Running head: DIAGNOSTICS OF ROBUST GROWTH CURVE MODELING

Diagnostics of Robust Growth Curve Modeling using Student's t Distribution

Xin Tong and Zhiyong Zhang

University of Notre Dame

Abstract

Growth curve models with different types of distributions of random effects and of intraindividual measurement errors for robust analysis are compared. After demonstrating the influence of distribution specification on parameter estimation, three methods for diagnosing the distributions for both random effects and intraindividual measurement errors are proposed and evaluated. The methods include (1) distribution checking based on individual growth curve analysis; (2) distribution comparison based on Deviance Information Criterion (DIC); and (3) post hoc checking of degrees of freedom estimates for t distributions. The performance of the methods is compared through simulation studies. When the sample size is reasonably large, the method of post hoc checking of degrees of freedom estimates works best. A web interface is developed to ease the use of the three methods. Application of the three methods is illustrated through growth curve analysis of mathematical ability development, using data on the Peabody Individual Achievement Test (PIAT) Mathematics assessment from the National Longitudinal Survey of Youth 1997 Cohort (Bureau of Labor Statistics, U.S. Department of Labor, 2005).

Diagnostics of Robust Growth Curve Modeling using Student's t Distribution

Growth curve modeling is one of the most frequently used analytic techniques for longitudinal data analysis with repeated measures (e.g., McArdle, 1988; Meredith & Tisak, 1990). A growth curve model generally consists of fixed effects and random effects where the random effects account for the interindividual variation. In traditional growth curve analysis, it is typically assumed that the random effects and intraindividual measurement errors are normally distributed. Although the normality assumption makes growth curve models easy to estimate and apply, data in social and behavioral science usually violate such an assumption (Micceri, 1989). Practically, data often have longer-than-normal tails and/or outliers. Ignoring the nonnormality of data may result in unreliable parameter estimates, their associated standard errors estimates, and thus, misleading tests and inference (Maronna, Martin, & Yohai, 2006; Yuan, Bentler, & Chan, 2004; Zu & Yuan, 2010). Some routine methods, such as deleting the outliers, may lead to serious problems (e.g., Lange, Little, & Taylor, 1989; Yuan & Bentler, 2002). For example, the resulting inferences can fail to reflect uncertainty in the exclusion process and the efficiency can be reduced.

In the past half century, researchers have become more keenly aware of the rather large influence that nonnormality has upon model estimations (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Huber, 1981) and have developed what are called robust methods aiming to provide reliable parameter estimates and inference when the normal distribution assumption is violated. The basic idea of the robust method is to assign a weight to each case according to its distance from the center of the majority of the data, so that the extreme cases can be downweighted (e.g., Huber, 1981; Zhong & Yuan, 2010; Yuan, Bentler, & Chan, 2004). A few studies have directly discussed the robust methods in growth curve analysis. For example, Pendergast & Broffitt (1985) and Singer & Sen (1986) proposed robust estimators based on M-methods for growth curve models with elliptically symmetric errors, and Silvapulle (1992) further extended the M-method to allow asymmetric errors for growth curve analysis. Recently, Zhang, Lai, Lu, & Tong (2012) suggested modeling heavy-tailed data and outliers using Student's t distributions and provided online software to conduct the robust analysis.

The use of the t distribution for describing a single sample dates back at least to Jeffreys (1939). A lot of work has been done to boost its application (Hampel et al., 1986; Little, 1988; Sutradhar & Ali, 1986). The wide adoption of the t distribution in robust statistical modeling was advanced by Lange, Little, & Taylor (1989). Lately, the t distribution has been applied in more complex data analysis, such as robust structural equation models (Lee & Xia, 2006, 2008), linear mixed-effects models (Pinheiro, Liu, & Wu, 2001; Song, Zhang, & Qu, 2007), and robust mixture modeling (Wang, Zhang, Luo, & Wei, 2004; Shoham, 2002). The advantages in using t distribution for robust data analysis are obvious: a) it has a parametric form and inference based on it can be carried out through maximum likelihood methods and Bayesian methods relatively easily; b) the degrees of freedom of t distributions control the weight of outliers and can be either set *a priori* or estimated; c) robust methods based on t distributions are easy to understand and can be considered as a direct extension of corresponding normal distribution methods for heavy-tailed data. Zhang et al. (2012) demonstrated that the robust growth curve modeling based on t distributions could be easy to understand and implement, and thus potentially would greatly promote the adoption of robust growth curve analysis.

In growth curve analysis, both the random effects and intraindividual measurement errors are equally likely to follow t distributions instead of normal distributions. However, Zhang et al. (2012) only focused on the case where the measurement errors were modeled by t distributions, although their software allows for simultaneously modeling both the random effects and the measurement errors by t distributions. Pinheiro et al. (2001) proposed a robust version of the linear mixed-effect model, in which the normal distributions for the random effects and the intraindividual errors were both replaced by multivariate t distributions. In fact, given a set of longitudinal data, one could fit one of the four possible types of distributional growth curve models to the data with (1) normal measurement errors and normal random effects, (2) t measurement errors and normal random effects, (3) normal measurement errors and t random effects, and (4) t measurement errors and t random effects. However, no study has been conducted to evaluate and compare the intrinsic characteristics of the four distributional models and no strategy has been developed to assist in choosing the appropriate model for a given set of data.

The purpose of this study is to evaluate and compare the four types of distributional growth curve models and develop strategies to select the appropriate model for growth curve analysis. In the following, we first illustrate the features of the four different types of distributional growth curve models. Then, we compare the four models in modeling different types of data through a simulation study. After the comparison, we develop three methods for selecting appropriate distributions for the measurement errors and random effects. Simulation studies are also carried out to demonstrate the effectiveness of those methods. Finally, we illustrate the use of the methods through an example with the real data from the National Longitudinal Survey of Youth (NLSY) 1997 Cohort. We end the paper with the discussion of the differences among the three methods.

Robust Growth Curve Models

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be a $T \times 1$ random vector and y_{ij} be an observation for individual i at time j ($i = 1, \ldots, N; j = 1, \ldots, T$). Here N is the sample size and T is the total number of measurement occasions. One form of growth curve models can be expressed as

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

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

Traditional growth curve models typically assume that both $\mathbf{e}_i$ and $\mathbf{u}_i$ follow multivariate normal distributions such that $\mathbf{e}_i \sim MN_T(0, \sigma^2 \mathbf{I}_{T\times T})$ and $\mathbf{u}_i \sim MN_q(0, \mathbf{D}_{q\times q})$, where MN denotes a multivariate normal distribution and the subscription denotes its dimension. Note that it is assumed that the errors are identically and independently distributed over time with mean 0 and variance σ^2 but this assumption can be relaxed. $\mathbf{D}$ is a $q \times q$ square matrix which represents the variance-covariance matrix of $\mathbf{u}_i$. Given the current specification of $\mathbf{u}_i$, $\mathbf{b}_i \sim MN_q(\boldsymbol{\beta}, \mathbf{D}_{q\times q})$. For convenience, we denote this model with normal errors and normal random effects as the Normal-Normal (N-N) distributional model.

Although the N-N distributional growth curve models are widely used, practical data often violate these normal assumptions. Zhang et al. (2012) proposed to model the measurement errors using the t distribution by letting $\mathbf{e}_i \sim MT_T(0, \sigma^2 \mathbf{I}_{T\times T}, df_e)$ and $\mathbf{u}_i \sim MN_q(0, \mathbf{D}_{q\times q})$ where MT_T denote a T-dimension multivariate t distribution. The resulting robust growth curve models have been shown to perform better in analyzing data with long-tailness and outliers than the N-N distributional model. Another type of robust growth curve models can be derived by letting $\mathbf{e}_i \sim MN_T(0, \sigma^2 \mathbf{I}_{T\times T})$ and $\mathbf{u}_i \sim MT_q(0, \mathbf{D}_{q\times q}, df_u)$. The third type of robust growth curve model can be obtained by letting $\mathbf{e}_i \sim MT_T(0, \sigma^2 \mathbf{I}_{T\times T}, df_e)$ and $\mathbf{u}_i \sim MT_q(0, \mathbf{D}_{q\times q}, df_u)$. We denote the three types of robust growth curve models as the T-N distributional model, the N-T distributional model, and the T-T distributional model, respectively.

To illustrate the differences among the N-N, T-N, N-T, and T-T distributional models, we generate and plot data from the four types of models (Figure 1). For each type of model, data on 20 subjects are generated for 4 occasions assuming a linear growth trend. Figure 1(a) displays the trajectories of the data generated by the N-N distributional model. No outlier can be observed. The overall trajectory looks clean and smooth. Figure 1(b) plots the data generated by the T-N distributional model. Noticeably, some observations stand out of the overall trajectory such as those labeled by 1 and 2. A close look at the two observations reveals that the reason why they deviate from the overall trajectory is that they are off their own growth trajectories. For example, at a given time, an individual performed much better than expected on a test. Figure 1(c) portrays data from the N-T distributional model. Some observations also deviate from the overall growth trajectory. However, it seems that those observations are still on their own growth trajectories. The reason why they stand out is that the rate of growth for the specific case is very different from the majority of cases. For example, for a participant, he or she could display an unusual rate of growth during the period of data collection. Figure 1(d) draws the trajectories for the data from the T-T distributional model. Clearly, the outlying observations are due to two sources - the trajectory of a case deviates from the overall trajectory and the observation for this specific case is off its own trajectory. For example, the observation 1 stands out because it is off the trajectory of the case and the case itself has a higher initial level.

In summary, Figure 1 suggests that the four types of distributional growth curve models can imply very different patterns in growth trajectories. For instance, if an individual's growth trajectory is within the normal range of the overall trajectory and an observation at certain times stands out, the data are more likely to come from the T-N distributional model. If, within an individual, his/her observations follow a smooth pattern but the trajectory itself differs from the overall trajectory, the data are more likely to come from the N-T distributional model. Therefore, given an empirical data set, it can be very important to specify the correct type of growth curve models. In order to concretely demonstrate the possible adverse effects of misspecification for finite samples, we will conduct a simulation study in the next section.

Influences of distribution misspecification

In this section, we conduct a simulation study to evaluate the effects of the misspecification of the four types of distributional growth curve models. In the simulation, we first generate data from the N-N, T-N, N-T, and T-T distributional models, respectively. The data are called N-N data, T-N data, N-T data, and T-T data, corresponding to the type of the models. Then, for each type of data, we fit all four types of models and compare their parameter estimates.

To estimate these models, we use a Bayesian method, the use of which for complex data analysis has become popular in the past two decades (Lee & Shi, 2000; Lee, 2007; Lee & Song, 2008; Lee & Xia, 2008). Zhang et al. (2012) applied the Bayesian method to the robust growth curve models with Student's t distributed measurement errors and normally distributed random coefficients. We introduce this method into our model estimation for all the four types of distributional models. The basic idea of the Bayesian method is to obtain the posterior distributions of model parameters based on the likelihood function and the prior distributions. Details are given in appendix A of Zhang et al. (2012).

The population parameters for the growth curve models in the simulation are given by

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ or \ \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \ or \ \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix},$$

and $\sigma^2 = 1$. From the factor loading matrix, the population growth trajectory follows a linear pattern. Furthermore, for the T-N distributional model, we let $df_e = 4$, for the N-T distributional model, $df_u = 4$, and for the T-T distributional model, both df_e and df_u are equal to 4. We vary the sample size of the simulation with $N = 50, 100, 300$ and 500 and the number of measurement occasions with $T = 4$, 5 and 10. Overall, 36 conditions of simulation are investigated.[1]

For each simulation study, a total of 1000 data sets are generated and analyzed. Two commonly used statistics, which examine more than one performance criterion (Collins, Schafer, & Kam, 2001), are calculated for each model parameter to compare the four types of growth curve models. The first statistic is the mean square error (MSE) based on 1000 sets of parameter estimates and standard errors, and the second one is the coverage probability (CP) of the 95% credible intervals[2]. The MSE and CP are then averaged over all model parameters (except the degrees of freedom) for each simulation condition. Table 1 summarizes the results for the analysis of each type of data by different types of distributional models with different sample sizes. In the table, on the rows are the different types of generated data and on the columns are the four types of distributional models used to analyze the generated data.

In almost all situations, the model used to generate the data provides the best estimations with smaller MSE and better credible interval coverage among the four types of growth curve models. For example, for the T-T data with N=300, the T-T distributional model gives the smallest MSE and the best coverage probability. Although in Table 1, the T-T distributional model seems to obtain very good results even when the data-generating model is not T-T, it is not always a good idea to simply apply the T-T distributional model to all data.

First, theoretically, although the t distribution approaches the normal distribution as the degrees of freedom increase, the degrees of freedom estimates of the t distribution will never be infinity when we fit a t distribution to normal data. Because the degrees of freedom parameter controls the kurtosis of the distribution, the estimates of kurtosis will be larger than the true kurtosis, which is the kurtosis of the normal distribution. Yuan, Bentler, & Zhang (2005) proved that with a larger kurtosis, the variance of the model parameter estimates would be larger. In this case, the statistical power will be low and the model efficiency will be reduced, as demonstrated in Zhang et al. (2012). Therefore, one will lose statistical power and efficiency by fitting the T-T distributional model to the N-N, T-N, or N-T data.

Second, numerically through our simulation studies, the T-T distributional growth curve model is not optimal all the time. From Table 1, as the sample size decreases, the MSE becomes larger and the coverage probability becomes further away from its nominal level 0.95, meaning that the accuracy of the T-T distributional model is getting worse. To better illustrate the influence of sample size, Table 2 provides the parameter estimates for the T-T distributional model when the sample size is 50. Although the fixed effects parameter estimates for the initial level and rate of growth are very close to their true values, other parameters such as the estimate for the variability of slope are largely biased. This is due to the difficulty of getting accurate estimates for the degrees of freedom of the t distributions. When the sample size is small, it is more likely to identify a t distribution as a normal distribution. Thus, the degrees of freedom estimates are positively biased, and accordingly, the other parameter estimates may be inaccurate. To solve this problem, it is suggested to supply a fixed value as the degree of freedom of the t distribution (Zhang et al., 2012; Lange et al., 1989). Namely, it is necessary to select the right model type at the first step. Then, we can fix the degrees of freedom based on that model type and estimate the other parameters in order to obtain more accurate parameter estimates and credible intervals.

Third, practically, estimating a T-T distributional model is more time consuming than other types of models. Especially, the multimode problem often occurs in estimating the t distribution (Box & Tiao, 1973), and a long Markov chain is often needed to ensure accurate parameter estimates. Therefore, it is often worth putting effort into determining the distributions of random effects and measurement errors.

In summary, the accuracy and efficiency of the estimation for a specific type of data closely depends on the correct specification of a model, especially when the sample size is small. Thus, in practical data analysis, it is important to choose the correct type of model.

Methods for distribution diagnostics

In the previous section, we have shown that the distribution misspecification influences both parameter estimates and credible interval coverage probabilities. To address this misspecification problem, we propose and compare three methods to aid in selecting appropriate distributions for measurement errors and random effects. The three methods are (1) distribution checking based on individual growth curve analysis; (2) distribution comparison based on Deviance Information Criterion (DIC); and (3) post hoc checking of degrees of freedom estimates. In the following, we first describe each method and then evaluate its performance through simulation studies. In the simulation studies, we use the same specifications as in the previous section to generate four types of data (N-N data, T-N data, N-T data, T-T data) from four types of models (N-N distributional model, T-N distributional model, N-T distributional model, T-T distributional model) with sample sizes of 50, 100, 300, and 500. Then we evaluate whether the proposed methods can correctly identify the correct models and distributions.

Distribution checking based on individual growth curve analysis

In this method, an individual growth curve ($\mathbf{y}_i = \mathbf{\Lambda} \cdot \mathbf{b}_i + \mathbf{e}_i$) is first fitted to data from each individual through the least squares method. The individual coefficients (random effects) $\mathbf{b}_i = (b_{i0}, \ldots, b_{iq})^T$ and the measurement errors $\mathbf{e}_i = (e_{i1}, \ldots, e_{iT})^T$ are estimated and retained. Let $\mathbf{b} = (\hat{\mathbf{b}}_1, \cdots, \hat{\mathbf{b}}_N)^T$ and $\mathbf{e} = (\hat{\mathbf{e}}_1, \cdots, \hat{\mathbf{e}}_N)^T$ where $\mathbf{b}$ is a $N \times q$ matrix of individual coefficients estimates and $\mathbf{e}$ is a $N \times T$ matrix of estimated errors. Then, we test whether $\mathbf{e}$ and $\mathbf{b}$ follow normal distributions. If all q columns of $\mathbf{b}$ follow normal distributions, we consider the individual coefficients to be normally distributed. Otherwise, we consider them nonnormally distributed. Similarly, if all T columns of $\mathbf{e}$ are normally distributed, the errors are viewed as from normal distributions. If $\mathbf{e}$ and $\mathbf{b}$ are not normally distributed, t distributions are recommended. Based on the combination of the distributions for $\mathbf{e}$ and $\mathbf{b}$, the decision can be made according to Table 3.

Note that four null hypotheses are tested in order to select a model. They are: (1) both measurement errors $\mathbf{e}$ and individual coefficients $\mathbf{b}$ are normallly distributed, (2) $\mathbf{e}$ is normal while $\mathbf{b}$ is nonnormal, (3) $\mathbf{e}$ is nonnormal and $\mathbf{b}$ is normal, and (4) both $\mathbf{e}$ and $\mathbf{b}$ are nonnormal. Among these four tests, only three are independent because if it is certain that one data set is not generated from three of the models, it has to be generated from the remaining model. Thus, a Bonferroni correction should be used for multiple testing. The new alpha level should be $\alpha/3 = 0.05/3 \doteq 0.017$.

Two tests are employed in this study to check the normality of random coefficients and errors – the Shapiro-Wilk normality test (Shapiro & Wilk, 1965) and the Anscombe-Glynn test of kurtosis (Anscombe & Glynn, 1983). The Shapiro-Wilk normality test checks the normality based on both skewness and kurtosis. If data exhibit either skewness or kurtosis, the normality is rejected. It is applied here because its test statistic exhibits sensitivity to nonnormality over a wide range of alternative distributions and in most cases it has power as good as or better than many other normality test statistics (Shapiro et al., 1968). The Anscombe-Glynn test checks whether the kurtosis for the current data is larger than the kurtosis of normal data. It is good at detecting nonnormality caused by nonnormal tail heaviness. If a distribution is symmetric but heavy-tailed, the test for kurtosis may be more powerful than the Shapiro-Wilk test. The Anscombe-Glynn test is used in the simulation study because of the heavy-tailed property of our generated data.

Table 4 summarizes the proportions of times that a model was selected based on the two tests for different types of data and sample sizes. For example, the last row of the table tells us that when the data (T-T data) are generated from a T-T distributional model with a sample size 100, the Shapiro-Wilk test selects the T-T distributional model 44.7% of the time and the Anscobme-Glynn test selects the T-T distributional model 49.0% of the time. This means that based on the Shapiro-Wilk test and the Anscobme-Glynn test, one can make the correct decision on the coefficients and errors distributions 44.7% and 49.0% of the time, respectively.

Overall, both tests have a high false positive rate. For example, when the true model is the T-N distributional model, only 41.9% of the time the T-N distributional model is identified and about 58.1% of the time the model is mis-identified as the other three types of models based on the Shapiro-Wilk test with a sample size of 500. For the Anscombe-Glynn test, 64.3% of the time the correct model is not identified. Furthermore, it seems that for different types of data, the tests perform differently. For example, the two tests work best for the N-N data and worst for the T-N data. Finally, the Shapiro-Wilk test works slightly better for the N-N and T-N data than the Anscombe-Glynn test that, on the other hand, works slightly worse for the N-T and T-T data.

A closer look at the results also reveals some useful details. For the N-N data, it is most likely to be mis-classified as being generated from the T-N distributional model and least likely to be mis-classified as being generated from the T-T distributional model. For the T-N data, it is most likely to be mis-classified as the data generated from the T-T distributional model and least likely to be mis-classified as being generated from the N-T distributional model. When the sample size is small, for the N-T data, it is most and least likely to be mis-classified as being generated from the N-N distributional model and the T-N distributional model, respectively. Finally, for the T-T data, it is more likely to be mis-classified as the data generated from the T-N distributional model and the N-T distributional model than the N-N distributional model. The results are consistent for both the Shapiro-Wilk test and the Anscombe-Glynn test.

Distribution comparison based on Deviance Information Criterion (DIC)

The Deviance Information Criterion (DIC, Spiegelhalter, Best, Carlin, & Linde, 2002a) is a widely used criterion for model selection in the Bayesian framework. It is defined as a Bayesian measure of goodness of model fit with a penalty for model complexity,

$$DIC = \overline{D\left(\boldsymbol{\Theta}\right)} + p_D = D\left(\bar{\boldsymbol{\Theta}}\right) + 2p_D$$

where $\boldsymbol{\Theta}$ represents a vector of all the unknown parameters in the model, $\overline{D\left(\boldsymbol{\Theta}\right)}$ is the posterior mean of -2(LogLikehood function) and $D\left(\bar{\boldsymbol{\Theta}}\right)$ is -2(LogLikehood function) calculated at the posterior mean of $\boldsymbol{\Theta}$. The complexity measure, p_D, is defined as the difference between the posterior mean of deviance and the deviance evaluated at the posterior mean of the parameters. DIC can be used in the same way as AIC and BIC. The model with the minimum DIC will make the best short-term predictions and thus indicates the best model among the evaluated ones.

To determine the appropriate distributions for random effects and errors for the given data, models with different combinations of distributions can be fitted to the data. The model with the smallest DIC can be retained and the corresponding distributions for random coefficients and errors can be considered to be reasonable. In order to evaluate the performance of DIC in determining the correct distributions, a simulation is conducted and the results are given in Table 5. In the simulation, all the four types of distributional models are applied to the same set of data and the DIC values are obtained and compared. The proportions of the model with the smallest DIC are calculated and provided.

The results in Table 5 clearly suggest that the performance of DIC in determining the random effects and error distributions depends on both the sample size and the underlying true models. With sample sizes 300 and 500, the DIC method works very well for the N-T and T-T data because the rate of correctly specifying a model is close to 1. However, the same method does not seem to work well for N-N and T-N data especially for the simulation of sample size 100. A closer look at the results reveals that DIC has trouble correctly identifying the distribution for random coefficients. To be specific, for N-N data, DIC has a large chance to select the N-T distributional model and for T-N data, it has a tendency to select the T-T distributional model.

Post hoc checking of degrees of freedom estimates

As the degrees of freedom increase, a t distribution approaches a normal distribution. Thus, if a t distribution is fitted to a set of data and the estimated degree of freedom is large, it indicates that a normal distribution may be good enough to model the data. For the robust growth curve modeling, the same strategy can be applied. Specifically, the T-T distributional model can be fitted to the data, and the degrees of freedom for random effects and errors can be estimated. If both df_u and df_e are larger than a certain cutoff (e.g., 30), it indicates that $\mathbf{u}$, so does $\mathbf{b}$, and $\mathbf{e}$ follow normal distributions. Thus, the N-N distributional model can be used. If df_u is larger than the cutoff while df_e is not, it implies that $\mathbf{u}$ follows a normal distribution and $\mathbf{e}$ follows a t distribution. Accordingly, the T-N distributional model is the appropriate one to select. In the opposite, if df_e is larger than the cutoff while df_u is not, the N-T distributional model should be selected. Finally, if both df_u and df_e are smaller than the cutoff, the T-T distributional model should be applied.

To evaluate the performance of the method through simulation study, we fit the T-T distributional model to all the generated data and compare the degrees of freedom for random effects and errors to a given cutoff. Because there is no universally accepted cutoff, we investigate several levels of cutoff at 20, 25, 30 and 45. The results for the simulation are summarized in Table 6. In the table, the proportions to select each type of model based on the degrees of freedom are presented.

Overall, a larger cutoff leads to a higher likelihood of accepting $\mathbf{u}$ and $\mathbf{e}$ to follow the t distribution than the normal distribution as shown in the table. The post hoc degrees of freedom estimates checking method seems to be very effective especially when the sample size is large, for example, $N \geq 300$. Furthermore, a cutoff of 30 seems to perform well under most conditions. For example, when the sample size is as large as 300, more than 98% of times the correct distributions can be identified. When the sample size is small (e.g. $N \leq 100$), its performance is not as good as for the large sample size situation but still comparable to or better than the distribution checking and DIC methods. For the small sample size situation, increasing the cutoff tends to improve the performance of the method in identifying the T-N, N-T, and T-T distributional models but hinder the performance in identifying the N-N distributional model.

Implementation

In order to facilitate the application of the distribution diagnostics methods, we implement them in free software R (R Development Core Team, 2011) and OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009). A web interface is also developed to ease the use of the three methods. An end user is only required to upload a set of longitudinal data to be analyzed. The web interface can be accessed at `http://webstats.psychstat.org/wiki/m01/robdiag`. The interface for the implementation is shown in Figure 2.

To use the software online, one first uploads the data file in which the data are organized as a matrix with data from the first occasion on the first column, data from the second occasion on the second column, and so on. Then, one of the three methods can be selected. For the second method, one can supply a non-negative number as the degrees of freedom. The default 0 requests the estimation of the degrees of freedom. The burn-in period and the length of Markov chain can be controlled by supplying them in the respective fields.

In the next section, the example of real data analysis is carried out using this online software. The format of the original output can be seen on our website.

An Example

To demonstrate the application of the three methods for diagnosing distributions for random effects and measurement errors, we investigate a subset of data from the National Longitudinal Survey of Youth (NLSY) 1997 Cohort (Bureau of Labor Statistics, U.S. Department of Labor, 2005). A sample of N = 512 school children was administered the Peabody Individual Achievement Test (PIAT) Mathematics subset yearly from grade 7 to grade 10. The individuals' trajectory plot (Figure 3) suggests a linear growth pattern and a T-N distributional model because the pattern of the figure is similar to the pattern in Figure 1(b). The following illustration is to determine the best distributions for the random coefficients and measurement errors for the linear growth curve model.

First, a simple linear regression, $y_{it} = b_{i0} + b_{i1}t + e_{it}$, is fitted to data of each individual, and the individual regression coefficients $\hat{b}_{i0}$ and $\hat{b}_{i1}$ and measurement (residual) errors $\hat{e}_{it}$ are estimated for $t = 1, \ldots, 4$ and $i = 1, \ldots, 512$. The histograms and Q-Q plots for the errors are portrayed in Figure 4. Both histograms and Q-Q plots indicate that the errors have longer-than-normal tails. The results from the Shapiro-Wilk normality test and the Anscombe-Glynn test of kurtoses are given in Table 7(a). Based on the Shapiro-Wilk normality test, the normality of the errors at all grades are rejected. Furthermore, the kurtoses from all grades are significantly larger than the normal kurtosis, ranging from 4.536 to 10.270 from grade 7 to grade 10. Thus, it seems more reasonable to specify t distributions for the errors than normal distributions. Figure 5 presents the histograms and Q-Q plots for the random coefficients and Table 7(b) summarizes the results of the Shapiro-Wilk test and the Anscombe-Glynn test. From either the plots or the tests, both random coefficients appear to be normally distributed. Overall, it seems that the T-N distributional model is a good choice for the current sets of data based on the checking of the distributions of the errors and random coefficients.

Second, to select appropriate distributions for random coefficients and errors based on DIC, we fit the four distributional models to data and obtain the DIC for each model. The DICs for the

N-N, T-N, N-T and T-T distributional models are 5752.711, 5487.288, 5753.621, and 5487.579, respectively. Overall, the T-N distributional model has the smallest DIC, meaning that it fits the data the best. Although the DIC of the T-T distributional model is very close to that of the T-N distributional model, the selection of the T-N distributional model is decisive because our simulation results in Table 5 show that it is unlikely that a T-T distributional model is mistaken as a T-N model using DIC when the sample size is as large as 500.

Finally, we check the magnitude of the degrees of freedoms for the random coefficients and errors by fitting a T-T distributional model to the data. The estimate of df_u for random coefficients is 59.281 and the estimate of df_e for errors is 4.172. Given the large sample size (N>500), a cutoff of 30 seems to give reliable results based on our simulation results in Table 6. Because df_u is larger and df_e is smaller than the cutoff, it is reasonable to believe that random coefficients follow normal distributions and the errors follow t distributions. Thus, the T-N distributional model is the right one to choose.

To summarize, all three methods lead to the same conclusion that the T-N distributional model with normally distributed random coefficients and t distributed errors fits the data best in this example. This is consistent with the pattern we observe in Figure 3. Individuals' growth trajectoryies are within the normal range of the overall trajectory, while some observations at certain times stand out of their own developmental pathways. The parameter estimates for the T-N distributional model are given in Table 8. To ensure convergence, we have calculated the Geweke statistics (Geweke, 1992) for all parameters. All Geweke statistics are smaller than 2, suggesting the convergence of the Markov chains. Furthermore, for all parameters, the Monte Carlo errors are smaller than 5% of their corresponding standard deviations, indicating accurate parameter estimates. Based on the results, the average initial mathematical ability at grade 7 is about 6.153 with an average total change of 0.319 from grade 7 to grade 10. By conducting z-tests on D_{11} and D_{22}, it is found that there are also significant individual differences in both initial ability and change. Compared to the parameter estimates using the N-N distributional model, it is found that the estimate for D_{12}, the covariance between the intercept and the slope, from the robust T-N distributional model is substantially different from that of the N-N distributional model, as it is significant for the T-N distributional model and insignificant for the N-N distributional model. Because the correlation between the intercept and slope parameters across sessions are interested in many studies (e.g. Zhang, Davis, Salthouse, & Tucker-Drob, 2007), the N-N distributional model should not be used as a substitute.

Discussion

Although robust growth curve models based on t distributions have been developed, no study has discussed possible ways to diagnose the distributions for the random effects and intraindividual measurement errors in the models. To fill the gap, we evaluated three methods to select appropriate distributions for both random effects and errors including (1) distribution checking based on individual growth curve analysis; (2) distribution comparison based on Deviance Information Criterion (DIC); and (3) post hoc checking of degrees of freedom estimates. The performance of each method was evaluated through simulation studies. The application of the three methods was also illustrated through a real data example.

The method of distribution checking based on individual growth curve analysis has many advantages. First, it is easy to implement. The random effects and errors can be obtained through the OLS method. Second, the distributions of random effects and errors can be displayed visually through histograms and Q-Q plots. The normality of random effects and errors can be further tested through the available methods such as the Shapiro-Wilk test and the Anscombe-Glynn test. Third, the diagnostics of the distributions are not confined to the t distributions and can be extended to any other distribution. Fourth, the use of this method usually does not require a large sample size. The main disadvantage of this method is that the distribution diagnostics based on the Shapiro-Wilk test and the Anscombe-Glynn test do not perform very well for some types of data under certain conditions. For example, the T-T data can be mistakenly classified into other three types of data easily when the sample size is small (e.g. $N \leq 100$). Not correcting the type I error for the multiple testing problems improves the results for the data with a small sample size, because correction makes it more difficult to reject the null hypothesis and more likely to conclude that distributions are normal.

The distribution comparison method based on DIC is essentially a model comparison method. This method requires the fit of four distributional models in order to determine the distributions for random effects and errors. It is straightforward and can also generate model parameter estimates when obtaining DIC. The DIC method works very well for the N-T distributional model and the T-T distributional model with a reasonably large sample size. For the small sample size problem, it is not recommended to estimate the degrees of freedom for t distributions. Alternatively, the degrees of freedom can be fixed to a small value. We only estimate the rest of the model parameters and compute DIC (Lange et al., 1989). The degree of freedom determines how much one plans to downweight the outliers. If one would downweight more of the outlier, a smaller degree of freedom can be used. Venables & Ripley (1999) suggested that df = 5 is often a good choice.

The method of post hoc checking of degrees of freedom estimates works best among the three methods when the sample size is reasonably large. Compared to the second method, it only involves the fit of the T-T distributional model. Although this method seems to be simpler than the DIC method, it is not allowed to fix the degrees of freedom at a given value. For the small sample size problems, this method may have difficulty in estimating the degrees of freedom for both the random effects and the measurement errors components. Furthermore, it requires the determination of a cutoff. Generally speaking, a cutoff of 30 is recommended based on our simulation study.

To conclude, our aim in this article has been to compare the four types of distributional growth curve models and evaluate the three proposed model selection methods. When the sample size is large, the post hoc checking of degrees of freedom estimates method can be directly used as the best model selection method. When the sample size is relatively small, this method is not suggested anymore. In contrast, the method of distribution checking based on individual growth curve analysis, combined with the distribution comparison method based on DIC, should be used together to make a decision. Note that, to achieve a higher correct selection rate for data with a small sample size, one may consider adopting a less strict alpha level in the first method and using the second method with fixed degrees of freedom.

Limitations

The study was conducted for data with longer-than-normal tailed distributions but not with skewed distributions. Like the normal distribution, the t distribution is symmetric and can be sensitive to skewed data. If the skewness in the data is caused by outliers, the robust growth curve models can still perform very well. However, if the skewness is because of skewed distributions such as the lognormal distribution, robust growth curve models can break down, although they still work relatively better than the normal growth curve models (Zhang et al., 2012). In this case, skew-t distributions may be applied to accommodate the skewness in the data (Azzalini & Genton, 2008; Lachos, Dey, & Cancho, 2009). But the performance of the skew-t distribution in robust growth curve modeling needs further evaluation.

For robust growth curve model selection, we have used a distribution comparison method based on DIC. Although DIC is very widely used, it has received much criticism since it was proposed (Spiegelhalter, Best, Carlin, & Linde, 2002b). Celeux, Forbers, Robert, & Titterington (2006) argued that the DIC introduced by Spiegelhalter et al. (2002b) for model assessment and model comparison was directly inspired by linear and generalized linear models, but it was open to different possible variations in the setting of models involving random effects, as in our robust growth curve models. A number of ways of computing DICs are proposed in Celeux et al. (2006), and their advantages and disadvantages are discussed. Thus, more sophisticated fit indices should be carried out for comparison in the future, since a better computation of DIC may lead to better results of the model selection method.

The study focuses on linear robust growth curve models for demonstration. However, the same methods should work for nonlinear growth curve models as well. The performance of the nonlinear robust growth curve models (e.g. quadratic models) and the proposed model selection methods can be studied in the future.

References

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

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

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Wiley.

Bureau of Labor Statistics, U.S. Department of Labor. (2005). National longitudinal survey of youth 1997 cohort, 1997-2003 (rounds 1-7) [computer file] [Computer software manual]. Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, Ohio.

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

Collins, L., Schafer, J., & Kam, C. (2001). A comparison of inclusive and restrictive missing-data strategies in modern missing-data procedures. *Psychological Methods*, *6*, 330–351.

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

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

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

Jeffreys, H. (1939). *Theory of probability*. Oxford,U.K.: Clarendon Press.

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

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

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

Lee, S. Y., & Shi, J. Q. (2000). Joint bayesian analysis of factor scores and structural parameters in the factor analysis model. , vol. 52, 2000, 722-736. *Annal of the Institute of Statistical Mathematics*, *52*, 722–736.

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

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

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

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

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The bugs project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, *28*, 3049–3082.

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

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

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166.

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

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

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Shapiro, S. S., Wilk, M. B., & Chen, M. H. (1968). A comparative study of various tests for normality. *American Statistical Association Journal*, 1343–1372.

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

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

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

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

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

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

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

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

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

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

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

Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods and Research*, *34*.

Zhang, Z., Davis, H. P., Salthouse, T. A., & Tucker-Drob, E. A. (2007). Correlates of individual, and age-related, differences in short-term learning. *Learning and Individual Differences*, *17(3)*, 231–240.

Zhang, Z., Lai, K., Lu, Z., & Tong, X. (2012). Bayesian inference and application of robust growth curve models using student's t distribution. *Structural Equation Modeling*, *(in press)*.

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

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

Footnotes

[1]Only the results for 4 measurement occasions will be reported. The results under other conditions are similar and are omitted for the sake of saving space. Those results are available on request or can be found at `http://webstats.psychstat.org/models/lgcm/robdiag/MBR.supp.zip`.

[2]The posterior credible interval in Bayesian analysis, also called credible interval or Bayesian confidence interval, is analogical to the frequentist confidence interval.

Table 1

Mean squared errors and coverage probabilities for different data and models ($T = 4$, $\mathbf{D} = diag(1, 1)$)

N			N-N Model	T-N Model	N-T Model	T-T Model
500	N-N data	MSE	**0.005**	0.005	0.006	0.006
		CP	**0.948**	0.943	0.944	0.938
	T-N data	MSE	0.021	**0.009**	0.021	0.009
		CP	0.874	**0.943**	0.854	0.938
	N-T data	MSE	0.066	0.032	**0.008**	0.008
		CP	0.836	0.827	**0.951**	0.942
	T-T data	MSE	0.072	0.043	0.017	**0.011**
		CP	0.808	0.839	0.893	**0.945**
300	N-N data	MSE	**0.008**	0.009	0.009	0.009
		CP	**0.95**	0.948	0.948	0.945
	T-N data	MSE	0.037	**0.016**	0.036	0.016
		CP	0.887	**0.944**	0.876	0.936
	N-T data	MSE	0.084	0.051	**0.014**	0.015
		CP	0.848	0.85	**0.951**	0.949
	T-T data	MSE	0.114	0.062	0.029	**0.021**
		CP	0.815	0.86	0.911	**0.945**
100	N-N data	MSE	**0.026**	0.025	0.025	0.025
		CP	**0.951**	0.956	0.95	0.952
	T-N data	MSE	0.073	**0.054**	0.062	0.054
		CP	0.901	**0.919**	0.905	0.918
	N-T data	MSE	0.532	0.153	**0.058**	0.058
		CP	0.861	0.87	**0.936**	0.938
	T-T data	MSE	0.304	0.172	0.094	**0.086**
		CP	0.834	0.868	0.912	**0.925**

Note. N: sample size; MSE: mean square error; CP: coverage probability. In the table, on the rows are the different types of generated data with sample size = 500, 300, and 100. On the columns are the four types of distributional models used to analyze the generated data. For each type of the generated data, four distributional models are fitted to them. The average MSE and CP for the model parameters (except the degrees of freedom) are obtained, as displayed in the table.

Table 2

Parameter estimates for the T-T distributional model when sample size is 50, $T = 4$, and $\mathbf{D} = diag(4, 2)$

		True	Est.	Absolute bias	Relative bias(%)	SE	MSE	CP
T-T data	β_0	6.000	6.003	0.003	0.056	0.380	0.144	0.944
	β_1	2.000	1.994	-0.006	-0.284	0.257	0.063	0.952
	D_{11}	4.000	4.810	0.810	20.240	1.687	4.066	0.904
	D_{22}	2.000	2.405	0.405	20.247	0.793	0.873	0.900
	D_{12}	0.000	0.026	0.026	2.574	0.661	0.466	0.927
	σ^2	1.000	1.426	0.426	42.590	0.351	0.322	0.748
	df_u	4.000	23.745	19.745	493.622	19.051	708.921	0.770
	df_e	4.000	29.043	25.043	626.070	21.760	946.364	0.659

Note. Est.: parameter estimates; SE: standard error.

Table 3

Distribution checking based on individual growth curve analysis

Errors	Individual Coefficients	Model
normal	normal	N - N distributional model
nonnormal	normal	T - N distributional model
normal	nonnormal	N - T distributional model
nonnormal	nonnormal	T - T distributional model

Table 4

Distribution diagnostics based on the Shapiro-Wilk test and the Anscombe-Glynn test ($T = 4$, $\boldsymbol{D} = diag(1,1)$). The number in each cell represents the proportion of selecting the fitted model listed at the column heading for the type of generated data listed at the row heading.

N		Shapiro-Wilk test				Anscombe-Glynn test			
		N-N model	T-N model	N-T model	T-T model	N-N model	T-N model	N-T model	T-T model
500	N-N data	**0.898**	0.068	0.032	0.002	**0.911**	0.055	0.032	0.002
	T-N data	0.000	**0.419**	0.000	0.581	0.000	**0.357**	0.000	0.643
	N-T data	0.003	0.000	**0.926**	0.071	0.000	0.000	**0.937**	0.063
	T-T data	0.000	0.002	0.000	**0.998**	0.000	0.000	0.000	**1.000**
300	N-N data	**0.911**	0.057	0.030	0.002	**0.900**	0.060	0.038	0.002
	T-N data	0.011	**0.583**	0.007	0.399	0.007	**0.537**	0.006	0.450
	N-T data	0.022	0.001	**0.933**	0.044	0.012	0.001	**0.922**	0.065
	T-T data	0.001	0.036	0.022	**0.941**	0.000	0.021	0.014	**0.965**
100	N-N data	**0.901**	0.060	0.037	0.002	**0.898**	0.066	0.034	0.002
	T-N data	0.232	**0.585**	0.052	0.131	0.218	**0.586**	0.053	0.143
	N-T data	0.306	0.022	**0.628**	0.044	0.280	0.021	**0.650**	0.049
	T-T data	0.106	0.272	0.175	**0.447**	0.086	0.262	0.162	**0.490**

Table 5

Determining the distributions based on DIC ($T = 4$, $\boldsymbol{D} = diag(1, 1)$). The number in each cell represents the proportion of selecting the fitted model listed at the column heading for the type of generated data listed at the row heading.

N		Distributional Model			
		N-N model	T-N model	N-T model	T-T model
500	N-N data	**0.587**	0.000	0.413	0.000
	T-N data	0.000	**0.508**	0.000	0.492
	N-T data	0.000	0.000	**1.000**	0.000
	T-T data	0.000	0.000	0.000	**1.000**
300	N-N data	**0.670**	0.000	0.330	0.000
	T-N data	0.002	**0.474**	0.007	0.517
	N-T data	0.014	0.005	**0.981**	0.000
	T-T data	0.001	0.006	0.029	**0.964**
100	N-N data	**0.680**	0.004	0.316	0.000
	T-N data	0.083	**0.401**	0.100	0.416
	N-T data	0.132	0.003	**0.864**	0.001
	T-T data	0.027	0.087	0.240	**0.646**

Table 6

Simulation results of the post hoc degrees of freedom estimates checking method ($T = 4$, $\boldsymbol{D} = diag(1,1)$). The number in each cell represents, for the type of generated data listed at the row heading, the proportion of selecting the fitted model listed at the column heading.

N		df cutoff = 20				df cutoff = 25				df cutoff = 30				df cutoff = 45			
		N-N	T-N	N-T	T-T	N-N	T-N	N-T	T-T	N-N	T-N	N-T	T-T	N-N	T-N	N-T	T-T
500	N-N data	**1.000**	0.000	0.000	0.000	**0.998**	0.000	0.002	0.000	**0.994**	0.002	0.004	0.000	**0.939**	0.023	0.038	0.000
	T-N data	0.000	**1.000**	0.000	0.000	0.000	**0.999**	0.000	0.001	0.000	**0.997**	0.000	0.003	0.000	**0.972**	0.000	0.028
	N-T data	0.000	0.000	**0.999**	0.001	0.000	0.000	**0.998**	0.002	0.000	0.000	**0.994**	0.006	0.000	0.000	**0.974**	0.026
	T-T data	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	**1.000**
300	N-N data	**0.998**	0.001	0.001	0.000	**0.996**	0.002	0.002	0.000	**0.992**	0.004	0.004	0.000	**0.908**	0.042	0.047	0.003
	T-N data	0.001	**0.997**	0.000	0.002	0.001	**0.997**	0.000	0.002	0.001	**0.993**	0.000	0.006	0.000	**0.939**	0.000	0.061
	N-T data	0.014	0.000	**0.983**	0.003	0.007	0.000	**0.988**	0.005	0.006	0.000	**0.987**	0.007	0.000	0.000	**0.951**	0.049
	T-T data	0.000	0.017	0.003	**0.980**	0.000	0.011	0.001	**0.988**	0.000	0.006	0.001	**0.993**	0.000	0.001	0.000	**0.999**
100	N-N data	**0.997**	0.002	0.001	0.000	**0.995**	0.004	0.001	0.000	**0.991**	0.006	0.003	0.000	**0.877**	0.061	0.056	0.006
	T-N data	0.205	**0.795**	0.000	0.000	0.160	**0.839**	0.000	0.001	0.126	**0.870**	0.001	0.003	0.036	**0.905**	0.001	0.058
	N-T data	0.277	0.000	**0.723**	0.000	0.235	0.001	**0.764**	0.000	0.197	0.001	**0.795**	0.007	0.068	0.005	**0.856**	0.071
	T-T data	0.071	0.301	0.163	**0.465**	0.039	0.268	0.146	**0.547**	0.028	0.239	0.124	**0.609**	0.002	0.101	0.036	**0.861**

Table 7

(a) The Shapiro-Wilk test and Anscombe-Glynn test for residuals; (b) The Shapiro-Wilk test and Anscombe-Glynn test of kurtosis for the coefficients

(a)

	Shapiro-Wilk test		Anscombe-Glynn test		
Grade	Statistic	p-value	Kurtosis	Statistic	p-value
7	0.984	2.594E-05	4.757	4.764	1.898E-06
8	0.982	5.262E-06	4.536	4.408	1.045E-05
9	0.914	<2.2E-16	10.270	8.874	<2.2E-16
10	0.959	8.467E-11	6.349	6.603	4.034E-11

(b)

	Shapiro-Wilk test		Anscombe-Glynn test		
	Statistic	p-value	Kurtosis	Statistic	p-value
Intercept	0.994	0.053	2.982	0.075	0.941
Slope	0.996	0.210	3.085	0.546	0.586

Table 8

Estimates of the T-N and N-N distributional models in NLSY97 data analysis

	Parameters	Mean	s.d.	MC error	2.5%	97.5%	Geweke
T-N model	β_0	6.153	0.056	2.510E-04	6.043	6.263	0.017
	β_1	0.319	0.018	7.865E-05	0.284	0.353	-0.025
	D_{11}	1.197	0.102	4.550E-04	1.009	1.409	-0.017
	D_{22}	0.041	0.007	3.217E-05	0.029	0.057	0.023
	D_{12}	-0.048	0.023	1.046E-04	-0.095	-0.004	0.011
	σ^2	0.377	0.032	1.411E-04	0.316	0.440	-0.039
	df_e	4.155	0.532	0.002	3.239	5.342	-0.042
N-N model	β_0	6.157	0.057	5.737E-04	6.047	6.269	0.012
	β_1	0.312	0.019	2.944E-04	0.275	0.349	-0.021
	D_{11}	1.125	0.102	1.256E-03	0.937	1.337	-0.024
	D_{22}	0.035	0.006	1.229E-04	0.024	0.049	-0.097
	D_{12}	-0.034	0.023	4.109E-04	-0.081	0.009	0.053
	σ^2	0.748	0.029	2.209E-04	0.694	0.806	0.019

Note. s.d.: standard deviation; MC error: Monte Carlo error; Geweke: Geweke statistic.

Figure Captions

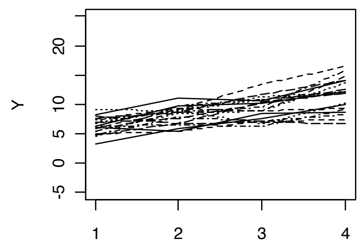
Figure 1. Trajectory plots of data generated by the 4 different types of growth curve models. Data on 20 subjects are generated for 4 measurement occasions. Figure 1(a), (b), (c), and (d) display the trajectories of the data generated by the N-N, T-N, N-T, and T-T distributional model, respectively.

Figure 2. The web interface for distribution diagnostics for the four types of distributional growth curve models

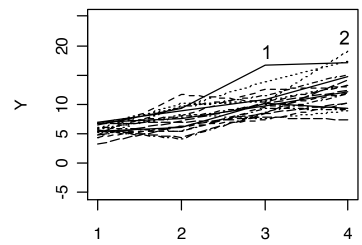
Figure 3. A collection of individual trajectories (a spaghetti plot) for the NLSY97 data. 512 school children are measured at 4 occasions.

Figure 4. Histograms and Q-Q plots of estimated residuals at the 4 measurement occasions

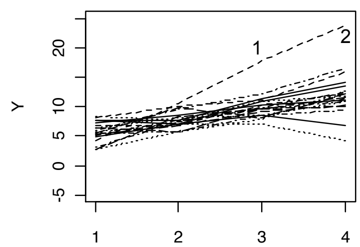
Figure 5. Histogram and Q-Q plot of regression coefficients



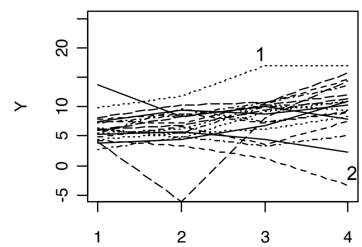
(a) N-N



(b) T-N



(c) N-T



(d) T-T

Distribution diagnostics for robust growth curve models

Please select the data file you want to use

Upload data Choose File No file chosen

(less than 1Mb & only .txt and .dat allowed.):

Select the method to use

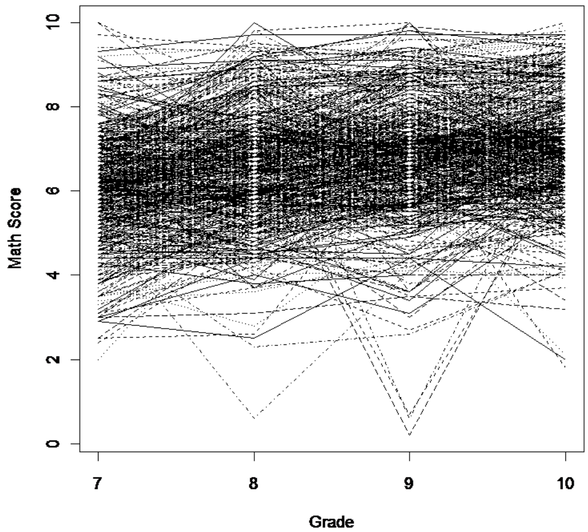
- (•) Distribution checking based on individual growth curve analysis
- ( ) Distribution comparison based on Deviance Information Criterion (DIC)
  - Fix degrees of freedom at: 0
- ( ) Post hoc checking of degrees of freedom estimates

Control MCMC

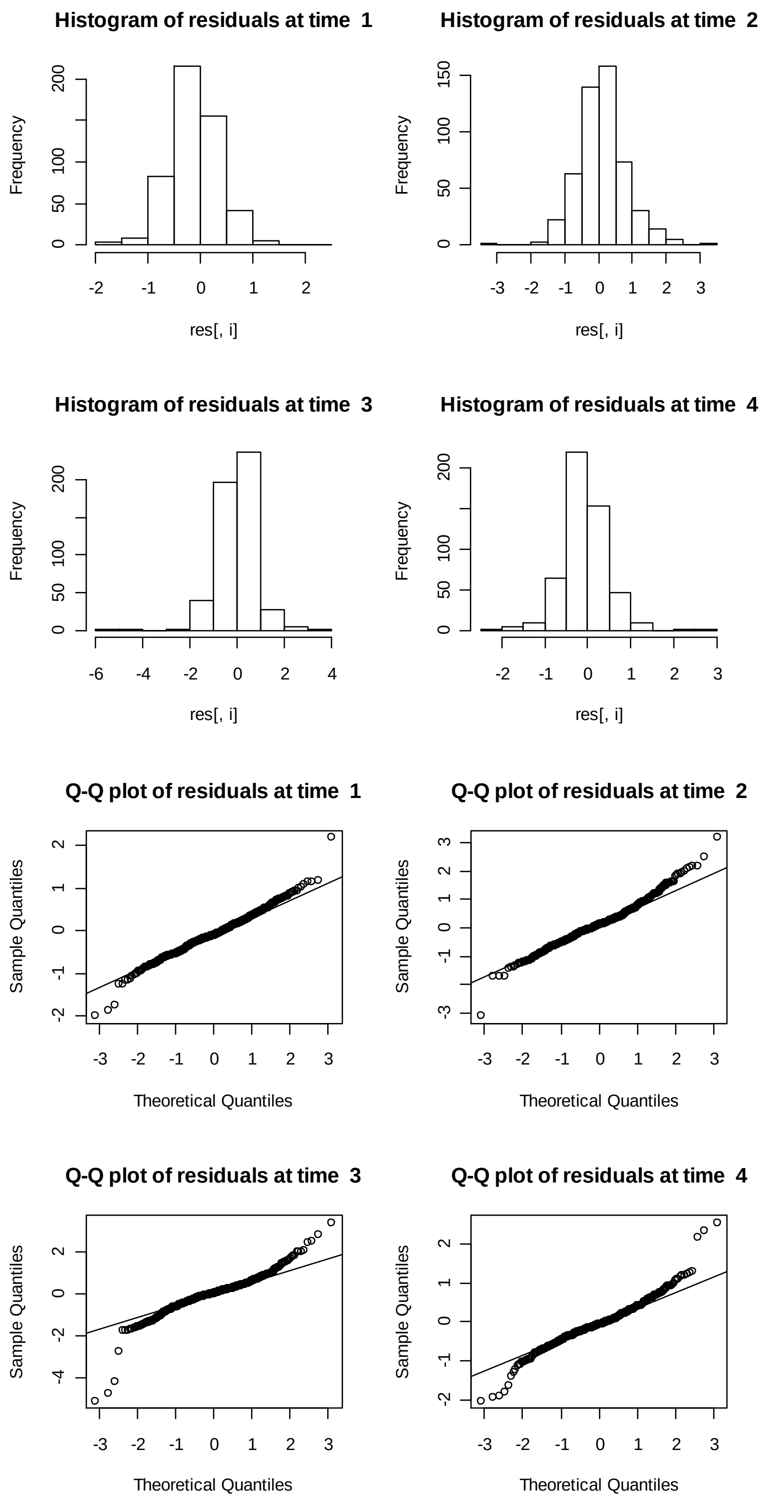
Burn-in period: 1000

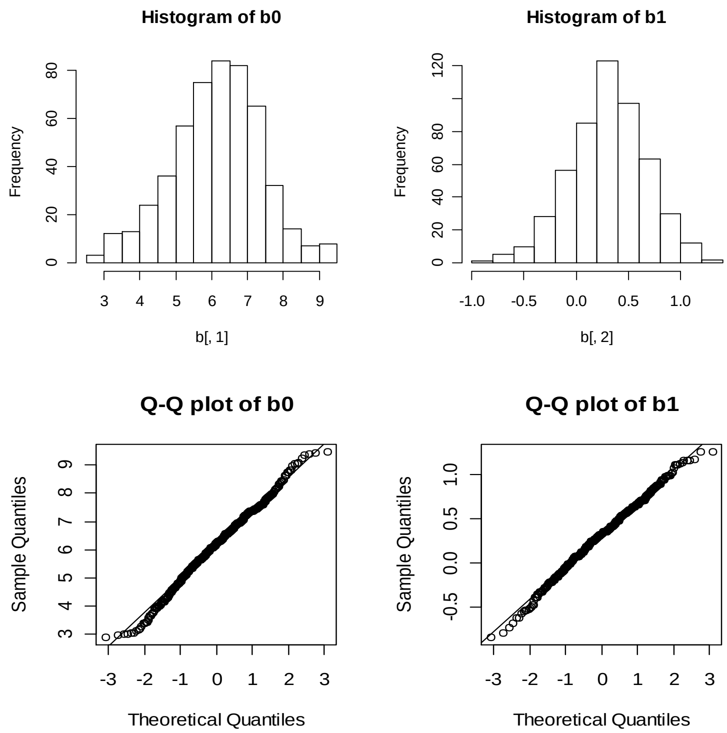
MCMC length: 10000

Submit



Math Score
0
2
4
6
8
10
Grade
7
8
9
10


Histogram of residuals at time 1
Frequency
res[, i]
Histogram of residuals at time 2
Frequency
res[, i]
Histogram of residuals at time 3
Frequency
res[, i]
Histogram of residuals at time 4
Frequency
res[, i]
Q-Q plot of residuals at time 1
Sample Quantiles
Theoretical Quantiles
Q-Q plot of residuals at time 2
Sample Quantiles
Theoretical Quantiles
Q-Q plot of residuals at time 3
Sample Quantiles
Theoretical Quantiles
Q-Q plot of residuals at time 4
Sample Quantiles
Theoretical Quantiles


Histogram of b0
Frequency
b[, 1]
Histogram of b1
Frequency
b[, 2]
Q-Q plot of b0
Sample Quantiles
Theoretical Quantiles
Q-Q plot of b1
Sample Quantiles
Theoretical Quantiles