Residual Diagnostics for General Growth Mixture Models: Examining the Impact of a Preventive Intervention on Multiple Trajectories of Aggressive Behavior [*]

This draft is under revision. Do not distribute.

Chen-Pin Wang
C. Hendricks Brown
University of South Florida

February 7, 2002

---

[*] An earlier version of this paper was presented at the 2001 Society for Prevention Research meeting. Work on this paper was supported by both the National Institute of Mental Health and the National Institute on Drug Abuse under grant MH40859, and by NIMH under grants No. MH01259, No. MH38725, and No. MH42968. The work has benefitted from many helpful discussions within the Prevention Science and Methodology Group. Especially, we wish to acknowledge Drs. Bengt Muthén, Karen Bandeen-Roche, Sheppard Kellam, and Nick Ialongo for their insightful comments. We are also thankful for Dr. Wei Wang's generous consultation in shaping up the proof of the theorem.

Abstract

This paper proposes three graphical methods for diagnosing general growth mixture models (GGMM; see Muthén and Shedden, 1999; Muthén et al., in press). GGMMs are built on a finite mixture of growth trajectory classes with individuals' scores deviating around the mean of one such curve. GGMM allows for covariates to predict class membership and to influence a curve's trajectory over time. They can be used to assess differential effects of an intervention on repeated measurements as well as on more distal outcomes (Muthén et al., in press). We develop new diagnostic methods to detect misspecification in GGMM regarding growth trajectory, covariance structure, and the number of classes. Our procedure applies the pseudo class technique (Bandeen-Roche, Miglioretti, Zeger, and Rathouz, 1997) to impute multiple class membership for individuals in the sample, and then for each pseudo class, forms diagnostic plots based on the empirical Bayes residuals at both level one and level two of the GGMM hierarchy. These methods are tested in simulation studies involving two classes of linear growths, each having a distinct covariance structure. They are then applied to longitudinal data from a randomized field intervention trial that tested whether children's trajectories of aggressive behavior could be changed elementary and middle school. Our diagnostics lead to a solution involving a mixture of three classes of growth curves.

Key words: Mixture modeling, growth modeling, preventive interventions, latent variables, trajectory classes, pseudo classes, maximum likelihood, empirical Bayes.

1 Introduction

Growth modeling has become a standard tool in the examination of intervention impact across time. Growth trajectories can be fitted using random effect modeling (REM; Laird and Ware, 1982; Longford, 1993), hierarchical modeling (Bryk and Raudenbush, 1992), or latent growth modeling (Muthén, 1983; Muthén, 1984; Muthén, 1996; Muthén, 1997; Muthén and Curran, 1997). These overlapping literatures use random effects to represent an individual's linear or otherwise smooth growth curve; covariates, such as intervention conditions, can be used to account for variations in these smooth curves. Recently, new methods have been developed to examine multiple developmental pathways in a population (Muthén and Shedden, 1999; Nagin, 1999; Muthén et al., in press). Such modeling of growth mixtures is consisted of three levels: level 1 involving individual level measurement errors at each point in time across time, level 2 involving continuous latent variables that specify the variation within each growth class, and level 3 involving discrete latent classes that represent different growth patterns. Different covariates may be included at all three levels. This approach thus provides an explicit way to identify different patterns of growth trajectories and the impact of an intervention on different subsets of a population. General growth mixture modeling provides a broad class of models to accommodate both time dependent and independent predictors as well as continuous, categorical, or time to event distal outcomes (GGMM; see Muthén et al., in press). However, model checking for these growth mixture models are inherently challenging because their involvement with both unobserved latent discrete and continuous variables.

There are different ways that model fit can be assessed, including overall goodness-of-fit indices (e.g., the likelihood ratio test), model comparison measures (e.g., Bayesian information criterion; Schwarz, 1978 and C_p, Mallows, 1973), and residual diagnostics, the subject of this paper. Residual diagnostics involve identifying patterns of model departures, detecting specific forms of model misspecification, and directing analysts towards to a better fitting model. Such methods have been well developed for linear regression models (Cook and Weisberg, 1982; Hoaglin, Mosteller and Tukey, 1982), generalized linear models (McCullagh and Nelder, 1989), as well as for REMs (Waternaux, Laird, and Ware, 1989; Bates and Pinheiro, 1998). They can be effective for probing the influence of single observation (Beckman, et al., 1987; Lange and Ryan, 1989; Bradlow and Zaslavsky, 1997; Cook and Weisberg, 1982), collinearity or lack of covariates (Hodge, 1998; Cook, 1998; Atkinson and Riani, 1998), and departure from normality (Dempster and Ryan, 1985; Lange and Ryan, 1989) or variance assumptions (McCullagh and Nelder, 1989, pp. 400-401). These cited techniques cannot be applied directly to GGMM because the growth classes are not observed. We adapt the pseudo class technique (Bandeen-Roche, Miglioretti, Zeger, and Rathouz; 1997; Nagin, 1999), and extend its use to our GGMM framework.

Model diagnostics have also been used to detect a single or a cluster of outliers (Waternaux, Laird, and Ware, 1989; Weiss and Lazaro, 1992; Longford and Lewis, 1998), and guide the fit for these outlying data. Often outliers form a cluster which in GGMM, can be represented by a new latent growth class. Consequently our approach relies on checking the adequacy of the mixture components rather than testing for outlying observations, a perspective similar to that in Lindsay and Roeder (1992).

Section 2 presents the GGMM formulation using notations similar to that of Laird and Ware (1982) for REMs. This differs from the original GGMM framework (Muthén et al., 2001), which was developed from the structural equation tradition (Bollen, 1989).

The major contribution of this paper includes deriving asymptotic properties associated with pseudo classes in the GGMM context and developing diagnostic procedures for examining GGMM misspecification in growth trajectories, covariance structures, and the number of growth classes. Section 3 presents the calculation for pseudo class probabilities and residuals based on empirical Bayes methods. We also examine the asymptotic distributions of pseudo-class adjusted residuals. In Section 4, three graphical tools are developed to examine the distributions of pseudo class adjusted residuals: one being the time trend plot for residuals, and two involving the quantile-quantile plot technique (Chambers, Cleveland, Kleiner, and Tukey, 1983). These methods have also been tested in simulations (see Section 4), and were then applied to longitudinal data from a randomized preventive trial aimed at reducing aggression in elementary school children (see Section 5).

We adopt the following notational convention throughout the paper: i.i.d. for "independently and identically distributed"; p.d.f. for "probability density function"; p.m.f. for "probability mass function"; $E_{\mathsf{p}}(.)$ for "expectation of . with respect to probability measure p"; $\mathbf{0}_l$ for the $l \times 1$ zero vector; $\mathbf{I}_l$ for the identity matrix of dimension $l \times l$; χ_l for a Chi-square variable with degree of freedom l; $stac(., \cdots, .)$ for a vector that stacks all elements in the parentheses; $\mathcal{MN}(\mathbf{u}, \mathbf{V})$ for the multivariate normal distribution with mean $\mathbf{u}$ and covariance $\mathbf{V}$; $\phi(\mathbf{u}, \mathbf{V}) \equiv \{(\sqrt{2}\pi)^{l_u}|\mathbf{V}|\}^{-.5} \exp\{-\frac{1}{2}\mathbf{u}'\mathbf{V}^{-1}\mathbf{u}\}$, where l_u is the length of $\mathbf{u}$; $\xrightarrow{L}$ for "converges in law to as $n \to \infty$"; $\xrightarrow{c}$ for "converges completely as $n \to \infty$"; and $\xrightarrow{a.s.}$ for " converges almost surely to as $n \to \infty$".

2 The General Growth Mixture Model

The general growth mixture model (GGMM; Muthén et al., 2001) can be used to examine multiple developmental trajectories within a population. It assumes a mixture of parametric distributions to characterize a population's multiple growth trajectories. The discrete developmental pathways are labelled as latent trajectory classes since they are all unknown. Within each trajectory class, it also hypothesizes that individuals' curves depart from the class average by the makeup of their unique measurement errors and the so-called latent growth variables.

We assume measurement times are the same for everyone, and initially assume no missing data. Denote $\mathbf{y}_i$ ($l_y \times 1$) as the longitudinal outcomes for subject i for $i = 1, \cdots, n$. Let C_i represent the discrete latent growth class for subject i, which takes values of $1, \cdots, K$ for $i = 1, \cdots, n$. At the first level in the model, let each $\mathbf{y}_i$, given $C_i = k$, be modeled as a sum of a fixed component $\boldsymbol{\mu}_{ik}$, a random component $\mathbf{r}_{ik}$ associated with within class variation, and a random error component $\mathbf{e}_{ik}$

$$\mathbf{y}_i|_{[C_i=k]} \quad = \quad \boldsymbol{\mu}_{ik} + \mathbf{r}_{ik} + \mathbf{e}_{ik}, \tag{1}$$

where $\mathbf{e}_{ik} \overset{i.i.d.}{\sim} \mathcal{MN}(\mathbf{0}_{l_y}, \boldsymbol{\Theta}_k)$.

The second level involves modeling $\boldsymbol{\mu}_{ik}$ and $\mathbf{r}_{ik}$ as functions of level two covariates $\mathbf{X}_{ik}$ and $\mathbf{Z}_{ik}$, and their associated fixed parameters $\boldsymbol{\alpha}_k$ ($l_{\alpha_k} \times 1$) and latent growth variables $\mathbf{b}_{ik}$ ($l_{b_k} \times 1$), respectively, such that

$$\boldsymbol{\mu}_{ik} = \mathbf{X}_{ik}\boldsymbol{\alpha}_k, \text{ and } \mathbf{r}_{ik} = \mathbf{Z}_{ik}\mathbf{b}_{ik}, \tag{2}$$

where $\mathbf{b}_{ik} \overset{i.i.d.}{\sim} \mathcal{MN}(\boldsymbol{\beta}_k, \boldsymbol{\Psi}_k)$.

Using polytomous logistic modeling, the marginal probability of C_i is expressed in terms of individual level covariates $\mathbf{W}_i^c = (\mathbf{W}_{i1}^c, \cdots, \mathbf{W}_{iK}^c)$ and unknown fixed parameters $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K)$ such that

$$Pr(C_i = k | \mathbf{W}_i^c, \boldsymbol{\gamma}) = \frac{\exp(\mathbf{W}_{ik}^c \boldsymbol{\gamma}_k)}{\sum_{k'=1}^{K} \exp(\mathbf{W}_{ik'}^c \boldsymbol{\gamma}_{k'})}, \tag{3}$$

where $\boldsymbol{\gamma}_1 = \mathbf{0}_{l_{\gamma_1}}$ is set to assure model identifiability. We also assume that $C_1, \cdots, C_n$ are mutually independent given $(\mathbf{W}_1^c, \cdots, \mathbf{W}_n^c)$ and $\boldsymbol{\gamma}_1$.

In addition, we write for $i = 1, \cdots, n$, and $k = 1, \cdots, K$,

$$\mathbf{X}_{ik} = \boldsymbol{\Lambda}_k^f \mathbf{W}_{ik}^f \text{ and } \mathbf{Z}_{ik} = \boldsymbol{\Lambda}_k^r \mathbf{W}_{ik}^r, \tag{4}$$

where $\boldsymbol{\Lambda}_k^f$ and $\boldsymbol{\Lambda}_k^r$ are time measurements associated with $\boldsymbol{\mu}_{ik}$ and $\mathbf{r}_{ik}$, and $\mathbf{W}_{ik}^f$ and $\mathbf{W}_{ik}^r$ are individual level covariate matrices associated with these effects.

In situations where some $\mathbf{y}_i$s contain missing values due to an ignorable missing mechanism (Little and Rubin, 1987), we replace equations in (2) and (4) with their appropriate marginal structures for the nonmissing components. Thus in general, we write (1), (2) and (4) together as

$$\mathbf{y}_{ik} = \boldsymbol{\mu}_{ik} + \mathbf{r}_{ik}, \quad \boldsymbol{\mu}_{ik} = \boldsymbol{\Lambda}_{ik}^f \mathbf{W}_{ik}^f \boldsymbol{\alpha}_k, \quad \text{and} \quad \mathbf{r}_{ik} = \boldsymbol{\Lambda}_{ik}^r \mathbf{W}_{ik}^r \mathbf{b}_k, \tag{5}$$

where $\mathbf{e}_{ik} \sim \mathcal{MN}(\mathbf{0}_{l_y}, \boldsymbol{\Theta}_{ik})$ and $\mathbf{b}_{ik} \sim \mathcal{MN}(\boldsymbol{\beta}_k, \boldsymbol{\Psi}_k)$.

Unless otherwise mentioned, we will assume (5) and (3).

As an example, we construct a K-class GGMM for the intervention study involving treatment and control groups. Let w_i denote the indicator of intervention status for individual i, where $w_i = 1$ if the i^{th} individual is assigned to the intervention setting, and $w_i = 0$ otherwise. Let $0, 1, \cdots, (l_y - 1)$ be the time points at which the outcomes are measured. Suppose that data are completely observed at all l_y occasions. Assume quadratic growth curves for all trajectory classes, and that both intercepts and slopes to be random and quadratics to be fixed. Then the GGMM components in (1)-(3) are given as follows:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i(l_y-1)} \end{bmatrix}. \ \boldsymbol{\Theta}_k = \begin{bmatrix} \theta_{k11} & \theta_{k12} & \cdots & \theta_{k1l_y} \\ \theta_{k21} & \theta_{k22} & \cdots & \theta_{k2l_y} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{kl_y1} & \theta_{kl_y2} & \cdots & \theta_{kl_yl_y} \end{bmatrix}. \ \boldsymbol{\Psi}_k = \begin{bmatrix} \psi_{k11} & \psi_{k12} \\ \psi_{k21} & \psi_{k22} \end{bmatrix}.$$

$$\boldsymbol{\Lambda}^f_{ik} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ \vdots & & \\ 1 & (l_y-1) & (l_y-1)^2 \end{bmatrix}. \ \mathbf{W}^f_{ik} = \begin{bmatrix} 1 & 0 & 0 & w_i & 0 & 0 \\ 0 & 1 & 0 & 0 & w_i & 0 \\ 0 & 0 & 1 & 0 & 0 & w_i \end{bmatrix}. \ \boldsymbol{\alpha}_k = \begin{bmatrix} \alpha_{k1} \\ \alpha_{k2} \\ \alpha_{k3} \\ \alpha_{k4} \\ \alpha_{k5} \\ \alpha_{k6} \end{bmatrix}.$$

$$\boldsymbol{\Lambda}^r_{ik} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \\ 1 & (l_y-1) \end{bmatrix}. \ \mathbf{W}^r_{ik} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \ \boldsymbol{\beta}_k = \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \end{bmatrix}.$$

$$\mathbf{W}^c_{ik} = [w_i]. \ \boldsymbol{\gamma}_k = [\gamma_k].$$

In the above specification, $(\beta_{k1}+\alpha_{k1}, \alpha_{k2}+\beta_{k2}, \alpha_{k3})'$ corresponds to the class-k mean growth factors (intercepts, slopes and quadratics) for the control group; $(\alpha_{k4}, \alpha_{k5}, \alpha_{k6})'$ corresponds to the vector of intervention impact on class-k growth; and γ_k is interpreted as the log-odds of class k probability relative to class 1 probability. For randomized trials, we would typically assume that intervention and control groups have the same baseline distributions by setting $\alpha_{k4} = 0$.

All GGMMs considered in this paper can be fitted using Mplus software (Muthén and Muthén, 1999), which applies the EM algorithm (Dempster, Laird, and Rubin, 1985) to compute maximum likelihood estimates (MLE) for parameters based on the marginal likelihood,

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\Psi}; \mathbf{y}) = \prod_{i=1}^{n} f(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i^c, \boldsymbol{\Xi}), \tag{6}$$

where $\boldsymbol{\Xi} = stac(\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_K)$, $\boldsymbol{\Xi}_k = stac(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \boldsymbol{\Psi}_k, \boldsymbol{\Theta}_k)$, $\mathbf{X}_i = stac(\mathbf{X}_{i1}, \cdots, \mathbf{X}_{iK})$, $\mathbf{Z}_i = stac(\mathbf{Z}_{i1}, \cdots, \mathbf{Z}_{iK})$, and $\boldsymbol{\Omega}_{ik} = \mathbf{X}'_{ik}\boldsymbol{\Psi}_k\mathbf{X}_{ik} + \boldsymbol{\Theta}_{ik}$, $f(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i^c, \boldsymbol{\Xi}) = \left\{ \sum_{k=1}^{K} \left(\frac{e^{\mathbf{W}_{ik}^c \boldsymbol{\gamma}_k}}{\sum_{k'=1}^{K} e^{\mathbf{W}_{ik'}^c \boldsymbol{\gamma}_{k'}}} f_k(\mathbf{y}_i) \right) \right\}$, and $f_k(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i^c, \boldsymbol{\Xi}) = \phi(\mathbf{y}_i - \mathbf{Z}_{ik}\boldsymbol{\beta}_k - \mathbf{X}_{ik}\boldsymbol{\alpha}_k, \boldsymbol{\Omega}_{ik})$.

3 Pseudo Classes and Empirical Bayes Residuals

3.1 Pseudo Classes

To obtain residual diagnostics for GGMM, we will need to assign growth classes to individuals. The "pseudo class" technique (Bandeen-Roche et al.; 1997) is employed for such estimation. Originally introduced in the latent class framework, it assigns individuals to classes with weights determined by the conditional probability that a subject comes from that class given the data, with unknown parameters being evaluated at their maximum likelihood solution. That is, if $C_i^*(\equiv C_i|_{[\mathbf{y}_i]})$ denotes the pseudo class for subject i, then its p.m.f. under (3) and (5) is given by

$$\begin{aligned} & Pr(C_i^* = k | \mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\ = & \frac{\phi(\mathbf{y}_i - \mathbf{X}_{ik}\hat{\boldsymbol{\beta}}_k - \mathbf{Z}_{ik}\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\Omega}}_{ik}) Pr(C_i = k | \mathbf{W}_i^c, \hat{\gamma})}{\sum_{k'=1}^{K} \phi(\mathbf{y}_i - \mathbf{X}_{ik'}\hat{\boldsymbol{\beta}}_{k'} - \mathbf{Z}_{ik'}\hat{\boldsymbol{\alpha}}_{k'}, \hat{\boldsymbol{\Omega}}_{ik'}) Pr(C_i = k' | \mathbf{W}_i^c, \hat{\gamma})}. \end{aligned} \tag{7}$$

The asymptotic properties associated with a single replication of $\mathbf{C}^*$ has been examined by Bandeen-Roche et al. (1997) in the latent class model context; data from each pseudo class behave just like samples drawn from the true classes if the model is correctly specified. As shown below, their result can be extended to our GGMM diagnostics. In addition, we demonstrate the improvement by using multiple $\mathbf{C}^*$.

Theorem Suppose that $\mathbf{y}_i, \cdots, \mathbf{y}_n$ are independently distributed according to (3) and (5), where the parameters $\boldsymbol{\Xi}$ indexing these distributions belong to a set $\mathbf{Q}$ and satisfy the standard identifiability condition (up to permutation of mixture components). Consider fitting a model for $\mathbf{y}_i, \cdots, \mathbf{y}_n$ by obtaining $\hat{\boldsymbol{\Xi}}$ that maximizes (6) within the interior of $\mathbf{Q}$. Define a general diagnostic transformation τ such that $\tau_i = \tau(\mathbf{y}_i, C_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\Xi})$ for $i = 1, \cdots, n$, where $\tau_i|_{[C_i=k,\boldsymbol{\Xi}]} \overset{i.i.d.}{\sim} G_k$. Also, assume that τ is continuous with respect to $\mathbf{y}_i$s and $\boldsymbol{\Xi}$. Denote $\tilde{G}_{kh}$ as the empirical c.d.f. of $\tau_i|_{[C^*_{i_h}=k,\hat{\boldsymbol{\Xi}}]}$, and $\tilde{\xi}_{kh}$ as the associated pth quantiles. Then

$$\tilde{G}_{kh} \xrightarrow{c} G_k, \tag{8}$$

and

$$n^{\frac{1}{2}}(\tilde{\xi}_{kh} - \xi_k) \xrightarrow{L} \mathcal{N}(0, \sigma^2_{\xi}) \tag{9}$$

if $|\frac{\partial g_k}{\partial \boldsymbol{\Xi}}| < \infty \;\; a.s.$, where $\sigma^2_{\xi} = [\frac{\partial \xi_k}{\partial \boldsymbol{\Xi}}]' E_f[\frac{\partial^2 \log(6)}{\partial \boldsymbol{\Xi}^2}][\frac{\partial \xi_k}{\partial \boldsymbol{\Xi}}]$ and $g_k(.) = \frac{\partial G_k}{\partial .}$.

Moreover, denote $\mu_k = \int \tau \, dG_k(\tau)$ and $\tilde{\mu}_{kh} = \int \tau \, d\tilde{G}_{kh}(\tau)$, where τ is uniformally integrable with respect to both G_k and $\tilde{G}_{kh}$, and both μ_k and $\tilde{\mu}_{kh}$ are continuous with respect to $\boldsymbol{\Xi}$. Then for $1 \leq h \leq H$ and $1 \leq k \leq K$,

$$\tilde{\mu}_{kh} \xrightarrow{a.s.} \mu_k. \tag{10}$$

A proof is given in the Appendix.

To apply the theorem to our residual diagnostics, we consider special τs such that for each $\mathbf{y}_i$, $\tau_i|_{[C_i=k]} = \mathbf{m}'_{ik}[\mathbf{y}_i - E_{f_k}(\mathbf{y}_i)]$ or $\tau_i|_{[C_i=k]} = [\mathbf{y}_i - E_{f_k}(\mathbf{y}_i)]'\mathbf{M}_{ik}[\mathbf{y}_i - E_{f_k}(\mathbf{y}_i)]$, where $\mathbf{m}_{ik}$ and $\mathbf{M}_{ik}$ are a weight vector of length l_y and a weight matrix of rank l_y, respectively. Results given in (8) and (10) imply that for each class k, its associated residual distribution G_k and the residual mean μ_k can be well approximated via $\tilde{G}_{kh}$ and $\tilde{\mu}_{kh}$ despite the absence of $\mathbf{C}$.

Further, to improve diagnostics, we recommend using the residuals averaged over multiple pseudo class draws, similar in spirit to multiple imputation. According to Rubin (1987) and Schafer (1997), 2 to 5 multiple imputations are usually sufficient to achieve satisfactory relative efficiency (RE) for most missing data problems where the fraction of missing information λ (Orchard and Woodbury, 1972) is below 0.2; the asymptotic RE associated with H imputation is given as $(1 + \lambda/H)^{-1}$. Such is not the case here since completely missingness of $\mathbf{C}$ results in a high λ. In the examples discussed in Sections 3 and 4, λs range from 0.5 to 0.8, and therefore, taking 20 pseudo class draws will ensure to achieve 95% relative efficiency (RE).

Next we establish appropriate τ for our GGMM diagnostics.

3.2 Residuals at Level 1 and Level 2

In line with the approach by Gruttola, Ware, and Louis (1987) and Weiss (1995), we consider decomposing $\mathbf{y}_i - \mathbf{X}_{ik}\boldsymbol{\alpha}_k - \mathbf{Z}_{ik}\boldsymbol{\beta}_k$ as $\mathbf{R}^{(1)}_{ik} + \boldsymbol{\Lambda}^r_{ik}\mathbf{R}^{(2)}_{ik}$, where $\mathbf{R}^{(1)}_{ik} = \mathbf{y}_i - \mathbf{X}_{ik}\boldsymbol{\alpha}_k - \mathbf{Z}_{ik}\mathbf{b}_{ik}$ and $\mathbf{R}^{(2)}_{ik} = \mathbf{W}^r_{ik}\mathbf{b}_{ik} - \mathbf{W}^r_{ik}\boldsymbol{\beta}_k$. We define $\mathbf{R}^{(1)}_{ik}$ and $\mathbf{R}^{(2)}_{ik}$ as level 1 residuals and level 2 residuals for individual i with respect to class k.

Note that each component of $\mathbf{R}^{(1)}_{ik}$ and $\mathbf{R}^{(2)}_{ik}$ constitutes a valid τ stated in the above Theorem provided $\mathbf{W}^f_{1k} = \cdots = \mathbf{W}^f_{nk}$, $\boldsymbol{\Lambda}^f_{1k} = \cdots, \boldsymbol{\Lambda}^f_{nk}$, $\mathbf{W}^r_{1k} = \cdots = \mathbf{W}^r_{nk}$, and $\boldsymbol{\Lambda}^r_{1k} = \cdots =$

$\boldsymbol{\Lambda}^r_{nk}$ for all k. We therefore consider rescaling these residuals so that the resulting τs (i.e., $S^{(\cdot)}_{\cdot\cdot}$s and $\mathcal{D}^{(\cdot)}_{\cdot\cdot}$s defined later in this Section) have distributions that are free from $\mathbf{W}^{\cdot}_{\cdot\cdot}$s and $\boldsymbol{\Lambda}^{\cdot}_{\cdot\cdot}$s.

Since the calculation for $\mathbf{R}^{(1)}_{ik}$ and $\mathbf{R}^{(2)}_{ik}$ involves unknown quantities $\{\mathbf{b}_{ik} : i = 1, \cdots, n, \;\; k = 1, \cdots, K\}$, we use the EB formula given in Lange and Ryan (1989) to calculate these estimates.

The posterior distribution of $\mathbf{R}^{(2)}_{ik}$ under (3) and (5) is

$$\mathbf{R}^{(2)}_{ik}|_{[\mathbf{y}_i,\mathbf{X}_i,\mathbf{Z}_i,\mathbf{W}^c_{ik},\boldsymbol{\Xi}_k,C_i=k]} \sim \mathcal{MN}(\mathbf{W}^r_{ik}\mathbf{b}^B_{ik} - \mathbf{W}^r_{ik}\boldsymbol{\beta}_k, \mathbf{V}^{(2)}_{ik}(\mathbf{y}_i)), \tag{11}$$

where

$$\begin{aligned}
\mathbf{W}^r_{ik}\mathbf{b}^B_{ik} &= \mathcal{V}^{(2)}_{ik}\left(\{\mathbf{B}_{ik}\boldsymbol{\Theta}_{ik}[\mathbf{B}_{ik}]'\}^{-1}\mathbf{B}_{ik}[\mathbf{y}_i - \mathbf{X}_{ik}\boldsymbol{\alpha}_k] + \{\mathbf{W}^r_{ik}\boldsymbol{\Psi}_k[\mathbf{W}^r_{ik}]'\}^{-1}\mathbf{W}^r_{ik}\boldsymbol{\beta}_k\right), \\
\mathbf{V}^{(2)}_{ik}(\mathbf{y}_i) &= \left(\{\mathbf{B}_{ik}\boldsymbol{\Theta}_{ik}[\mathbf{B}_{ik}]'\}^{-1} + \{\mathbf{W}^r_{ik}\boldsymbol{\Psi}_k[\mathbf{W}^r_{ik}]'\}^{-1}\right)^{-1}. \\
\mathbf{B}_{ik} &= \{[\boldsymbol{\Lambda}^r_{ik}]'\boldsymbol{\Lambda}^r_{ik}\}^{-1}[\boldsymbol{\Lambda}^r_{ik}]'.
\end{aligned}$$

Using (11) and the asymptotic property of $\hat{\boldsymbol{\Xi}}$, we have

$$[\hat{\mathbf{V}}^{(2)}_{ik}]^{-\frac{1}{2}}\hat{\mathbf{R}}^{(2)}_{ik}|_{[C_i=k]} \xrightarrow{L} \mathcal{MN}(\mathbf{0}_{l_{b_k}}, \mathbf{I}_{l_{b_k}}), \tag{12}$$

where $\hat{\mathbf{R}}^{(2)}_{ik} = \mathbf{W}^r_{ik}\hat{\mathbf{b}}^B_{ik} - \mathbf{W}^r_{ik}\hat{\boldsymbol{\beta}}_k$ and

$$\hat{\mathbf{V}}^{(2)}_{ik} = \left([\mathbf{V}^{(2)}_{ik}(\mathbf{y}_i)]^{-1}\{\mathbf{B}_{ik}\hat{\boldsymbol{\Theta}}_{ik}[\mathbf{B}_{ik}]'\}^{-1}\mathbf{B}_{ik}\right)\hat{\boldsymbol{\Omega}}_{ik}\left([\mathbf{V}^{(2)}_{ik}(\mathbf{y}_i)]^{-1}\{\mathbf{B}_{ik}\hat{\boldsymbol{\Theta}}_{ik}[\mathbf{B}_{ik}]'\}^{-1}\mathbf{B}_{ik}\right)'.$$

We standardize these level 2 residual estimates in two ways based on (12). For each subject i, we define the componentwise standardization for level 2 residuals $\hat{\mathbf{R}}^{(2)}_{ik}$ to be $\hat{\mathbf{S}}^{(2)}_{ik} = (\hat{\mathbf{V}}^{(2)}_{ik})^{-\frac{1}{2}}\mathbf{I}_{l_{b_k}}\hat{\mathbf{R}}^{(2)}_{ik}$. The multivariate standardization for $\hat{\mathbf{R}}^{(2)}_{ik}$ is defined as $\hat{D}^{(2)}_{ik} = [\hat{\mathbf{R}}^{(2)}_{ik}]'[\hat{\mathbf{V}}^{(2)}_{ik}]^{-1}[\hat{\mathbf{R}}^{(2)}_{ik}]$, which is also the Mahalanobis distance for $\hat{\mathbf{R}}^{(2)}_{ik}$.

Similarly, the EB level 1 residuals are distributed as

$$\mathbf{R}^{(1)}_{ik}|_{[\mathbf{y}_i,\mathbf{X}_i,\mathbf{Z}_i,\mathbf{W}^c_{ik},\boldsymbol{\Xi}_k,C_i=k]} \sim \mathcal{MN}(\mathbf{y}_i - \mathbf{X}_{ik}\boldsymbol{\alpha}_k - \mathbf{Z}_{ik}\mathbf{b}^B_{ik}, \mathbf{V}^{(1)}_{ik}(\mathbf{y}_i)), \tag{13}$$

where $\mathbf{V}_{ik}^{(1)}(\mathbf{y}_i) = \left([\boldsymbol{\Theta}_{ik}]^{-1} + \{\mathbf{Z}_{ik}\boldsymbol{\Psi}_k[\mathbf{Z}_{ik}]'\}^{-1}\right)^{-1}$.

Consequently, their associated EB estimates are $\hat{\mathbf{R}}_{ik}^{(1)} = \mathbf{y}_i - \mathbf{X}_{ik}\hat{\boldsymbol{\alpha}}_k - \mathbf{Z}_{ik}\hat{\mathbf{b}}_{ik}^{B}$, and

$$[\hat{\mathbf{V}}_{ik}^{(1)}]^{-\frac{1}{2}}\hat{\mathbf{R}}_{ik}^{(1)}|_{[C_i=k]} \xrightarrow{L} \mathcal{MN}(\mathbf{0}_{l_y}, \mathbf{I}_{l_y}), \tag{14}$$

for $i = 1, \cdots, n$ and $k = 1, \cdots, K$, where $\hat{\mathbf{V}}_{ik}^{(1)} = [\mathbf{A}_{ik}]\,\hat{\boldsymbol{\Omega}}_{ik}\,[\mathbf{A}_{ik}]'$ and

$$\mathbf{A}_{ik} = \mathbf{I}_{l_y} - \boldsymbol{\Lambda}_{ik}^{r}\mathbf{V}_{ik}^{(1)}(\mathbf{y}_i)\{\mathbf{B}_{ik}\boldsymbol{\Theta}_{ik}[\mathbf{B}_{ik}]'\}^{-1}\mathbf{B}_{ik}.$$

The componentwise standardization for level 1 residuals $\hat{\mathbf{R}}_{ik}^{(1)}$ and its corresponding Mahalanobis distance are defined as $\hat{\mathbf{S}}_{ik}^{(1)} = (\hat{\mathbf{V}}_{ik}^{(1)})^{-\frac{1}{2}}\mathbf{I}_{l_{y_i}}\hat{\mathbf{R}}_{ik}^{(1)}$ and $\hat{D}_{ik}^{(1)} = [\hat{\mathbf{R}}_{ik}^{(1)}]'[\hat{\mathbf{V}}_{ik}^{(1)}]^{-1}\hat{\mathbf{R}}_{ik}^{(1)}$.

Note that the multivariate Mahalanobis residuals are asymptotically multiples of chi-square with degree of freedom depending on the amount of data observed for that case. To overcome this problem, we recommend applying the Wilson-Hilferty (W-H) transformation (Johnson and Kotz, 1970) to $\hat{D}_{..}^{(.)}$. Specifically, W-H cube-root transformation yields $\hat{\mathcal{D}}_{ik}^{(j)} = \left((\frac{\hat{D}_{ik}^{(j)}}{\nu})^{\frac{1}{3}} - 1 + \frac{2}{9\nu}\right) * (\frac{9\nu}{2})^{\frac{1}{2}}$, where ν is the number of elements in the residual vector, and $\hat{\mathcal{D}}_{ik}^{(j)}|_{[C_i=k]} \xrightarrow{L} \mathcal{MN}(0,1)$ for all possible i, k, j when the assumed model is correct. This transformation also reduces clear skewness in the Mahalanobis residuals and improves the readability of Q-Q plots.

In what follows, we rely on the pseudo-class adjusted residuals $\hat{\mathbf{R}}_{.k}^{(.)}|_{[C_.^*=k]}$, $\hat{\mathbf{S}}_{.k}^{(.)}|_{[C_.^*=k]}$, and $\hat{\mathcal{D}}_{.k}^{(.)}|_{[C_.^*=k]}$ to make inference about $\mathbf{R}_{.k}^{(.)}|_{[C_.=k]}$, $\mathbf{S}_{.k}^{(.)}|_{[C_.=k]}$, and $\mathcal{D}_{.k}^{(.)}|_{[C_.=k]}$, respectively.

4 Methods and Simulation Studies

Three types of diagnostic plots are proposed to detect model misspecification in growth trajectory, covariance structure, and the number of classes. All the diagnostic quantities calculated below are the average of multiple pseudo class draws with H chosen to be sufficient for 95% RE. Details of these methods are described below.

We use three simulation studies to test each particular model misspecification. The GGMM that generated the simulated data (of size 500) assumes $\gamma_1 = \gamma_2 = 0$, $\boldsymbol{\beta}_1 = (0, .3)'$, $\boldsymbol{\beta}_2 = (2, 1)'$, $\boldsymbol{\Theta}_1 = .70 * \mathbf{I}_5$, $\boldsymbol{\Theta}_2 = .35 * \mathbf{I}_5$, $\boldsymbol{\Psi}_1 = \begin{bmatrix} 0.30 & 0 \\ 0 & 0.15 \end{bmatrix}$, $\boldsymbol{\Psi}_2 = \begin{bmatrix} 0.15 & 0 \\ 0 & 0.075 \end{bmatrix}$, and $\boldsymbol{\mu}_{i1} = \boldsymbol{\mu}_{i2} = \mathbf{0}_5$, $\mathbf{W}^f_{i1} = \mathbf{W}^r_{i1} = \mathbf{W}^c_{i1} = \mathbf{W}^f_{i2} = \mathbf{W}^r_{i2} = \mathbf{W}^c_{i2} = \mathbf{I}_2$, and $(\boldsymbol{\Lambda}^r_{i1})' = (\boldsymbol{\Lambda}^r_{i2})' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$, for $i = 1, \cdots, n$.

These data have been fitted with both correct and incorrect models using Mplus (Muthén and Muthén, 1999, Appendix 8). In our Mplus program, we placed no restrictions on $\boldsymbol{\beta}$s and $\boldsymbol{\gamma}$s for the correct model while forcing both $\boldsymbol{\Theta}$s and $\boldsymbol{\Psi}$s to be arbitrary diagonal matrices that may vary by class. The specifications for the incorrect models are given in the following subsections.

4.1 Examining Growth Trajectories

Two diagnostic procedures are proposed to determine whether the existing growth factors are adequate for characterizing the data trajectories. In the illustrative example, the incorrect model assumes no slope growth factor in both classes, that is, $\beta_{12} = \beta_{22} = 0$ and $\psi_{1,22} = \psi_{2,22} = 0$.

The first method examines the time trends of level 1 residual means averaging over H pseudo class draws, $\bar{\tilde{\boldsymbol{\mu}}}_k = H^{-1}\left\{\sum_{h=1}^{H} \tilde{\boldsymbol{\mu}}_{kh}\right\}$, for $k = 1, \cdots, K$, where $\tilde{\boldsymbol{\mu}}_{kh} = \frac{\sum_{i=1}^{n} \hat{\mathbf{R}}^{(1)}_{ik} \mathbf{1}_{[C^*_{i_h} = k]}}{\sum_{i=1}^{n} \mathbf{1}_{[C^*_{i_h} = k]}}$, and $\hat{\mathbf{R}}^{(1)}_{ik} = \mathbf{R}^{(1)}_{ik}|_{[C_i = k, \hat{\boldsymbol{\Xi}}]}$. For each growth class k, the vector $\boldsymbol{\mu}_k$ is plotted against the time axis. To address the variability of these estimates, we also add error bars to represent the variation associated with both within and between pseudo class draws using the calculation by Rubin and Schenker's (1986; see equation (1.3)). Should any of $\bar{\bar{\boldsymbol{\mu}}}_k$ exhibit noticeable departure from 0 , this points to misspecified fitting of time dependence on growth for that class.

The second method aims at examining the entire distribution of each element of $\hat{\mathbf{R}}_{ik}^{(1)}|_{[C_i=k]}$ using the quantile-quantile (Q-Q) norm plot. To do so, we compute empirical quantiles for each class [1]. According to (8), each of such, say the $p-$quantile for class k, can be well approximated by $\bar{\tilde{\xi}}_k = H^{-1}\sum_{h=1}^{H}\tilde{\xi}_{kh}$, where $\tilde{\xi}_{kh} = \inf\left\{\hat{R}_{ik.}^{(1)} : \frac{\sum_{C^*_{i'_h}=k}\mathbf{1}_{[\hat{R}_{i'k.}^{(1)}<\hat{R}_{ik.}^{(1)}]}}{\sum_{i'=1}^{n}\mathbf{1}_{[C^*_{i'_h}=k]}} = p\right\}$. Then for $k = 1, \cdots, K$, the Q-Q norm curves are formed by plotting those mean quantiles ($\bar{\tilde{\xi}}_k$s) against their standard normal counter parts. Associated with each u, such curve should align with $\mathsf{y} = \sqrt{[\hat{\mathbf{V}}_{ik}^{(1)}]_{uu}}\ \mathsf{x}$ (i.e, the straight line crossing the origin with slope $\sqrt{[\hat{\mathbf{V}}_{ik}^{(1)}]_{uu}}$) if the assumed model is correct. Otherwise, we suggest considering alternative growth curves. This approach, compared with the former, thus can be more insightful since it reveals the information that is beyond the first two moments of $\hat{R}_{\cdots}^{(1)}$ (see Section 4.3).

Note that the result based on $\hat{\mathbf{R}}_{.k}^{(.)}|_{[C_.^*=k]}$ is trustworthy in our simulations since it assumes $\mathbf{W}_{1k}^{f} = \cdots, \mathbf{W}_{nk}^{f}$ and $\boldsymbol{\Lambda}_{1k}^{f} = \cdots, \boldsymbol{\Lambda}_{nk}^{f}$, $\mathbf{W}_{1k}^{r} = \cdots, \mathbf{W}_{nk}^{r}$ and $\boldsymbol{\Lambda}_{1k}^{r} = \cdots, \boldsymbol{\Lambda}_{nk}^{r}$, for $k = 1, \cdots, K$.

Simulation Results

Figure 1 presents the time trend plots for $\bar{\bar{\boldsymbol{\mu}}}_k$s derived from the incorrect model (the error bars that present the total variation however, are too slim to be acknowledged). The result suggested that residual means at the first two time points were remarkedly lower than expected,

---

[1] The number of points chosen for producing the Q-Q plots is either the class size for the class-specific calculation (e.g., Sections 4.1 and 4.3), or the size of the smaller class when involving two classes (e.g., Section 4.2).

and the opposite for the last two time points. Figure 2 presents Q-Q norm plots for $\hat{R}^{(1)}_{iku}|_{[C^*_{i_h}=k]}$s under the incorrect model. It concluded similarly as Figure 1. Both time trend plots and Q-Q norm plots behave appropriately under the correct model.

[Insert Figures 1 & 2 Here.]

In accordance with the usual advice on these departure patterns, one may consider to add linear time predictors for model improvement, especially for the trajectory class associated with $\boldsymbol{\beta}_2$, $\boldsymbol{\Psi}_2$, and $\boldsymbol{\Theta}_2$.

Notice that GGMM incorrectly concludes that level 1 variance estimates are quadratic in time despite no improper constraints on the covariances being imposed in our Mplus program (see Table 1). In fact, such aliasing effect is common to all procedures in fitting longitudinal data not just GGMM when the underlying growth trajectories is incorrectly specified (Laird and Lange, 1989). In this example, the quadratic pattern in $\hat{\theta}_{k,uu}$ is due to the absolute bias $|\beta_{k,2}(u - \frac{l_y-1}{2})|$ in estimating the mean of y_{iu} given $C_i = k$.

[Insert Table 1 Here.]

4.2 Examining Covariance Structure

We proposed the paired-class empirical Q-Q plot (E-Q-Q plot) to investigate whether covariance structures should vary by growth class. The three incorrect models considered in our demonstration are called level 1 *homogeneous* GGMM, level 2 *homogeneous* GGMM, and two-level *homogeneous* GGMM. The term *homogeneous* is referred to equal covariances across trajectory class, while the within class covariances can be any that is positive definite.

This method was derived based on the comparison of $\{\hat{\mathcal{D}}_{ik}^{(j)} : C_{i_h}^{*} = k\}$s between possible class pairs. We justify this approach with the following observations. First, note that each $\hat{\mathcal{D}}_{ik}^{(j)}$ represents a weighted sum of squares of $\mathbf{y}_i - E_{f_k}(\mathbf{y}_i|\hat{\mathbf{\Xi}})$, where the weight functions depend on $\hat{\mathbf{\Psi}}$ and $\hat{\mathbf{\Theta}}$. According to Lange and Laird (1989) and our simulation result shown in Table 2, the impact on $(\hat{\mathbf{\Psi}}, \hat{\mathbf{\Theta}})$ due to misspecification of $\mathbf{\Psi}$ and $\mathbf{\Theta}$ can be quite substantial, and that may vary across growth classes. This impact then passes immediately on to $\hat{\mathcal{D}}_{ik}^{(j)}$. The test based on univariate type of residuals, such as $\hat{S}_{\cdots}^{(\cdot)}$, however, has been found not to be helpful for the diagnostics. Second, these quantities, irrespect to growth classes, all have a common asymptotic distribution $\mathcal{MN}(0, 1)$ when the GGMM is correctly specified. This thus motivates the paired-class comparison.

For each level j $(j = 1, 2)$ and class pair (k, k'), our E-Q-Q method draws the mean quantiles obtained from $\{\hat{\mathcal{D}}_{ik}^{(j)} : C_{i_1}^{*} = k\}$, …, $\{\hat{\mathcal{D}}_{ik}^{(j)} : C_{i_H}^{*} = k\}$ against those from $\{\hat{\mathcal{D}}_{ik'}^{(j)} : C_{i_1}^{*} = k'\}$, …, $\{\hat{\mathcal{D}}_{ik'}^{(j)} : C_{i_H}^{*} = k'\}$. The mean quantile calculation and the Q-Q plot technique follow the instruction given in Section 4.1. If such a curve deviates from $\mathsf{y} = \mathsf{x}$, it suggests that assuming $\mathbf{\Theta}_k = \mathbf{\Theta}_{k'}$ or $\mathbf{\Psi}_k = \mathbf{\Psi}_{k'}$ can be questionable.

Simulation Results

The results of paired-class E-Q-Q comparison based on $\hat{\mathcal{D}}_{\cdot\cdot}^{(\cdot)}$s are given in Figure 3. It shows that our method has faithfully reflected the source of heterogeneity (from either $\mathbf{\Theta}$s or $\mathbf{\Psi}$s) that was not adequately addressed by the model. More evidential departure from $\mathsf{y} = \mathsf{x}$ was found in the comparison between $\hat{\mathcal{D}}_{1.}^{(1)}$ and $\hat{\mathcal{D}}_{2.}^{(1)}$ for the level 1 *homogeneous* model, while such evidence was marked in the comparison between $\hat{\mathcal{D}}_{1.}^{(2)}$ and $\hat{\mathcal{D}}_{2.}^{(2)}$ for the level 2 *homogeneous* model.

[Insert Figure 3 and Table 2 Here.]

4.3 Examining The Number of Growth Classes K

The class-specific Q-Q norm method is employed for testing the adequacy of K based on the hypothesis that the normality of the standardized residuals $\hat{S}^{(j)}_{iku}|_{[C^*_{i_h}=k]}$ $(j = 1, 2)$ could be perturbed if K being under-specified. Also, we anticipate that such impact on level 1 or level 2 may depend on the degree each attributing to the mixture nature.

The incorrect model considered below is a 1-class growth model reduced from the correct model.

Simulation Results

The incorrect model discussed below is a 1-class growth model reduced from the 2-class correct model. The level 2 Q-Q norm diagnostic result for the 1-class model is given in Figure 4. Unlike those Q-Q norm curvatures that present a pure shift in location parameter or scale parameter from the standard normal (e.g., Figure 2), these, in fact, can be recognized as Q-Q norm curves for a mixture of normal variables. However, the level 1 assessment hardly informed any evidence regarding class inadequacy. This result, in conjuction with Table 3, suggests that misspecification at level 3 has stronger impact on level 2 than that on level 1 since the distinguishable nature among the growth classes arises mainly from level 2, which also happens to be a realistic assumption for most application data.

[Insert Figure 4 and Table 3 Here.]

4.4 Computation

The residual calculation and diagnostic plots were carried out using the statistical software Splus 2000 (1999). We have posted our Splus codes for these work under the Prevention Science Methodology Group website at

http : //www.biostat.coph.usf.edu/research/psmgold/ggmmdiag/ggmmdiagplot.html.

In the next Section, we demonstrate how these methods can be extended to more complex GGMMs involving missing data, covariates, and three classes.

5 An Application Example

In 1985 a large randomized field trial, involving 1196 children from the Baltimore City Public School System, was initiated through a partnership between the school system and Johns Hopkins University. For two years, children received the same intervention, either a classroom-based Good Behavior Game (GBG) intervention designed to reduce aggression and disruption, an enhanced classroom-based Mastery Learning intervention, or a control condition. Prior to the intervention, during the intervention, and at follow-up we obtained teacher ratings of aggressive behavior using the Teachers' Observation of Classroom Adaptation-Revised (TOCA-R; Kellam et al., 1975). Being the average of ten items, this scale ranged from 1 to 6 with higher scores indicating higher aggression. These were measured once in first grade before the intervention began, and three times during the intervention period of first and second grade: Spring of first grade and Fall and Spring of second grade. Thereafter until grade seven we interviewed teachers to obtain annual aggressive behavior ratings on the TOCA-R. Thus except for missing data, children were measured 9 times on the same aggression measure. For the purposes of this paper, we will focus on the 119 males who received the GBG intervention and their 80 matched controls. Previous analyses had revealed an early beneficial impact of the GBG intervention by the end of first grade.GBG reduced aggressive behavior more in boys who began the year as aggressive compared to boys who were less aggressive. Longer term impact was also seen at the sixth grade follow-up (ref??). Additional growth analyses had suggested that GBG's improvement in aggression was strongest among those who initially began with high aggression and essentially negligible for those with low aggression (Muthén and Curran, 1997). These effects are elucidated in this paper through the comprehensive GGMM analyses.

5.1 Setting Up A Preliminary GGMM for the Baltimore Prevention Trial

Again, we use $w_1, \cdots, w_n$ to denote the indicators of individual's intervention statuses. We labeled 0 as the entry time and 1 as the increment between adjacent school years. The 9 time points thus were coded as 0, 0.5, 1, 1.5, 2.5, 3.5, 4.5, 5.5 ,6.5, accordingly. To form a GGMM for analyzing this prevention trial, we first screened through individual development trajectories using the usual technique for data exploring. This preliminary result suggested the following: the growth shape should be no simpler than a quadratic; level 1 measurement errors may change with time; and variances of both intercept(s) and slope(s) at level 2 are nonzeros. At level one, we also posited autocorrelation(s) between Fall and Spring semesters of the first two grades to address the rater effect. No autocorrelation was assumed anywhere else.

5.2 Determining Number of Growth Class and Covariance Structures

We started with a 1-class latent growth model (or a standard random effect model). In this case, checking for the number of growth class was of primary interest. We used the Q-Q norm method to examine distributions of $\hat{S}^{(2)}_{\cdot\cdot\cdot}$. Since violation of level two normality was profound (see Figure 5), we investigated both 2-class and 3-class GGMMs for improvement. Here we only inspect the 3-class solutions since they appeared to fit the data better than the 2-class ones.

Below we entertain three potential 3-class GGMM candidates. All the three growth classes described below are termed as "high" class, "medium" class, and "low" class according to their overall TOCA-R scores (see Figure 10). Model 1 assumed *homogeneous* covariances at both level 1 and level 2. Model 2 differed from Model 1 by allowing the "low" class to have different variances than the other two. Model 3 further relaxed the *homogeneous* restriction put on Model 2. The differences among these three models are given in Table 4 in terms of the constrains specified for their parameters.

[Insert Figure 5 & 10 Here.]

Diagnosis for Model 1. Using the level 2 class-specific Q-Q norm plot, we first tested $K = 3$ assumption. Since there exhibited no dramatic violation in normality among residuals from either intervention group, increasing K beyond 3 was not strongly encouraged (see Figures 6 and 7).

We then examined the *homogeneous* covariance assumption using the paired-class E-Q-Q method based on $\hat{\mathcal{D}}^{(1)}_{\cdot\cdot}$s (see Figure 8). Among these comparisons, we found sound evidence indicating that level 1 covariance structure for the "low" class ($\mathbf{\Theta}_3$) was likely to be different from those of the rest two classes ($\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$), while such evidence was not significant at level 2 (see Figure 9).

Diagnosis for Model 2. Similar to our finding in Model 1, the result again showed decent support for $K = 3$ (see Figures 6 and 7). As being expected by our previous diagnosis, the modification in the covariance had turned to a better fit model (see Figures 8 and 9).

Diagnosis for Model 3. To confirm the adequacy of Model 2, Model 3 was further investigated. In fitting this model using Mplus, it was suggested that the likelihood was maximized when slope variance for the "high" class was zero ($\hat{\theta}_{1,22} = 0$). Thus this optimization is sought in a confined parameter space differing from that for Model 2 (see Table 4). This effort, according to Figures 8 and 9, did not seem to add new improvement from Model 2, but eluded that choosing proper constraints among parameters could be a crucial step in the search for an optimal model.

Based on the above evidence, we concluded Model 2 as our final model.

[Insert Figures 6, 7, 8, and 9 Here.]

Tables 5 and 6 summarize parameter estimates and their associated standard errors for the three fitted 3-class GGMMs.

[Insert Tables 5 & 6 Here.]

6 Discussion

The diagnostic methods proposed in this paper were developed based on pseudo-class adjusted residuals at level 1 and level 2 of GGMM and the theorem associated with these quantities (see Section 3). The argument established for multiple draws (or imputation, the terminology used in the missing data literature) of pseudo classes contrasts somewhat from the general property of multiple imputations whereby increasing the number of imputations beyond 5 does not result in much improvement for standard model parameter inferences (Rubin and Schenker, 1986). The reason we chose a large number of pseudo class draws was because we found it to be beneficial, particularly, for estimating extreme residual quantiles, whose distributions often tended to have wide spread due to their highly sensitivity to class assignments. However, by taking the results averaged across multiple draws, we can improve the reliability of such estimation, hence the quality of the diagnostics.

As demonstrated in Section 4, these diagnostic methods vary in terms of their abilities to tackle different types of model misspecification. Both time trend plot and class-specific Q-Q norm plot based on $\tilde{\hat{\mathbf{R}}}_{ik}^{(1)}|_{[C_i^*=k]}$ are useful for correcting growth trajectories. The paired-class comparison based on $\hat{\mathcal{D}}_{ik}^{(1)}|_{C_i^*=k}$ and $\hat{\mathcal{D}}_{ik}^{(2)}|_{C_i^*=k}$ is most helpful for checking the *homogeneity* of covariances across growth classes. The class-specific Q-Q norm plot based on $\hat{\mathbf{S}}_{ik}^{(2)}|_{C_i^*=k}$ can be used to check the sufficiency of growth classes. Specially, with the employment of W-H transformed Mahalanobis distances, our methods can be suitable even when the different patterns and percentages of missingness, due to some missing at random mechanism, are present in the data.

We would like to point out some features possessed by our residual diagnostics that are designated to complement those overall model fit indices. In terms of robustness, our methods are applicable for testing models that are not completely nested or parameter values are set at the boundary of their support while theory falls short to approve the likelihood ratio test (LRT) used in either instance. In addition to a model sorting mechanism built in most model fit indices, our methods carry the strength to point specific form(s) of model departure and to help correct the misspecification. Assume for the moment that LRT is roughly valid for testing the three 3-class GGMMs considered in Section 5. Formal MLE calculation for the three 3-class GGMMs yielded $(-2)*$(log-likelihood) equal to 3436.8, 3206.8, and 3186.4 with number of parameters 32, 41, and 53, respectively. This, according to LRT, suggests that the latter two models with varying covariances are far superior for the application data than the first one. In this sense, Both LRT and our laborious approach reach the same conclusion. What is gained from our residual plots are the rich insights about the model deficiency.

However, diagnostics for models involving both discrete and continuous random effects, like GGMM, are often more complicated than those for models containing simply fixed effects. Although this paper has demonstrated its appropriateness for examining certain features of growth mixture models, we consider that this and other related works (Nagin, 1999; Bandeen-Roche et al., 1997) are not the last word. For example, one promising avenue in line with our approach is to explore the use of generalized residuals as described by Dempster and Ryan (1985) and Lange and Ryan (1989). Also, a general procedure for determining the number of mixture components can be constructed upon Garel (2001) Lo, Mendell, and Rubin (2001), Ishwaran, James, and Sun (2001), and the references cited within.

Another concern is that many of our methods rely on the normality assumption. Serious departure from these assumptions may lead to erroneous results. For an example, Carlin et al. (2001) found a mixture of a point mass at infinity and a smooth distribution that best described the underlying random effects for smoking. Our methods would be hard pressed to identify such a model. Secondly, although not being found in our work, the ceiling or floor effects can be of concern for datasets when measurement scales are subject to bounds. For such situations, the modeling technique that includes such bounds, such as provided by Jones, Nagin, and Roeder (2001), may better compensate for certain departures from normality.

Appendix: *Proof of our Theorem*

Preceeding the main proof, we derive the following six propositions.

(i) $\hat{\boldsymbol{\Xi}}$ exists uniquely (up to the permutation of the mixture components), and $\hat{\boldsymbol{\Xi}} \overset{a.s.}{\to} \boldsymbol{\Xi}$. This is due to the assumption on $f(.)$ and Render (1981).

(ii) $Pr(C^*_{i_h} = k|\mathbf{y}_i, \mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \overset{a.s.}{\to} Pr(C_i = k|\mathbf{y}_i, \mathbf{W}_i, \boldsymbol{\Xi})$. This follows from (i), the invariance property of MLE, and continuity of $f(.)$ and $Pr(C_i|\mathbf{W}_i, \boldsymbol{\Xi})$.

(iii) $Pr(C^*_{i_h} = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \overset{a.s.}{\to} Pr(C_i = k|\mathbf{W}_i, \boldsymbol{\Xi})$. This is derived from

$$\begin{aligned}
& Pr(C^*_{i_h} = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\
\equiv\ & \int Pr(C^*_{i_h} = k|\mathbf{y}_i = \mathbf{v}, \mathbf{W}_i, \hat{\boldsymbol{\Xi}}) f(\mathbf{v}|\mathbf{W}_i, \hat{\boldsymbol{\Xi}})\, d\mathbf{v} \\
=\ & \int f_k(\mathbf{v}|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) Pr(C_i = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}})\, d\mathbf{v} \\
=\ & Pr(C_i = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\
\overset{a.s.}{\to}\ & Pr(C_i = k|\mathbf{W}_i, \boldsymbol{\Xi}) \quad \text{(according to (i))}.
\end{aligned}$$

(iv) $\mathbf{y}_i|_{[C^*_{i_h}=k, \hat{\boldsymbol{\Xi}}]} \sim F_k(.|\mathbf{W}_i, \hat{\boldsymbol{\Xi}})$. That is because

$$\begin{aligned}
& Pr(\mathbf{y}_i = \mathbf{v}, C^*_{i_h} = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\
=\ & Pr(C^*_{i_h} = k|\mathbf{y}_i = \mathbf{v}, \mathbf{W}_i, \hat{\boldsymbol{\Xi}}) f(\mathbf{v}|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\
=\ & f_k(\mathbf{v}|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) Pr(C_i = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}) \\
=\ & Pr(\mathbf{y}_i = \mathbf{v}, C_i = k|\mathbf{W}_i, \hat{\boldsymbol{\Xi}}).
\end{aligned}$$

(v) $\tau_i|_{[C^*_{i_h}=k, \hat{\boldsymbol{\Xi}}]} \sim \hat{G}_k$, where $\hat{G}_k(.)$ denotes the c.d.f. of $\tau_i|_{[C_i=k, \hat{\boldsymbol{\Xi}}]}$. This is due to (iv), the continuity of τ (with respect to $\mathbf{y}_i$), and for every measurable $\mathbf{A}$

$$Pr(\mathbf{y}_i \in \mathbf{A}, C^*_{i_h} = k, \mathbf{W}_i, \hat{\boldsymbol{\Xi}})$$

$$
\begin{aligned}
&= \frac{\int_{\mathbf{v}\in\mathbf{A}} Pr(C^*_{i_h}=k|\mathbf{y}_i=\mathbf{v},\mathbf{W}_i,\hat{\mathbf{\Xi}})f(\mathbf{v}|\mathbf{W}_i,\hat{\mathbf{\Xi}})\,d\mathbf{v}}{Pr(C^*_{i_h}=k|\mathbf{W}_i,\hat{\mathbf{\Xi}})} \\
&= \int_{\mathbf{v}\in\mathbf{A}} f_k(\mathbf{v}|\mathbf{W}_i,\hat{\mathbf{\Xi}})Pr(C_i=k|\mathbf{W}_i,\hat{\mathbf{\Xi}})\,d\mathbf{v}.
\end{aligned}
$$

(vi) $\tau_i|_{[C^*_{i_h}=k,\hat{\mathbf{\Xi}}]} \xrightarrow{L} \tau_i|_{[C_i=k,\mathbf{\Xi}]}$. It follows

$$
\begin{aligned}
& Pr(\mathbf{y}_i\in\mathbf{A}|C^*_{i_h}=k,\mathbf{W}_i,\hat{\mathbf{\Xi}}) \\
&= \frac{\int_{\mathbf{v}\in\mathbf{A}} f(\mathbf{v}|C_i=k,\mathbf{W}_i,\hat{\mathbf{\Xi}})Pr(C_i=k|\mathbf{W}_i,\hat{\mathbf{\Xi}})\,d\mathbf{v}}{Pr(C^*_{i_h}=k|\mathbf{W}_i,\hat{\mathbf{\Xi}})} \\
&\overset{a.s.}{\rightarrow} \frac{\int_{\mathbf{v}\in\mathbf{A}_y} f(\mathbf{v}|C_i=k,\mathbf{W}_i,\mathbf{\Xi})Pr(C_i=k|\mathbf{W}_i,\mathbf{\Xi})\,d\mathbf{v}}{Pr(C_i=k|\mathbf{W}_i,\mathbf{\Xi})} \\
&= Pr(\mathbf{y}_i\in\mathbf{A}|\mathbf{W}_i,\mathbf{\Xi},C_i=k),
\end{aligned}
$$

according to (i)-(v), $0 < Pr(C^*_{i_h}=k|\mathbf{W}_i,\mathbf{\Xi}) < 1$, and that f is tight.

The three main results follow from (i)-(vi) as shown below.

First, applying the Glivenko-Cantelli theorem based on (v) and (vi), one has

$$
\sup_{\omega}|\tilde{G}_{kh}(\omega)-\hat{G}_k(\omega)| \overset{a.s.}{\rightarrow} 0,
$$

for $h=1,\cdots,H$.Stretching this result further according to (i), one obtains

$$
\sup_{\omega}|\tilde{G}_{kh}(\omega)-G_k(\omega)| \le \sup_{\omega}|\tilde{G}_{kh}(\omega)-\hat{G}_k(\omega)| + \sup_{\omega}|\hat{G}_k(\omega)-G_k(\omega)| \overset{a.s.}{\rightarrow} 0, \tag{15}
$$

which immediately implies (8).

Based on Bahadur's (1966) representation for $\tilde{\xi}_{kh}$, we write

$$
\begin{aligned}
& \tilde{\xi}_{kh}-\xi_k \\
&= (\tilde{\xi}_{kh}-\hat{\xi}_k)+(\hat{\xi}_k-\xi_k) \\
&= \frac{p-\tilde{G}_{kh}(\hat{\xi}_k)}{g_k(\hat{\xi}_k)} + o_p(n^{-\frac{1}{2}}) + (\hat{\xi}_k-\xi_k)
\end{aligned}
$$

Then due to the continuity of $\hat{\xi}_k$, we can find an ellipsoid $\mathcal{E}$ which contains both $\hat{\xi}_k$ and ξ_k and $\tilde{G}_{kh}(\xi_0) = \tilde{G}_{kh}(\hat{\xi}_k) = \tilde{G}_{kh}(\xi_k)$ for every $\xi_0 \in \mathcal{E}$. Thus

$$\begin{aligned}
&\tilde{\xi}_{kh} - \xi_k \\
= \;& \frac{p - \tilde{G}_{kh}(\xi_0)}{g_k(\xi_k)} + (\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi})' \left[\frac{-\{p - \tilde{G}_{kh}(\xi_0)\} \frac{\partial g_k(\xi_k)}{\partial \boldsymbol{\Xi}}}{(g_k(\xi_k))^2} + \frac{\partial \xi_k}{\partial \boldsymbol{\Xi}} \right] + o_p(n^{-\frac{1}{2}}) \\
= \;& (\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi})' \frac{\partial \xi_k}{\partial \boldsymbol{\Xi}} + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

The last equality holds true since $p - \tilde{G}_{kh}(\xi_0) = O(n^{-1})$ according to (15), and $(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}) = O_p(n^{-\frac{1}{2}})$. Therefore, (9) is concluded.

Lastly, result (10) follows (8) and the uniform integrability of τ.

References

[1] Atkinson, A. and Riani, M. (2000). Robust Diagnostic Regression Analysis. New York: Springer.

[2] Beckman, R. J., Nachtsheim, C. J., and Cook, R. D. (1987) Diagnostics for mixed model analysis of variance. *Technometrics*, **29**, 412-426.

[3] Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., and Rathoutz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, **92**, 1375-1386.

[4] R. R. Bahadur R. R. (1966). A Note on Quantiles in Large Samples. *Annals of Mathematical Statistics*, **37**, 577-580.

[5] Bollen, K.A. (1989). Structural equations with latent variables. New York: John Wiley.

[6] Browne, M.W., and Arminger, G. (1995). Specification and estimation of mean-and-covariance structure models. In G. Arminger, C.C.

[7] Bradlow, E. T. and Zaslavsky, A. M. (1997). Case influence analysis in Bayesian inference. *J. Compute. Graph. Statist.*, **6**, 1-18.

[8] Bryk, A. S. and Raudenbush, S. W. (1992). Hierarchical linear models: applications and data analysis methods. Sage Publications. International Educational and Professional Publisher. Newsbury Park, CA.

[9] Carlin, J. B., Wolfe, R., Brown, C. H., Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics*, **2**, 397-416.

[10] Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983). Graphical Methods for Data Analysis. Wadsworth, Belmont, California.

[11] Cook, R. D. and Weisberg, S. (1982). Residuals and Influence in Regression. New York: Chapman and Hall.

[12] Cook R. D. (1998). Regression graphics: ideals for studying regressions through graphics. New York: John & Sons Wiley.

[13] Data Analysis Products Division of Mathsoft (1999). Splus 2000 Programmer's Guide. Mathsoft Inc..

[14] Dempster, A. P. and Ryan, L. M. (1985). Weighted normal plots. *Journal of the American Statistical Association*, **80** 845-850.

[15] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). Bayesian Data Analysis. Boca Raton: C R C Press LLC.

[16] Gruttola, V. D., Ware, J. H., and Louis, T. A. (1987). Influence Analysis of Generalized Least Squares Estimates. *Journal of the American Statistical Association*, **82**, 911-917.

[17] Goldstein, H. (1995). Multilevel Statistical Models, 2nd edn. London: Arnold.

[18] Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1982). Understanding Robust and Exploratory Data Analysis. New York: John Wiley & Sons.

[19] Hodge, J. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. *J. R. Statist. Soc. B*, **60**, 497-536.

[20] Ishwaran, H., James, L. F., and Sun, J. (2001).Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association*, **96**, 1316-1332.

[21] Jones, B. L., Nagin, D. S., and Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods and Research.* **29**, 374-393.

[22] Schwarz, G. (1978). Estimating the dimension of a model. *Annals Statistics*, **6**, 461-464.

[23] Kellam, S. G., Branch, J. D., Agrawal, K. C., and Ensminger, M. E. (1975). Mental health and going to school: The Woodlawn program of assessment, early intervention, and evaluation. Chicago: University of Chicago Press.

[24] Laird, M. N. and Ware, J. H (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.

[25] Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *The Annals of Statistics*, **17**, 624-642.

[26] Lindsay, G. L. and Roeder, K. (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association*, **87**, 985-794.

[27] Little, R. J. A. and Rubin, D. B. (1987). Statistical Analysis with Missing Data. J. Wiley & Sons, New York.

[28] Lo, Y., Mendell, N., and Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, **88**, 767-778.

[29] Longford, N. T. (1993). Random coefficient models. New York: Oxford University Press.

[30] Longford, I. H. and Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society, A*, **161**, 121-160.

[31] Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, **15**, 661-676.

[32] McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Model 2nd Ed. New York : Chapman and Hall.

[33] Muthén, B. O. (1983). Latent variable structure equation modeling with categorical data. *Journal of Econometrics Psychological Methods*, **22**, 48-65.

[34] Muthén, B. O. (1984). A general structure equation model with dichomotous, ordered categorical and continuous latent variable indicators. *Psychometrika*, **49**, 115-132.

[35] Muthén, B. O. (1996). Growth modeling with binary responses. In A.V. Eye and C. Clogg, *Categorical variables in developmental research: Methods of analysis* (pp. 37-54). San Diego CA: Academic Press.

[36] Muthén, B. O. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery, *Sociological Methodology* (pp. 453-480). Boston: Blackwell Publishers.

[37] Muthén, B. O. and Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, **2**, 371-402.

[38] Muthén, B. O. and Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, **2**, 371-402.

[39] Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C. C., Wang, C.-P., Kellam, S. G., Carlin, J. B., and Liao, J. (2001). General growth mixture modeling for randomized preventive interventions. To be appeared in Biostatistics.

[40] Muthén, L. K. and Muthén, B. O. (1999). Mplus user's guide. Muthén and Muthén Publisher.

[41] Muthén, B. O. and Shedden, K.(1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, **55**, 463-469.

[42] Nagin, D. S. (1993). Analyzing developmental trajectories: a semiparametric, grouped-based approach (1999). *Psychological Methods*, **4**, 139-157.

[43] Redner R. A. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Annals of Statistics.*, **9**, 225-228.

[44] Rubin, B. R., and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samles with ignorable nonresponse. *Journal of the American Statistical Association*, **81**, 366-374.

[45] Schafer, J., L. (1997). *Analysis of incomplete multivariate data.* Chapman and Hall, New York.

[46] Wald R. A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics.*, **9**, 595-601.

[47] Waternaux, C., Laird N. M., and Ware, J. H. (1989). Methods for analysis of longitudinal data: blood lead concentrations and cognitive development. *Journal of the American Statistical Association*, **84**, 33-41.

[48] Weiss, R. E. and Lazaro, C. G. (1992). Residual plots for repeated measures. *Statistics in Medicine*, **11**, 115-124.

[49] Weiss, R. E. (1995). Residuals and outliers in repeated measures random effects models. *Technical Report*, University of California at Los Angeles, Department of Biostatistics.

[50] Weisberg, S. (1985). Applied Linear Regression, 2nd.. New York : John Wiley & Sons.

Table 1: GGMM parameter values that generate independent 100 simulated datasets, and parameter estimates and standard errors fitting for the simulated data (averaged over 100 simulations) under the correct GGMM and a GGMM missing the linear growth factors (No-slope)

Class 1

Model	True	Correct	No-slope
$\beta_{1,1}$	0.000	0.005 (0.070)	0.401 (0.113)
$\beta_{1,2}$	0.300	0.302 (0.019)	0.000 (0.000)
$\phi_{1,11}$	0.300	0.291 (0.069)	0.288 (0.060)
$\phi_{1,22}$	0.150	0.147 (0.019)	0.000 (0.000)
$\theta_{1,11}$	0.700	0.720 (0.074)	1.064 (0.123)
$\theta_{1,22}$	0.700	0.711 (0.073)	0.682 (0.087)
$\theta_{1,33}$	0.700	0.694 (0.069)	0.652 (0.084)
$\theta_{1,44}$	0.700	0.717 (0.077)	1.269 (0.156)
$\theta_{1,55}$	0.700	0.702 (0.118)	2.094 (0.255)
p_1	0.500	0.500 (0.048)	0.460 (0.059)

Class 2

Model	True	Correct	No-slope
$\beta_{2,1}$	2.000	2.006 (0.035)	3.880 (0.121)
$\beta_{2,2}$	1.000	0.997 (0.010)	0.000 (0.000)
$\phi_{2,11}$	0.150	0.151 (0.033)	0.350 (0.059)
$\phi_{2,22}$	0.075	0.073 (0.010)	0.000 (0.000)
$\theta_{2,11}$	0.350	0.352 (0.047)	5.029 (0.247)
$\theta_{2,22}$	0.350	0.351 (0.036)	1.730 (0.131)
$\theta_{2,33}$	0.350	0.352 (0.033)	0.153 (0.043)
$\theta_{2,44}$	0.350	0.348 (0.038)	1.730 (0.131)
$\theta_{2,55}$	0.350	0.351 (0.042)	4.915 (0.278)
p_2	0.500	0.500 (0.048)	0.540 (0.059)

Table 2: GGMM parameter values that generate independent 100 simulated datasets, and parameter estimates and standard errors fitting for the simulated data (averaged over 100 simulations) under three GGMMs with misspecified covariances

Class 1

Model	True	Level 1 *Homogeneous*	Level 2 *Homogeneous*	Two-level *Homogeneous*
$\beta_{1,1}$	0.000	0.021 (0.059)	-0.039 (0.054)	-0.059 (0.051)
$\beta_{1,2}$	0.300	0.305 (0.030)	0.285 (0.028)	0.279 (0.027)
$\phi_{1,11}$	0.300	0.395 (0.071)	0.196 (0.032)	0.210 (0.035)
$\phi_{1,22}$	0.150	0.155 (0.018)	0.104 (0.010)	0.108 (0.011)
$\theta_{1,11}$	0.700	0.504 (0.044)	0.746 (0.084)	0.529 (0.046)
$\theta_{1,22}$	0.700	0.521 (0.039)	0.729 (0.075)	0.525 (0.040)
$\theta_{1,33}$	0.700	0.524 (0.040)	0.699 (0.077)	0.522 (0.040)
$\theta_{1,44}$	0.700	0.535 (0.047)	0.719 (0.092)	0.531 (0.048)
$\theta_{1,55}$	0.700	0.517(0.065)	0.789 (0.013)	0.528 (0.067)
p_1	0.500	0.506 (0.049)	0.481 (0.046)	0.471 (0.047)

Class 2

Model	True	Level 1 *Homogeneous*	Level 2 *Homogeneous*	Two-level *Homogeneous*
$\beta_{2,1}$	2.000	2.012 (0.043)	1.972 (0.042)	1.954 (0.048)
$\beta_{2,2}$	1.000	1.003 (0.023)	0.987 (0.024)	0.984 (0.026)
$\phi_{2,11}$	0.150	0.082 (0.034)	0.196 (0.032)	0.210 (0.035)
$\phi_{2,22}$	0.075	0.070 (0.011)	0.104 (0.010)	0.108 (0.011)
$\theta_{2,11}$	0.350	0.504 (0.044)	0.335 (0.047)	0.529 (0.046)
$\theta_{2,22}$	0.350	0.521 (0.039)	0.343 (0.043)	0.525 (0.040)
$\theta_{2,33}$	0.350	0.524 (0.040)	0.357 (0.037)	0.522 (0.040)
$\theta_{2,44}$	0.350	0.535 (0.047)	0.360 (0.045)	0.531 (0.048)
$\theta_{2,55}$	0.350	0.517 (0.065)	0.335 (0.066)	0.528 (0.067)
p_2	0.500	0.494 (0.049)	0.529 (0.046)	0.529 (0.047)

Table 3: GGMM parameter values that generate independent 100 simulated datasets, and parameter estimates and standard errors (averaged over 100 simulation) fitting for the simulated data under a GGMM with inadequate growth classes

Model	True		1-Class
Class= k	$k=1$	$k=2$	$k=1$
$\beta_{k,1}$	0.000	2.000	1.000 (0.058)
$\beta_{k,2}$	0.300	1.000	0.650 (0.025)
$\phi_{k,11}$	0.300	0.150	1.448 (0.108)
$\phi_{k,22}$	0.150	0.075	0.276 (0.020)
$\theta_{k,11}$	0.700	0.350	0.395 (0.055)
$\theta_{k,22}$	0.700	0.350	0.534 (0.044)
$\theta_{k,33}$	0.700	0.350	0.534 (0.041)
$\theta_{k,44}$	0.700	0.350	0.539 (0.051)
$\theta_{k,55}$	0.700	0.350	0.465 (0.075)

Table 4: Model restrictions for the three selected 3-class GGMMs fitting for the Baltimore Prevention Research data

Model 1	Model 2	Model 3	Remark
$\alpha_{k,2} = \alpha_{k,3} = \alpha_{k,4} = 0$	$\alpha_{k,2} = \alpha_{k,3} = \alpha_{k,4} = 0$	$\alpha_{k,2} = \alpha_{k,3} = \alpha_{k,4} = 0$	$k = 1, 2, 3$
$\theta_{1,uu} = \theta_{2,uu} = \theta_{3,uu}$ $\theta_{1,uu'} = \theta_{2,uu'} = \theta_{3,uu'}$	$\theta_{1,uu} = \theta_{2,uu}$ $\theta_{1,uu'} = \theta_{2,uu'} = \theta_{3,uu'}$	$-$ $\theta_{1,uu'} = \theta_{2,uu'}$	$u = 1, \cdots, 9$ $(u, u') = (1, 2)$ or $(3, 4)$
$\mathbf{\Psi}_1 = \mathbf{\Psi}_2 = \mathbf{\Psi}_3$	$\mathbf{\Psi}_1 = \mathbf{\Psi}_2$	$\psi_{1,22} = 0$	$-$

Table 5: Mplus estimates for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $Pr(C_i|\boldsymbol{\gamma})$ and their standard errors under the three selected 3-class GGMMs fitting for the Baltimore Prevention Research data

Parameters	Model 1	Model 2	Model 3
$\beta_{1,1}$	4.438 (0.190)	4.275 (0.420)	3.242 (0.341)
$\beta_{1,2}$	0.148 (0.206)	0.288 (0.405)	0.438 (0.245)
$\alpha_{1,1}$	-0.058 (0.033)	-0.072 (0.046)	-0.066 (0.031)
$\alpha_{1,2}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{1,3}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{1,4}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{1,5}$	-0.567 (0.229)	-0.411 (0.247)	-0.212 (0.384)
$\alpha_{1,6}$	0.092 (0.038)	0.057 (0.040)	0.019 (0.079)
$\boldsymbol{\gamma}_1$	-2.188 (0.449)	-0.969 (0.325)	-0.391 (5.300)
$Pr(C_i = 1)$	0.143 (0.014)	0.135 (0.065)	0.308 (0.065)
$\beta_{2,1}$	1.906 (0.208)	2.386 (0.127)	2.256 (0.470)
$\beta_{2,2}$	0.741 (0.185)	0.190 (0.123)	0.067 (0.195)
$\alpha_{2,1}$	-0.058 (0.028)	-0.018 (0.016)	-0.003 (0.023)
$\alpha_{2,2}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{2,3}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{2,4}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{2,5}$	0.573 (0.264)	-0.070 (0.135)	-0.121 (0.154)
$\alpha_{2,6}$	-0.154 (0.045)	0.008 (0.022)	0.022 (0.028)
$\boldsymbol{\gamma}_2$	-1.679 (0.417)	0.361 (0.263)	0.135 (3.231)
$Pr(C_i = 2)$	0.114 (0.014)	0.510 (0.091)	0.383 (0.081)
$\beta_{3,1}$	1.991 (0.070)	1.477 (0.063)	1.458 (0.065)
$\beta_{3,2}$	-0.003 (0.064)	0.047 (0.092)	-0.006 (0.065)
$\alpha_{3,1}$	-0.002 (0.010)	-0.007 (0.012)	-0.003 (0.009)
$\alpha_{3,2}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{3,3}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{3,4}$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\alpha_{3,5}$	0.013 (0.085)	-0.039 (0.104)	0.022 (0.068)
$\alpha_{3,6}$	-0.002 (0.010)	0.009 (0.014)	-0.003 (0.009)
γ_3	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$Pr(C_i = 3)$	0.743 (0.104)	0.355 (0.128)	0.309 (0.092)

Table 6: Mplus estimates for $\boldsymbol{\Psi}$ & $\boldsymbol{\Theta}$ and their standard errors under the three selected GGMMs fitting for the Baltimore Prevention Research data

Parameters	Model 1	Model 2	Model 3
$\psi_{1,11}$	0.248 (0.048)	0.079 (0.053)	0.245 (0.080)
$\psi_{1,22}$	0.004 (0.002)	0.010 (0.005)	0.000 (0.000)
$\theta_{1,11}$	0.473 (0.074)	0.941 (0.279)	1.584 (0.601)
$\theta_{1,22}$	0.381 (0.061)	0.597 (0.125)	0.840 (0.609)
$\theta_{1,33}$	0.499 (0.070)	0.593 (0.088)	1.161 (0.302)
$\theta_{1,44}$	0.604 (0.086)	0.722 (0.102)	0.998 (0.187)
$\theta_{1,55}$	0.902 (0.144)	1.277 (0.231)	1.274 (0.523)
$\theta_{1,66}$	0.642 (0.092)	0.942 (0.152)	0.193 (0.197)
$\theta_{1,77}$	0.506 (0.081)	0.773 (0.140)	1.219 (0.362)
$\theta_{1,88}$	0.831 (0.116)	1.128 (0.147)	1.493 (0.409)
$\theta_{1,99}$	0.621 (0.105)	0.932 (0.199)	1.153 (0.580)
$\theta_{1,12}$	0.049 (0.052)	0.136 (0.031)	0.229 (0.139)
$\theta_{1,34}$	0.258 (0.063)	0.183 (0.045)	0.388 (0.080)
$\psi_{2,11}$	0.248 (0.048)	0.079 (0.053)	0.003 (0.468)
$\psi_{2,22}$	0.004 (0.002)	0.010 (0.005)	0.016 (0.044)
$\theta_{2,11}$	0.473 (0.074)	0.941 (0.279)	0.818 (1.000)
$\theta_{2,22}$	0.381 (0.061)	0.597 (0.125)	0.574 (0.466)
$\theta_{2,33}$	0.499 (0.070)	0.593 (0.088)	0.570 (0.145)
$\theta_{2,44}$	0.604 (0.086)	0.722 (0.102)	0.535 (0.115)
$\theta_{2,55}$	0.902 (0.144)	1.277 (0.231)	1.028 (0.532)
$\theta_{2,66}$	0.642 (0.092)	0.942 (0.152)	0.567 (0.168)
$\theta_{2,77}$	0.506 (0.081)	0.773 (0.140)	0.353 (0.120)
$\theta_{2,88}$	0.831 (0.116)	1.128 (0.147)	0.964 (0.254)
$\theta_{2,99}$	0.621 (0.105)	0.932 (0.199)	0.820 (0.319)
$\theta_{2,12}$	0.049 (0.052)	0.136 (0.031)	0.229 (0.139)
$\theta_{2,34}$	0.258 (0.063)	0.183 (0.045)	0.388 (0.080)
$\psi_{3,11}$	0.248 (0.048)	0.004 (0.027)	0.005 (0.018)
$\psi_{3,22}$	0.004 (0.002)	0.005 (0.002)	0.001 (0.001)
$\theta_{3,11}$	0.473 (0.074)	0.161 (0.032)	0.149 (0.033)
$\theta_{3,22}$	0.381 (0.061)	0.190 (0.032)	0.173 (0.028)
$\theta_{3,33}$	0.499 (0.070)	0.239 (0.046)	0.192 (0.053)
$\theta_{3,44}$	0.604 (0.086)	0.247 (0.050)	0.209 (0.049)
$\theta_{3,55}$	0.902 (0.144)	0.197 (0.067)	0.170 (0.054)
$\theta_{3,66}$	0.642 (0.092)	0.320 (0.154)	0.267 (0.085)
$\theta_{3,77}$	0.506 (0.081)	0.210 (0.088)	0.226 (0.109)
$\theta_{3,88}$	0.831 (0.116)	0.154 (0.063)	0.124 (0.101)
$\theta_{3,99}$	0.621 (0.105)	0.104 (0.047)	0.077 (0.056)
$\theta_{3,12}$	0.049 (0.052)	0.136 (0.031)	0.119 (0.029)
$\theta_{3,34}$	0.258 (0.063)	0.183 (0.045)	0.136 (0.053)