study, both parameter estimates and their standard errors were very comparable. However, when the non-normal error distributions were misspecified as normal distribution, the loss in the efficiency of the standard error estimates can be large. Therefore, it seemed that moving from normal distribution to those that closely resemble the data is compelling. The similar conclusion is also found in Zhang et al. (2013).

We demonstrated how to conduct growth curve analysis with both normal and non-normal error distributions practically using the MCMC procedure. The normal distribution and *t*-distribution are built-in distributions of the MCMC procedure. Therefore, the extension of the growth curve model with the normal error in the MLE framework to the Bayesian framework is almost effortless. It is also logically natural to extend the models with the normal error to the ones with the *t* error. Although there are no built-in functions for exponential power distribution and skew normal distribution, the MCMC procedure allows the specification of new distributions. From our demonstration, potentially any distribution with known density function can be used to model the error of a GCM. Overall, Bayesian growth curve modeling seems to offer great flexibility in longitudinal data analysis. Although GCMs with non-normal error distributions potentially can be estimated in the MLE framework, we are not aware of easy to use software to carry out such analysis.

For demonstration purposes, we have focused our discussion on the linear GCMs. However, the methods proposed can be equally applied to the nonlinear GCMs (e.g., Grimm et al., 2011). The nonlinear model estimation can also be conducted using the the MCMC procedure.

Given the fact that standard error estimates, not parameter estimates, are influenced by distribution misspecification, one might wonder why not use the MLE with robust standard error or bootstrap standard error (e.g., Yuan & Hayashi, 2006). Based on the likelihood theory, the MLE is most efficient when the distribution is correctly specified (e.g., Casella & Berger, 2001). Asymptotically, Bayesian method is equivalent to MLE especially with uninformative priors (e.g., Gelman et al., 2003). Therefore, one would still expect the Bayesian method will provide more efficient standard error estimates than the robust or bootstrap method based on the misspecified normal error distribution. Our simulation study actually supported this in finding that, overall, modeling the error distribution is more efficient than the bootstrap method.

The current study can be extended in different ways. Firstly, in determining the error distribution, the individual growth curve analysis is used. However, the method assumes that the growth function is known a priori. In practice, this might not be possible. Although it is easy to distinguish a linear growth from a non-linear growth, it can be difficult to choose among several potential non-linear forms. Therefore, more rigorous method can and should be developed to determine the best growth function. Secondly, in exploring the error distribution, it is assumed that the error would follow a certain known distribution. It is likely that one cannot find a good distribution to approximate the error distribution from the individual growth analysis. One potential solution is to use a non-parametric error distribution as discussed by Tong and Zhang (2014). Thirdly, although it is shown that when a normal distribution is misspecified as a non-normal distribution, the loss of efficiency is minimum, it is not clearly what will happen if the error distribution is completely misspecified. For example, what will happen if a skewed normal distribution is misspecified as an exponential power distribution with negative kurtosis. Fourthly, the current study did not discuss how to evaluate a model fit and how to conduct model selection with the new error distributions. Evaluation of a model fit can inform whether a model fits the data adequately and model selection can help decide among more than one competing models. Especially, model selection can also help the choice of the growth function forms and the error distributions. Many criteria for model evaluation are available (e.g., Lu et al., 2015; Spiegelhalter et al., 2002) and can be potentially applied for the current study. Fifthly, the current study assumes that the error can be non-normal but

the random coefficients are still normal. There exist multivariate counterparts for the t , exponential power, and skew normal distributions. However, we are only aware of the extension of the multivariate t -distribution to the model the random coefficients (Tong & Zhang, 2012). Although the above concerns are out of the scope of the current study, that is to introduce a new way to model error distributions in growth curve analysis, they will be investigated in future studies.

Author Note We thank Dr. Dale Barr and three anonymous reviewers for their constructive comments that have significantly improved this research. This study was partially supported by a grant from US Department of Education to Zhiyong Zhang (R305D140037). Correspondence can be sent to Zhiyong Zhang, Department of Psychology, University of Notre Dame, 118 Haggard Hall, Notre Dame, IN 46556. zhiyongzhang@nd.edu.

References

- Anscombe, F.J., & Glynn, W.J. (1983). Distribution of kurtosis statistic for normal statistics. *Biometrika*, 70(1), 227–234.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.
- Box, G.E.P., & Tiao, G.C. (1973). *Bayesian inference in statistical analysis*. New York: John Wiley & Sons.
- Brooks, S.P., & Roberts, G.O. (1998). Assessing convergence of markov chain monte carlo algorithms. *Statistics and Computing*, 8, 319–335.
- Casella, G., & Berger, R.L. (2001). *Statistical inference*, 2nd edn. Pacific Grove, CA: Duxbury Press.
- Chakravarti, I., Laha, R.G., & Roy, J. (1967). *Handbook of methods of applied statistics* (Vol. I, pp. 392–394). New York: John Wiley and Sons.
- Chow, S.M., Tang, N., Yuan, Y., Song, X., & Zhu, H. (2011). Bayesian estimation of semiparametric dynamic latent variable models using the dirichlet process prior. *British Journal of Mathematical and Statistical Psychology*, 64, 69–106.
- Congdon, P. (2003). *Applied Bayesian modelling*. New York: Wiley.
- Cowles, M.K., & Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434), 883–904.
- D’Agostino, R.B. (1970). Transformation to normality of the null distribution of G1. *Biometrika*, 57(3), 679–68.
- Demidenko, E. (2004). *Mixed models: Theory and applications*. New York: Wiley.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2003). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford: Clarendon Press.
- Grimm, K.J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, 82, 1357–1371.
- Kruschke, J.K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press.
- Laird, N.M., & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.

- Lange, K.L., Little, R.J.A., & Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408), 881–896.
- Lee, S.Y. (2007). *Structural equation modeling: A Bayesian approach*. New York: Wiley.
- Lu, Z., Zhang, Z., & Cohen, A. (2015). Model selection criteria for latent growth models using Bayesian methods. In R.E. Millsap, D.M. Bolt, L.A. van der Ark, & W.-C. Wang (Eds.), *Quantitative psychology research*, (Vol. 89 pp. 319–341). New York: Springer International Publishing. doi:10.1007/978-3-319-07503-7_21
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.
- Maronna, R.A., Martin, R.D., & Yohai, V.J. (2006). *Robust statistics: Theory and methods*. New York: Wiley.
- McArdle, J.J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16, 163–200.
- McArdle, J.J. (2000). Structural equation modeling: Present and future. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.) (pp. 342–380). Lincolnwood: Scientific Software International.
- McArdle, J.J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Psychology*, 58, 110–133.
- 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.
- McArdle, J.J., & Nesselroade, J.R. (2003). Growth curve analysis in contemporary psychological research. In J. Schinka, & W. Velicer (Eds.), *Comprehensive handbook of psychology: Research methods in psychology*, (Vol. 2 pp. 447–480). New York: Wiley.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166.
- Muthén, B., & Asparouhov, T. (2012). Bayesian sem: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335.
- Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, 32(7), 685–694.
- Preacher, K.J., Wichman, A.L., MacCallum, R.C., & Briggs, N.E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.
- Robert, C.P., & Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.
- SAS. (2009). *SAS/STAT 9.2 user's guide: The MCMC procedure*, 2nd edn. Cary: Computer software manual.
- Song, X.Y., & Lee, S.Y. (2012). *Basic and advanced bayesian structural equation modeling: With applications in the medical and behavioral sciences*. New York: Wiley.
- Song, X.Y., Lee, S.Y., & Hser, Y.I. (2009). Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects. *Structural Equation Modeling*, 16, 245–266.
- Song, X.Y., Lu, Z.H., Hser, Y.I., & Lee, S.Y. (2011). A Bayesian approach for analyzing longitudinal structural equation models. *Structural Equation Modeling*, 18, 183–194.
- 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.
- Tanner, M.A., & Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528–540.
- Tong, X., & Zhang, Z. (2012). Diagnostics of robust growth curve modeling using Student's t distribution. *Multivariate Behavioral Research*, 47, 493–518.
- Tong, X., & Zhang, Z. (2014). Semiparametric Bayesian modeling with application in growth curve analysis. *Multivariate Behavioral Research*, 49, 299–299.
- Tucker, L.R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1), 19–23.
- Wang, L., & McArdle, J.J. (2008). A simulation study comparison of Bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling*, 15, 52–74.
- Wang, L., Zhang, Z., McArdle, J.J., & Salthouse, T.A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, 43, 476–496.
- Yuan, K.H., Bentler, P.M., & Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika*, 69, 421–436.
- Yuan, K.H., & Hayashi, K. (2006). Standard errors in covariance structure models: A symptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, 59, 397–417.
- Yuan, K.H., & Zhang, Z. (2012a). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, 77, 803–826.
- Yuan, K.H., & Zhang, Z. (2012b). Structural equation modeling diagnostics using R semdiag package and EQS. *Structural Equation Modeling*, 19, 683–702.
- Zhang, Z. (2013). Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics*, 40, 1779–1795.
- 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., Lai, K., Lu, Z., & Tong, X. (2013). Bayesian inference and application of robust growth curve models using Student's t distribution. *Structural Equation Modeling*, 20, 47–78.
- Zu, J., & Yuan, K.H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research*, 45, 1–44.