

1 Running head: DIAGNOSTICS OF ROBUST GROWTH CURVE MODELING

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

3 Xin Tong and Zhiyong Zhang

4 University of Notre Dame

5

Abstract

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

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

19
20
21 Growth curve modeling is one of the most frequently used analytic techniques for
22 longitudinal data analysis with repeated measures (e.g., McArdle, 1988; Meredith & Tisak, 1990).
23 A growth curve model generally consists of fixed effects and random effects where the random
24 effects account for the interindividual variation. In traditional growth curve analysis, it is typically
25 assumed that the random effects and intraindividual measurement errors are normally distributed.
26 Although the normality assumption makes growth curve models easy to estimate and apply, data in
27 social and behavioral science usually violate such an assumption (Micceri, 1989). Practically, data
28 often have longer-than-normal tails and/or outliers. Ignoring the nonnormality of data may result in
29 unreliable parameter estimates, their associated standard errors estimates, and thus, misleading
30 tests and inference (Maronna, Martin, & Yohai, 2006; Yuan, Bentler, & Chan, 2004; Zu & Yuan,
31 2010). Some routine methods, such as deleting the outliers, may lead to serious problems (e.g.,
32 Lange, Little, & Taylor, 1989; Yuan & Bentler, 2002). For example, the resulting inferences can
33 fail to reflect uncertainty in the exclusion process and the efficiency can be reduced.

34 In the past half century, researchers have become more keenly aware of the rather large
35 influence that nonnormality has upon model estimations (Hampel, Ronchetti, Rousseeuw, &
36 Stahel, 1986; Huber, 1981) and have developed what are called robust methods aiming to provide
37 reliable parameter estimates and inference when the normal distribution assumption is violated.
38 The basic idea of the robust method is to assign a weight to each case according to its distance
39 from the center of the majority of the data, so that the extreme cases can be downweighted (e.g.,
40 Huber, 1981; Zhong & Yuan, 2010; Yuan, Bentler, & Chan, 2004). A few studies have directly
41 discussed the robust methods in growth curve analysis. For example, Pendergast & Broffitt (1985)
42 and Singer & Sen (1986) proposed robust estimators based on M-methods for growth curve models
43 with elliptically symmetric errors, and Silvapulle (1992) further extended the M-method to allow
44 asymmetric errors for growth curve analysis. Recently, Zhang, Lai, Lu, & Tong (2012) suggested

45 modeling heavy-tailed data and outliers using Student's t distributions and provided online
46 software to conduct the robust analysis.

47 The use of the t distribution for describing a single sample dates back at least to Jeffreys
48 (1939). A lot of work has been done to boost its application (Hampel et al., 1986; Little, 1988;
49 Sutradhar & Ali, 1986). The wide adoption of the t distribution in robust statistical modeling was
50 advanced by Lange, Little, & Taylor (1989). Lately, the t distribution has been applied in more
51 complex data analysis, such as robust structural equation models (Lee & Xia, 2006, 2008), linear
52 mixed-effects models (Pineiro, Liu, & Wu, 2001; Song, Zhang, & Qu, 2007), and robust mixture
53 modeling (Wang, Zhang, Luo, & Wei, 2004; Shoham, 2002). The advantages in using t
54 distribution for robust data analysis are obvious: a) it has a parametric form and inference based on
55 it can be carried out through maximum likelihood methods and Bayesian methods relatively easily;
56 b) the degrees of freedom of t distributions control the weight of outliers and can be either set *a*
57 *priori* or estimated; c) robust methods based on t distributions are easy to understand and can be
58 considered as a direct extension of corresponding normal distribution methods for heavy-tailed
59 data. Zhang et al. (2012) demonstrated that the robust growth curve modeling based on t
60 distributions could be easy to understand and implement, and thus potentially would greatly
61 promote the adoption of robust growth curve analysis.

62 In growth curve analysis, both the random effects and intraindividual measurement errors
63 are equally likely to follow t distributions instead of normal distributions. However, Zhang et al.
64 (2012) only focused on the case where the measurement errors were modeled by t distributions,
65 although their software allows for simultaneously modeling both the random effects and the
66 measurement errors by t distributions. Pineiro et al. (2001) proposed a robust version of the linear
67 mixed-effect model, in which the normal distributions for the random effects and the
68 intraindividual errors were both replaced by multivariate t distributions. In fact, given a set of
69 longitudinal data, one could fit one of the four possible types of distributional growth curve models
70 to the data with (1) normal measurement errors and normal random effects, (2) t measurement

71 errors and normal random effects, (3) normal measurement errors and t random effects, and (4) t
 72 measurement errors and t random effects. However, no study has been conducted to evaluate and
 73 compare the intrinsic characteristics of the four distributional models and no strategy has been
 74 developed to assist in choosing the appropriate model for a given set of data.

75 The purpose of this study is to evaluate and compare the four types of distributional growth
 76 curve models and develop strategies to select the appropriate model for growth curve analysis. In
 77 the following, we first illustrate the features of the four different types of distributional growth
 78 curve models. Then, we compare the four models in modeling different types of data through a
 79 simulation study. After the comparison, we develop three methods for selecting appropriate
 80 distributions for the measurement errors and random effects. Simulation studies are also carried
 81 out to demonstrate the effectiveness of those methods. Finally, we illustrate the use of the methods
 82 through an example with the real data from the National Longitudinal Survey of Youth (NLSY)
 83 1997 Cohort. We end the paper with the discussion of the differences among the three methods.

84 Robust Growth Curve Models

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

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

88 where $\mathbf{\Lambda}$ is a $T \times q$ factor loading matrix determining the growth trajectories, \mathbf{b}_i is a $q \times 1$ random
 89 vector of random effects, and \mathbf{e}_i is a vector of intraindividual measurement errors. \mathbf{b}_i varies for
 90 each individual and its mean, $\boldsymbol{\beta}$, represents the fixed effects. \mathbf{u}_i is a residual vector that represents
 91 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
 92 type of distribution with different means.

93 Traditional growth curve models typically assume that both \mathbf{e}_i and \mathbf{u}_i follow multivariate
 94 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
 95 denotes a multivariate normal distribution and the subscription denotes its dimension. Note that it
 96 is assumed that the errors are identically and independently distributed over time with mean 0 and
 97 variance σ^2 but this assumption can be relaxed. \mathbf{D} is a $q \times q$ square matrix which represents the
 98 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
 99 convenience, we denote this model with normal errors and normal random effects as the
 100 Normal-Normal (N-N) distributional model.

101 Although the N-N distributional growth curve models are widely used, practical data often
 102 violate these normal assumptions. Zhang et al. (2012) proposed to model the measurement errors
 103 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
 104 MT_T denote a T -dimension multivariate t distribution. The resulting robust growth curve models
 105 have been shown to perform better in analyzing data with long-tailness and outliers than the N-N
 106 distributional model. Another type of robust growth curve models can be derived by letting
 107 $\mathbf{e}_i \sim MN_T(0, \sigma^2 \mathbf{I}_{T \times T})$ and $\mathbf{u}_i \sim MT_q(0, \mathbf{D}_{q \times q}, df_u)$. The third type of robust growth curve
 108 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
 109 denote the three types of robust growth curve models as the T-N distributional model, the N-T
 110 distributional model, and the T-T distributional model, respectively.

111 To illustrate the differences among the N-N, T-N, N-T, and T-T distributional models, we
 112 generate and plot data from the four types of models (Figure 1). For each type of model, data on 20
 113 subjects are generated for 4 occasions assuming a linear growth trend. Figure 1(a) displays the
 114 trajectories of the data generated by the N-N distributional model. No outlier can be observed. The
 115 overall trajectory looks clean and smooth. Figure 1(b) plots the data generated by the T-N
 116 distributional model. Noticeably, some observations stand out of the overall trajectory such as
 117 those labeled by 1 and 2. A close look at the two observations reveals that the reason why they
 118 deviate from the overall trajectory is that they are off their own growth trajectories. For example, at

119 a given time, an individual performed much better than expected on a test. Figure 1(c) portrays
120 data from the N-T distributional model. Some observations also deviate from the overall growth
121 trajectory. However, it seems that those observations are still on their own growth trajectories. The
122 reason why they stand out is that the rate of growth for the specific case is very different from the
123 majority of cases. For example, for a participant, he or she could display an unusual rate of growth
124 during the period of data collection. Figure 1(d) draws the trajectories for the data from the T-T
125 distributional model. Clearly, the outlying observations are due to two sources - the trajectory of a
126 case deviates from the overall trajectory and the observation for this specific case is off its own
127 trajectory. For example, the observation 1 stands out because it is off the trajectory of the case and
128 the case itself has a higher initial level.

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

138 **Influences of distribution misspecification**

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

144 To estimate these models, we use a Bayesian method, the use of which for complex data
 145 analysis has become popular in the past two decades (Lee & Shi, 2000; Lee, 2007; Lee & Song,
 146 2008; Lee & Xia, 2008). Zhang et al. (2012) applied the Bayesian method to the robust growth
 147 curve models with Student's t distributed measurement errors and normally distributed random
 148 coefficients. We introduce this method into our model estimation for all the four types of
 149 distributional models. The basic idea of the Bayesian method is to obtain the posterior distributions
 150 of model parameters based on the likelihood function and the prior distributions. Details are given
 151 in appendix A of Zhang et al. (2012).

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

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix},$$

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

158 For each simulation study, a total of 1000 data sets are generated and analyzed. Two
 159 commonly used statistics, which examine more than one performance criterion (Collins, Schafer,
 160 & Kam, 2001), are calculated for each model parameter to compare the four types of growth curve
 161 models. The first statistic is the mean square error (MSE) based on 1000 sets of parameter
 162 estimates and standard errors, and the second one is the coverage probability (CP) of the 95%
 163 credible intervals². The MSE and CP are then averaged over all model parameters (except the
 164 degrees of freedom) for each simulation condition. Table 1 summarizes the results for the analysis
 165 of each type of data by different types of distributional models with different sample sizes. In the
 166 table, on the rows are the different types of generated data and on the columns are the four types of

167 distributional models used to analyze the generated data.

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

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

183 Second, numerically through our simulation studies, the T-T distributional growth curve
184 model is not optimal all the time. From Table 1, as the sample size decreases, the MSE becomes
185 larger and the coverage probability becomes further away from its nominal level 0.95, meaning
186 that the accuracy of the T-T distributional model is getting worse. To better illustrate the influence
187 of sample size, Table 2 provides the parameter estimates for the T-T distributional model when the
188 sample size is 50. Although the fixed effects parameter estimates for the initial level and rate of
189 growth are very close to their true values, other parameters such as the estimate for the variability
190 of slope are largely biased. This is due to the difficulty of getting accurate estimates for the degrees
191 of freedom of the t distributions. When the sample size is small, it is more likely to identify a t
192 distribution as a normal distribution. Thus, the degrees of freedom estimates are positively biased,

193 and accordingly, the other parameter estimates may be inaccurate. To solve this problem, it is
194 suggested to supply a fixed value as the degree of freedom of the t distribution (Zhang et al., 2012;
195 Lange et al., 1989). Namely, it is necessary to select the right model type at the first step. Then, we
196 can fix the degrees of freedom based on that model type and estimate the other parameters in order
197 to obtain more accurate parameter estimates and credible intervals.

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

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

206 **Methods for distribution diagnostics**

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

218 identify the correct models and distributions.

219 *Distribution checking based on individual growth curve analysis*

220 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
 221 individual through the least squares method. The individual coefficients (random effects)
 222 $\mathbf{b}_i = (b_{i0}, \dots, b_{iq})^T$ and the measurement errors $\mathbf{e}_i = (e_{i1}, \dots, e_{iT})^T$ are estimated and retained.
 223 Let $\mathbf{b} = (\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_N)^T$ and $\mathbf{e} = (\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N)^T$ where \mathbf{b} is a $N \times q$ matrix of individual
 224 coefficients estimates and \mathbf{e} is a $N \times T$ matrix of estimated errors. Then, we test whether \mathbf{e} and \mathbf{b}
 225 follow normal distributions. If all q columns of \mathbf{b} follow normal distributions, we consider the
 226 individual coefficients to be normally distributed. Otherwise, we consider them nonnormally
 227 distributed. Similarly, if all T columns of \mathbf{e} are normally distributed, the errors are viewed as from
 228 normal distributions. If \mathbf{e} and \mathbf{b} are not normally distributed, t distributions are recommended.
 229 Based on the combination of the distributions for \mathbf{e} and \mathbf{b} , the decision can be made according to
 230 Table 3.

231 Note that four null hypotheses are tested in order to select a model. They are: (1) both
 232 measurement errors \mathbf{e} and individual coefficients \mathbf{b} are normally distributed, (2) \mathbf{e} is normal while
 233 \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
 234 these four tests, only three are independent because if it is certain that one data set is not generated
 235 from three of the models, it has to be generated from the remaining model. Thus, a Bonferroni
 236 correction should be used for multiple testing. The new alpha level should be
 237 $\alpha/3 = 0.05/3 \doteq 0.017$.

238 Two tests are employed in this study to check the normality of random coefficients and
 239 errors – the Shapiro-Wilk normality test (Shapiro & Wilk, 1965) and the Anscombe-Glynn test of
 240 kurtosis (Anscombe & Glynn, 1983). The Shapiro-Wilk normality test checks the normality based
 241 on both skewness and kurtosis. If data exhibit either skewness or kurtosis, the normality is
 242 rejected. It is applied here because its test statistic exhibits sensitivity to nonnormality over a wide

243 range of alternative distributions and in most cases it has power as good as or better than many
244 other normality test statistics (Shapiro et al., 1968). The Anscombe-Glynn test checks whether the
245 kurtosis for the current data is larger than the kurtosis of normal data. It is good at detecting
246 nonnormality caused by nonnormal tail heaviness. If a distribution is symmetric but heavy-tailed,
247 the test for kurtosis may be more powerful than the Shapiro-Wilk test. The Anscombe-Glynn test
248 is used in the simulation study because of the heavy-tailed property of our generated data.

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

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

264 A closer look at the results also reveals some useful details. For the N-N data, it is most
265 likely to be mis-classified as being generated from the T-N distributional model and least likely to
266 be mis-classified as being generated from the T-T distributional model. For the T-N data, it is most
267 likely to be mis-classified as the data generated from the T-T distributional model and least likely
268 to be mis-classified as being generated from the N-T distributional model. When the sample size is

269 small, for the N-T data, it is most and least likely to be mis-classified as being generated from the
 270 N-N distributional model and the T-N distributional model, respectively. Finally, for the T-T data,
 271 it is more likely to be mis-classified as the data generated from the T-N distributional model and
 272 the N-T distributional model than the N-N distributional model. The results are consistent for both
 273 the Shapiro-Wilk test and the Anscombe-Glynn test.

274 *Distribution comparison based on Deviance Information Criterion (DIC)*

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

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

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

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

292 The results in Table 5 clearly suggest that the performance of DIC in determining the

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

300 *Post hoc checking of degrees of freedom estimates*

301 As the degrees of freedom increase, a t distribution approaches a normal distribution. Thus,
 302 if a t distribution is fitted to a set of data and the estimated degree of freedom is large, it indicates
 303 that a normal distribution may be good enough to model the data. For the robust growth curve
 304 modeling, the same strategy can be applied. Specifically, the T-T distributional model can be fitted
 305 to the data, and the degrees of freedom for random effects and errors can be estimated. If both df_u
 306 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
 307 distributions. Thus, the N-N distributional model can be used. If df_u is larger than the cutoff while
 308 df_e is not, it implies that \mathbf{u} follows a normal distribution and \mathbf{e} follows a t distribution.

309 Accordingly, the T-N distributional model is the appropriate one to select. In the opposite, if df_e is
 310 larger than the cutoff while df_u is not, the N-T distributional model should be selected. Finally, if
 311 both df_u and df_e are smaller than the cutoff, the T-T distributional model should be applied.

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

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

328 *Implementation*

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

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

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

343

An Example

344 To demonstrate the application of the three methods for diagnosing distributions for random
 345 effects and measurement errors, we investigate a subset of data from the National Longitudinal
 346 Survey of Youth (NLSY) 1997 Cohort (Bureau of Labor Statistics, U.S. Department of Labor,
 347 2005). A sample of $N = 512$ school children was administered the Peabody Individual
 348 Achievement Test (PIAT) Mathematics subset yearly from grade 7 to grade 10. The individuals'
 349 trajectory plot (Figure 3) suggests a linear growth pattern and a T-N distributional model because
 350 the pattern of the figure is similar to the pattern in Figure 1(b). The following illustration is to
 351 determine the best distributions for the random coefficients and measurement errors for the linear
 352 growth curve model.

353 First, a simple linear regression, $y_{it} = b_{i0} + b_{i1}t + e_{it}$, is fitted to data of each individual,
 354 and the individual regression coefficients \hat{b}_{i0} and \hat{b}_{i1} and measurement (residual) errors \hat{e}_{it} are
 355 estimated for $t = 1, \dots, 4$ and $i = 1, \dots, 512$. The histograms and Q-Q plots for the errors are
 356 portrayed in Figure 4. Both histograms and Q-Q plots indicate that the errors have
 357 longer-than-normal tails. The results from the Shapiro-Wilk normality test and the
 358 Anscombe-Glynn test of kurtoses are given in Table 7(a). Based on the Shapiro-Wilk normality
 359 test, the normality of the errors at all grades are rejected. Furthermore, the kurtoses from all grades
 360 are significantly larger than the normal kurtosis, ranging from 4.536 to 10.270 from grade 7 to
 361 grade 10. Thus, it seems more reasonable to specify t distributions for the errors than normal
 362 distributions. Figure 5 presents the histograms and Q-Q plots for the random coefficients and Table
 363 7(b) summarizes the results of the Shapiro-Wilk test and the Anscombe-Glynn test. From either
 364 the plots or the tests, both random coefficients appear to be normally distributed. Overall, it seems
 365 that the T-N distributional model is a good choice for the current sets of data based on the checking
 366 of the distributions of the errors and random coefficients.

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

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

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

382 To summarize, all three methods lead to the same conclusion that the T-N distributional
383 model with normally distributed random coefficients and t distributed errors fits the data best in
384 this example. This is consistent with the pattern we observe in Figure 3. Individuals' growth
385 trajectoryies are within the normal range of the overall trajectory, while some observations at
386 certain times stand out of their own developmental pathways. The parameter estimates for the T-N
387 distributional model are given in Table 8. To ensure convergence, we have calculated the Geweke
388 statistics (Geweke, 1992) for all parameters. All Geweke statistics are smaller than 2, suggesting
389 the convergence of the Markov chains. Furthermore, for all parameters, the Monte Carlo errors are
390 smaller than 5% of their corresponding standard deviations, indicating accurate parameter
391 estimates. Based on the results, the average initial mathematical ability at grade 7 is about 6.153
392 with an average total change of 0.319 from grade 7 to grade 10. By conducting z -tests on D_{11} and
393 D_{22} , it is found that there are also significant individual differences in both initial ability and
394 change. Compared to the parameter estimates using the N-N distributional model, it is found that

395 the estimate for D_{12} , the covariance between the intercept and the slope, from the robust T-N
396 distributional model is substantially different from that of the N-N distributional model, as it is
397 significant for the T-N distributional model and insignificant for the N-N distributional model.
398 Because the correlation between the intercept and slope parameters across sessions are interested
399 in many studies (e.g. Zhang, Davis, Salhouse, & Tucker-Drob, 2007), the N-N distributional
400 model should not be used as a substitute.

401 Discussion

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

410 The method of distribution checking based on individual growth curve analysis has many
411 advantages. First, it is easy to implement. The random effects and errors can be obtained through
412 the OLS method. Second, the distributions of random effects and errors can be displayed visually
413 through histograms and Q-Q plots. The normality of random effects and errors can be further
414 tested through the available methods such as the Shapiro-Wilk test and the Anscombe-Glynn test.
415 Third, the diagnostics of the distributions are not confined to the t distributions and can be
416 extended to any other distribution. Fourth, the use of this method usually does not require a large
417 sample size. The main disadvantage of this method is that the distribution diagnostics based on the
418 Shapiro-Wilk test and the Anscombe-Glynn test do not perform very well for some types of data
419 under certain conditions. For example, the T-T data can be mistakenly classified into other three

420 types of data easily when the sample size is small (e.g. $N \leq 100$). Not correcting the type I error
421 for the multiple testing problems improves the results for the data with a small sample size,
422 because correction makes it more difficult to reject the null hypothesis and more likely to conclude
423 that distributions are normal.

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

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

442 To conclude, our aim in this article has been to compare the four types of distributional
443 growth curve models and evaluate the three proposed model selection methods. When the sample
444 size is large, the post hoc checking of degrees of freedom estimates method can be directly used as
445 the best model selection method. When the sample size is relatively small, this method is not

446 suggested anymore. In contrast, the method of distribution checking based on individual growth
447 curve analysis, combined with the distribution comparison method based on DIC, should be used
448 together to make a decision. Note that, to achieve a higher correct selection rate for data with a
449 small sample size, one may consider adopting a less strict alpha level in the first method and using
450 the second method with fixed degrees of freedom.

451 *Limitations*

452 The study was conducted for data with longer-than-normal tailed distributions but not with
453 skewed distributions. Like the normal distribution, the t distribution is symmetric and can be
454 sensitive to skewed data. If the skewness in the data is caused by outliers, the robust growth curve
455 models can still perform very well. However, if the skewness is because of skewed distributions
456 such as the lognormal distribution, robust growth curve models can break down, although they still
457 work relatively better than the normal growth curve models (Zhang et al., 2012). In this case,
458 skew- t distributions may be applied to accommodate the skewness in the data (Azzalini & Genton,
459 2008; Lachos, Dey, & Cancho, 2009). But the performance of the skew- t distribution in robust
460 growth curve modeling needs further evaluation.

461 For robust growth curve model selection, we have used a distribution comparison method
462 based on DIC. Although DIC is very widely used, it has received much criticism since it was
463 proposed (Spiegelhalter, Best, Carlin, & Linde, 2002b). Celeux, Forbers, Robert, & Titterington
464 (2006) argued that the DIC introduced by Spiegelhalter et al. (2002b) for model assessment and
465 model comparison was directly inspired by linear and generalized linear models, but it was open to
466 different possible variations in the setting of models involving random effects, as in our robust
467 growth curve models. A number of ways of computing DICs are proposed in Celeux et al. (2006),
468 and their advantages and disadvantages are discussed. Thus, more sophisticated fit indices should
469 be carried out for comparison in the future, since a better computation of DIC may lead to better
470 results of the model selection method.

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

475 **References**

- 476 Anscombe, F., & Glynn, W. (1983). Distribution of kurtosis statistic for normal statistics.
 477 *Biometrika*, 70(1), 227–234.
- 478 Azzalini, A., & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related
 479 distributions. *International Statistical Review*, 76(1), 106–129.
- 480 Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Wiley.
- 481 Bureau of Labor Statistics, U.S. Department of Labor. (2005). National longitudinal survey of
 482 youth 1997 cohort, 1997-2003 (rounds 1-7) [computer file] [Computer software manual].
 483 Produced by the National Opinion Research Center, the University of Chicago and distributed
 484 by the Center for Human Resource Research, The Ohio State University. Columbus, Ohio.
- 485 Celeux, G., Forbers, F., Robert, C., & Titterington, D. (2006). Deviance information criteria for
 486 missing data models. *Bayesian Analysis*, 1, 651–674.
- 487 Collins, L., Schafer, J., & Kam, C. (2001). A comparison of inclusive and restrictive missing-data
 488 strategies in modern missing-data procedures. *Psychological Methods*, 6, 330–351.
- 489 Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior
 490 moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian*
 491 *statistics 4* (pp. 169–193). Oxford, UK: Clarendon Press.
- 492 Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). *Robust statistics: The approach*
 493 *based on influence functions*. New York: John Wiley & Sons, Inc.

- 494 Huber, P. J. (1981). *Robust statistics*. New York: John Wiley & Sons, Inc.
- 495 Jeffreys, H. (1939). *Theory of probability*. Oxford, U.K.: Clarendon Press.
- 496 Lachos, V. H., Dey, D. K., & Cancho, V. G. (2009). Robust linear mixed models with
497 skew-normal independent distributions from a bayesian perspective. *Journal of Statistical*
498 *Planning and Inference*, 139, 4098–4110.
- 499 Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t
500 distribution. *Journal of the American Statistical Association*, 84(408), 881–896.
- 501 Lee, S. Y. (2007). *Structural equation modeling: A bayesian approach*. New York: John Wiley &
502 Sons, Inc.
- 503 Lee, S. Y., & Shi, J. Q. (2000). Joint bayesian analysis of factor scores and structural parameters in
504 the factor analysis model. , vol. 52, 2000, 722-736. *Annals of the Institute of Statistical*
505 *Mathematics*, 52, 722–736.
- 506 Lee, S. Y., & Song, X. Y. (2008). On bayesian estimation and model comparison of an integrated
507 structural equation model. *Computational Statistics and Data Analysis*, 52, 4814–4827.
- 508 Lee, S. Y., & Xia, Y. M. (2006). Maximum likelihood methods in treating outliers and
509 symmetrically heavy-tailed distributions for nonlinear structural equation models with missing
510 data. *Psychometrika*, 71, 541–565.
- 511 Lee, S. Y., & Xia, Y. M. (2008). A robust bayesian approach for structural equation models with
512 missing data. *Psychometrika*, 73, 343–364.
- 513 Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with
514 missing values. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37(1),
515 23–38.

- 516 Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The bugs project: Evolution, critique
517 and future directions (with discussion). *Statistics in Medicine*, 28, 3049–3082.
- 518 Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. New
519 York: John Wiley & Sons, Inc.
- 520 McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In
521 J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology*
522 (2nd ed., pp. 561–614). New York: Plenum Press.
- 523 Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122.
- 524 Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological*
525 *Bulletin*, 105(1), 156–166.
- 526 Pendergast, J. F., & Broffitt, J. D. (1985). Robust estimation in growth curve models.
527 *Communications in Statistics: Theory and Methods*, 14, 1919–1939.
- 528 Pinheiro, J. C., Liu, C., & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear
529 mixed-effects models using the multivariate t distribution. *Journal of Computational and*
530 *Graphical Statistics*, 10(2), 249–276.
- 531 R Development Core Team. (2011). R: A language and environment for statistical computing
532 [Computer software manual]. Vienna, Austria. Available from
533 <http://www.R-project.org> (ISBN 3-900051-07-0)
- 534 Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete
535 samples). *Biometrika*, 52, 591–611.
- 536 Shapiro, S. S., Wilk, M. B., & Chen, M. H. (1968). A comparative study of various tests for
537 normality. *American Statistical Association Journal*, 1343–1372.

- 538 Shoham, S. (2002). Robust clustering by deterministic agglomeration em of mixtures of
539 multivariate t-distributions. *Pattern Recognition*, 35, 1127–1142.
- 540 Silvapulle, M. J. (1992). On m-methods in growth curve analysis with asymmetric errors. *Journal*
541 *of Statistical Planning and Inference*, 32(3), 303–309.
- 542 Singer, J. M., & Sen, P. K. (1986). M-methods in growth curve analysis. *Journal of Statistical*
543 *Planning and Inference*, 13, 251–261.
- 544 Song, P., Zhang, P., & Qu, A. (2007). Maximum likelihood inference in robust linear
545 mixed-effects models using multivariate t distribution. *Statistica Sinica*, 17, 929–943.
- 546 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. v. d. (2002a). Bayesian measures of
547 model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical*
548 *Methodology)*, 64(4), 583-639.
- 549 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. v. d. (2002b). Bayesian measures of
550 model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical*
551 *Methodology)*, 64(4), 583–639.
- 552 Sutradhar, B. C., & Ali, M. M. (1986). Estimation of the parameters of a regression model with a
553 multivariate t error variable. *Communications in Statistics - Theory and Methods*, 15(2),
554 429–450.
- 555 Venables, W. N., & Ripley, B. D. (1999). *Modern applied statistics with s-plus*. Springer.
- 556 Wang, H.-X., Zhang, Q.-B., Luo, B., & Wei, S. (2004). Robust mixture modelling using
557 multivariate t-distribution with missing information. *Pattern Recognition Letters*, 25, 701–710.
- 558 Yuan, K.-H., & Bentler, P. M. (2002). On normal theory based inference for multilevel models
559 with distributional violations. *Pschometrika*, 67, 539–561.

- 560 Yuan, K.-H., Bentler, P. M., & Chan, W. (2004). Structural equation modeling with heavy tailed
561 distributions. *Psychometrika*, *69*, 421—436.
- 562 Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean
563 and covariance structure analysis: The univariate case and its multivariate implication.
564 *Sociological Methods and Research*, *34*.
- 565 Zhang, Z., Davis, H. P., Salthouse, T. A., & Tucker-Drob, E. A. (2007). Correlates of individual,
566 and age-related, differences in short-term learning. *Learning and Individual Differences*, *17*(3),
567 231–240.
- 568 Zhang, Z., Lai, K., Lu, Z., & Tong, X. (2012). Bayesian inference and application of robust growth
569 curve models using student's t distribution. *Structural Equation Modeling*, (*in press*).
- 570 Zhong, X., & Yuan, K.-H. (2010). Weights. In N. J. Salkind (Ed.), *Encyclopedia of research*
571 *design* (pp. 1617–1620). Thousand Oaks, CA: Sage.
- 572 Zu, J., & Yuan, K.-H. (2010). Local influence and robust procedures for mediation analysis.
573 *Multivariate Behavioral Research*, *45*, 1–44.

574

Footnotes

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

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

Table 1

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

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 $\mathbf{D} = \text{diag}(4, 2)$

	True	Est.	Absolute bias	Relative bias(%)	SE	MSE	CP
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 ($T = 4$, $D = \text{diag}(1, 1)$). 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 ($T = 4$, $\mathbf{D} = \text{diag}(1, 1)$). 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 ($T = 4$, $D = \text{diag}(1, 1)$). 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)					
Grade	Shapiro-Wilk test		Anscombe-Glynn test		
	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.

582

Figure Captions

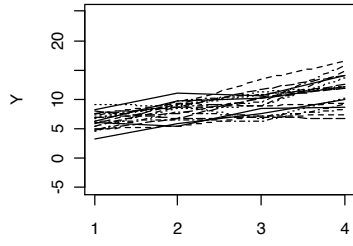
583 *Figure 1.* Trajectory plots of data generated by the 4 different types of growth curve models. Data
584 on 20 subjects are generated for 4 measurement occasions. Figure 1(a), (b), (c), and (d) display the
585 trajectories of the data generated by the N-N, T-N, N-T, and T-T distributional model,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

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

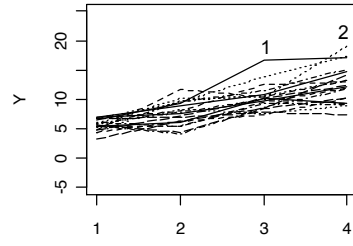
588 *Figure 3.* A collection of individual trajectories (a spaghetti plot) for the NLSY97 data. 512 school
589 children are measured at 4 occasions.

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

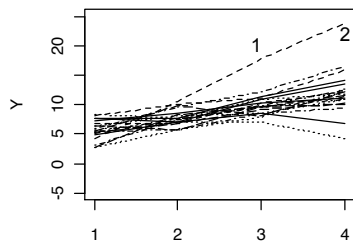
591 *Figure 5.* Histogram and Q-Q plot of regression coefficients



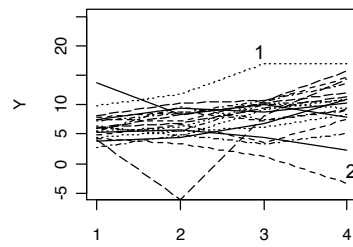
(a) N-N



(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 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)

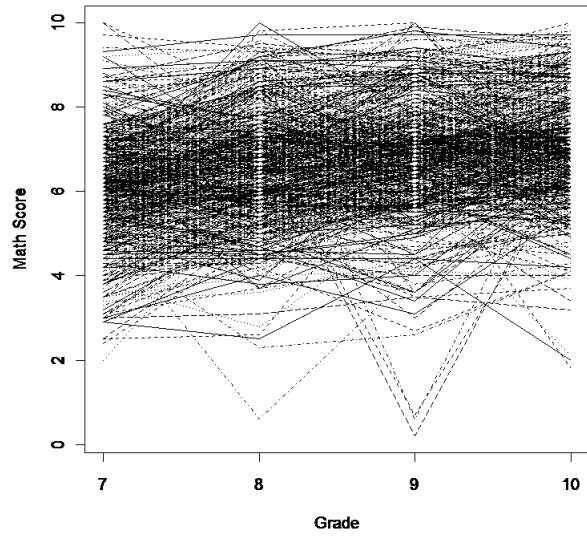
Fix degrees of freedom at:

- Post hoc checking of degrees of freedom estimates

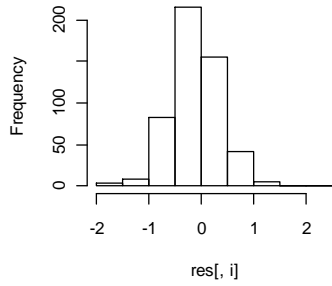
Control MCMC

Burn-in period:

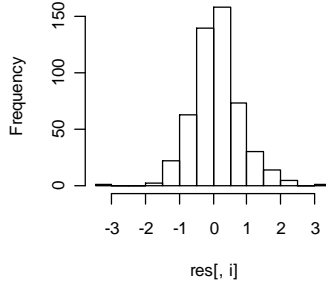
MCMC length:



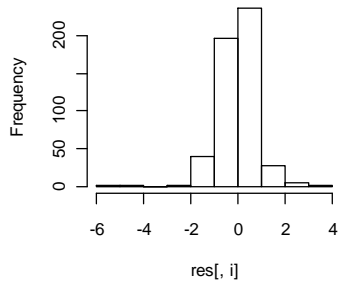
Histogram of residuals at time 1



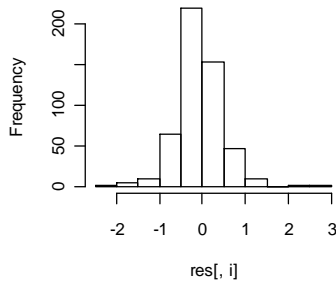
Histogram of residuals at time 2



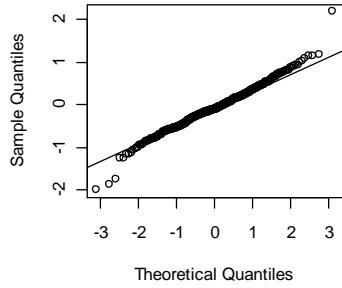
Histogram of residuals at time 3



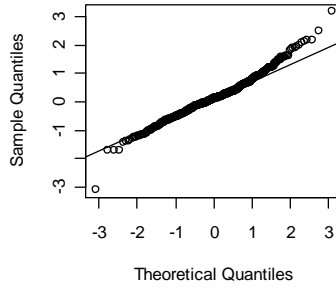
Histogram of residuals at time 4



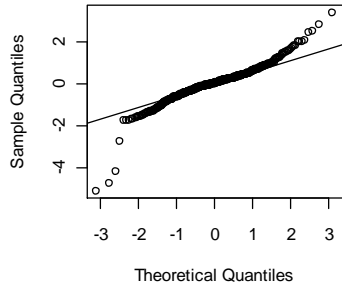
Q-Q plot of residuals at time 1



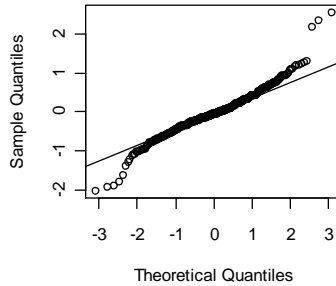
Q-Q plot of residuals at time 2



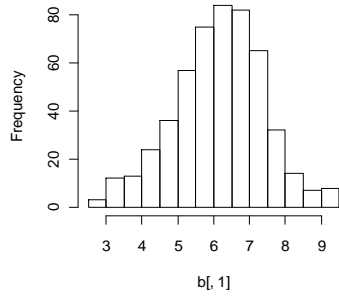
Q-Q plot of residuals at time 3



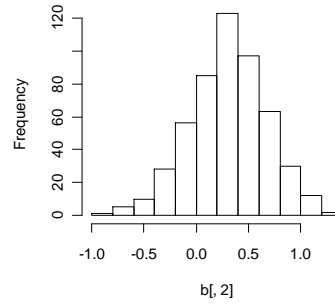
Q-Q plot of residuals at time 4



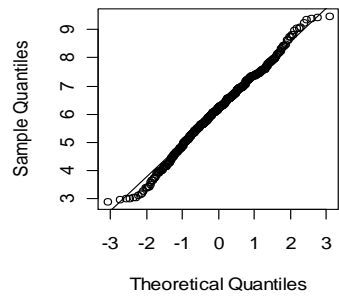
Histogram of b0



Histogram of b1



Q-Q plot of b0



Q-Q plot of b1

