First published online January 12, 2021

Published in Volume 30, Issue 4 of *Statistical Methods in Medical Research* on April 1, 2021

Improving Convergence in Growth Mixture Models Without Covariance Structure Constraints

Daniel McNeish[1] and Jeffrey Harring[2]

Abstract

Growth mixture models (GMMs) are a popular method to uncover heterogeneity in growth trajectories. Harnessing the power of GMMs in applications is difficult given the prevalence of nonconvergence when fitting GMMs to empirical data. GMMs are rooted in the random effect tradition and nonconvergence often leads researchers to modify their intended model with constraints in the random effect covariance structure to facilitate estimation. While practical, doing so has been shown to adversely affect parameter estimates, class assignment, and class enumeration. Instead, we advocate specifying the models with a marginal approach to prevent the widespread practice of sacrificing class-specific covariance structures to appease nonconvergence. A simulation is provided to show the importance of modeling class-specific covariance structures and builds off existing literature showing that applying constraints to the covariance leads to poor performance. These results suggest that retaining class-specific covariance structures should be a top priority and that marginal models like covariance pattern GMMs that model the covariance structure without random effects are well-suited for such a purpose, particularly with modest sample sizes and attrition commonly found in applications. An application to PTSD data with such characteristics is provided to demonstrate (a) convergence difficulties with random effect models, (b) how covariance structure constraints improve convergence but to the detriment of performance, and (c) how covariance pattern GMMs may provide a path forward that improves convergence without forfeiting class-specific covariance structures.

[1]Department of Psychology, Arizona State University, Tempe, USA
[2]Department of Human Development & Quantitative Methodology, University of Maryland, College Park, USA

Corresponding Author:
Daniel McNeish, Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ, USA, 85287.
Email: dmcneish@asu.edu

1 Introduction

Growth models are a common group of statistical methods applied to repeated measures data when the interest is in quantifying how the mean of an outcome changes over time.[1–3] One way to quantify this heterogeneity is to identify unobserved, latent classes of growth trajectories from the data.[4] The goal is similar to including a moderator for growth like sex or treatment condition to allow different growth trajectories for different types of people. However, the moderator in this case is latent and not known a priori. Discrete latent classes of growth trajectories are determined from the data by combining latent class analysis with growth modeling in what have been deemed *growth mixture models* (GMMs).[5,6]

Methodological and empirical literatures alike lament difficulties in estimating GMMs and applications can be fraught with nonconvergence. Sample size requirements for obtaining trustworthy estimates can exceed 1,000 with routine data characteristics.[7] Nonetheless, GMMs outperform alternative methods for clustering growth trajectories despite the possible reservations about estimation.[8] Criticisms notwithstanding, researchers continue to regularly employ GMMs to address questions about growth trajectory heterogeneity while attempting to parry nonconvergence difficulties along the way.

When nonconvergence arises with GMMs, a common practice is to impose equality constraints on the covariance structure across classes[9] or to switch to a more restrictive latent class growth model (LCGM).[10] Parameters in the covariance structure are the most difficult to estimate, so the logic is that nonconvergence can be avoided by simplifying this portion of the model. These methods sacrifice flexibility in the covariance structure to facilitate estimation and, though effective for reducing nonconvergence, they have been noted to change key conclusions of the model such as growth trajectories in each class,[11,12] how many classes are extracted,[13,14] the meaning of the classes,[15,16] or class assignment.[17] So whereas covariance structure misspecification is less injurious in non-mixture settings where it has no impact on consistency of regression coefficients,[18] adding mixture components results in covariance structure misspecifications permeating to all parts of the model.

Though GMMs typically follow the random effects tradition,[6] the marginal model tradition may have untapped utility in mixture contexts. Covariance pattern growth mixture models (CPGMM)[19,20] were recently developed as an alternative to covariance structure constraints to keep the covariance structure as flexible as possible while minimizing nonconvergence. Marginal models like covariance pattern models are less computationally demanding than random effect models for repeated measures data because they do not partition the covariance into between-person and within-person sources with random effects.[21] Furthermore, covariance pattern models are capable of providing the same marginal covariance structure as random effects models. However, marginal models have received little attention in the GMM literature, even though GMM applications focus on marginal quantities like average growth trajectories in each class and rarely have person-specific interests that would necessitate inclusion of random effects.[19,22–24]

Forgoing the random effects tradition in favor of a covariance pattern approach could promote flexible and class-specific covariance structures, potentially improving accuracy of parameter estimates and class assignment by reducing misspecifications induced via constraints imposed primarily to placate nonconvergence concerns. Proof-of-concept simulations with fairly idealistic conditions have been conducted with CPGMMs to show their potential utility.[19] The primary goal of this paper is to build from these initial simulations to explore how CPGMMs perform with messier and more realistic data that include features like attrition, modest sample sizes, and poor class separation. Ultimately, the goal is to determine if CPGMMs outperform popular covariance constraint approaches for circumventing nonconvergence with the data characteristics typically seen in applications. Specifically, we want to address the potential fallibility of prioritizing convergence over flexibly modeling the covariance structure and whether CPGMMs mitigate nonconvergence while maintaining class-specific covariance structures, thereby preempting researchers from having to choose between convergence and the covariance structure in the first place.

To outline the structure of this paper, Section 2 overviews the basics of GMMs and suggested alternatives when convergence problems are encountered. Section 3 reviews GMM applications in the PTSD literature to note the sample size, growth trajectories, and missing data characteristics of these studies. Section 4 presents a simulation study to compare the convergence, estimated class proportions, and parameter estimation accuracy of various models used for study growth trajectory heterogeneity. Section 5 presents the results of simulation. Section 6 provides an application to a PTSD study to demonstrate interpretational differences between approaches that sacrifice the covariance structure for convergence and CPGMMs. Section 7 provides discussion points and limitations.

2 Growth Mixture Models

Conceptually, GMMs can be thought of as an extension of growth models[25,26] with a discrete latent moderator.[27] In essence, this means that the model is a multiple-group growth model where the grouping variable is latent.[28] So whereas groups are split based on a known variable in a multiple group growth model, GMMs probabilistically uncover the groups into which the data are split.[29(p138)]

The most general unconstrained growth mixture model (GMMU) is based on a random effect growth model and many can be written as,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{T}_i \boldsymbol{\beta}_{ik} + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\beta}_{ik} &= \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \mathbf{x}_i + \mathbf{b}_i \end{aligned} . \tag{1}$$

In the first expression, $\mathbf{y}_i$ is a $t_i \times 1$ vector of responses where t_i is the number of observed repeated measures provided by person *i*, $\mathbf{T}_i$ is a $t_i \times q$ design matrix for *q*, the number of coefficients related to time, $\boldsymbol{\beta}_{ik}$ is a $q \times 1$ vector of growth coefficients for individual *i* in class *k* where $k = 1, ..., K$ for *K*, the preselected number of classes, and $\boldsymbol{\varepsilon}_i$ is a $t_i \times 1$ vector of multivariate normal residuals with a zero mean vector and class-specific covariance matrix $\boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \mathbf{R}_{ik})$. In the second expression, the growth coefficients $\boldsymbol{\beta}_{ik}$ are modeled by a $q \times 1$ vector of class-specific fixed effects $\boldsymbol{\alpha}_k$, a $q \times p$ matrix of class-specific coefficients $\boldsymbol{\Gamma}_k$ for *p* time-invariant covariates, a $p \times 1$ vector of time-invariant covariate values $\mathbf{x}_i$, and a $q \times 1$ vector of multivariate normal random effects with a zero mean vector and class-specific covariance matrix, $\mathbf{b}_i \sim MVN(\mathbf{0}, \mathbf{G}_k)$. This implies that every class will have its own fixed effects, random effect covariance matrix, residual covariance matrix, and time-invariant covariate coefficients. Rather than person-specific trajectories varying around a single marginal growth trajectory as in a traditional random effect growth model, person-specific growth trajectories vary around a class-specific marginal growth trajectory defined by the class-specific fixed effects, $\boldsymbol{\alpha}_k$.

The model-implied mean and covariance structures for each class are,

$$\boldsymbol{\mu}_{ik} = \mathbf{T}_i (\boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \bar{\mathbf{x}}) \tag{2}$$

$$\boldsymbol{\Sigma}_{ik} = \mathbf{T}_i (\boldsymbol{\Gamma}_k \boldsymbol{\Phi} \boldsymbol{\Gamma}_k^{\mathrm{T}} + \mathbf{G}_k) \mathbf{T}_i^{\mathrm{T}} + \mathbf{R}_{ik} \tag{3}$$

where $\boldsymbol{\Phi}$ is the covariance matrix of the time-invariant covariates. The model-implied mean and covariance structures for the full model cannot be expressed a priori because it depends on the latent class proportions, which serve as weights for each of the class-specific models. The general form of the composite density of the full model for $\mathbf{y}_i$ can be written as

$f\left(\mathbf{y}_i \mid \boldsymbol{\varphi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \sum_{k=1}^{K}\left[\varphi_k f_k\left(\mathbf{y}_i \mid \boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}\right)\right]$ where f_k is the component normal probability density function for the *k*th class, and φ_k is the proportion of people in the *k*th class where $0 \le \varphi_k \le 1$ and $\varphi_K = 1 - \sum_{k=1}^{K-1} \varphi_k$.

2.1 Convergence Difficulties

The model allows for between-person variability (captured by $\mathbf{G}_k$) around the class-specific mean trajectory (defined by $\boldsymbol{\alpha}_k$) and within-person variability around the person-specific growth trajectory (denoted by $\mathbf{R}_{ik}$). This helps to fully partition the variability, but it also means that there are many random effects in the model. In growth models, covariance structure parameters are the most difficult to estimate,[30] so the source of convergence difficulties with GMMUs is not hard to conceive when one considers the number of covariance structure parameters present when each of the *K* classes has unique covariance structure parameters.[31,32]

More specifically, the presence of so many covariance parameters across latent classes creates many singularities in the likelihood surface whereby the likelihood spikes to infinity.[33] If an optimization algorithm encounters one of these singularities, the algorithm will simply fail to converge because the gradient is undefined. A subtler but damaging issue with singularities occurs when the optimization algorithm encounters values near the singularity, which often results in a local maxima that terminates the optimization algorithm but whose solution does not represent the global maximum of the likelihood surface.[34,35] Therefore, when nonconvergence with GMMUs is encountered, the common remedy is to reduce the number of covariance parameters to avoid complications related to singularities in the likelihood surface.[33] The two most common ways to do this are with a Constrained GMM (GMMC) or with an LCGM. We also discuss recently proposed covariance pattern GMMs (CPGMM) as an alternative. These methods are overviewed in Sections 2.2 through 2.4.

2.2 Constrained Growth Mixture Models

Removing the *k* subscripts from $\mathbf{G}$ and $\mathbf{R}_i$ such that $\mathbf{G}_1 = \mathbf{G}_2 = \cdots = \mathbf{G}_K = \mathbf{G}$ and $\mathbf{R}_{i1} = \mathbf{R}_{i2} = \cdots = \mathbf{R}_{iK} = \mathbf{R}_i$ retains the concept behind GMMUs while reducing the number of covariance structure parameters. The logic is that, if the number of parameters composing the covariance structure make estimation difficult, then estimating fewer of these parameters should mollify the problem.[36] The advantage with this approach is that the model retains person-specific trajectories within each class, estimation is easier since there are fewer of the difficult-to-estimate covariance parameters in the model, and the frequency of singularities in the likelihood is greatly reduced.[13,33]

The disadvantage is that the random effect covariance matrix and residual covariance matrix must be equal across all classes. Though commonly implemented in empirical settings to aid convergence (e.g., GMMCs are the default in the M*plus* software),[37] the approach of constraining covariance matrices across classes has been widely criticized. Bauer and Curran[38] explicitly question the choice to apply constraints across classes by stating,

> Although [constraints across classes] are statistically expedient, we do not regard these equality constraints as optimal from a theoretical standpoint, and in our experience, they are rarely found to be tenable in practice. Indeed, implementing these constraints is in some ways inconsistent with the spirit of the analysis, because one is forcing the majority of the parameter estimates to be the same over classes (permitting only mean differences in the within-class trajectories) (p. 346).

The model will attempt to classify individuals while satisfying the within-class growth characteristics defined by the model. So to the extent that the covariance structure differs across classes, enumeration and classification errors will increase.[11,13,17,39,41] Further, these constraints are often applied in response to nonconvergence rather than theory calling for an assumption of covariance homogeneity.[37,40] So although this approach can aid in improving convergence, it has been criticized for being inconsistent with the goal of conducting such an analysis while also having a high probability of adversely impacting the resulting class enumeration and parameter estimates.[35]

2.3 Latent Class Growth Models

LCGMs provide another option to simplify the estimation of GMMUs.[10,42] In the random effect growth modeling framework, each person has a person-specific growth trajectory. Nagin and colleagues note that interpreting *N* separate trajectories can be surfeit and that *N* trajectories can be reduced to a handful of prototypical trajectories. They do not argue that only these prototypical trajectories matter, quite the opposite.[10(p140)] The argument is that the underlying information from a few prototypical trajectories is easier to digest than *N* person-specific trajectories. Instead of fitting a model with continuous random effects, LCGMs do not allow the growth coefficients to vary across people. Instead, discrete latent classes are used rather than continuous random effects to capture between-person differences in growth trajectories. Rather than person-specific growth trajectories, people are assigned to the class whose trajectory formed by class-specific fixed effects most closely match what the person-specific trajectory would have been. The model does not allow for between-person differences within classes and people within a class are considered interchangeable, so any deviation from the class-specific trajectory is absorbed into the residual term.

The LCGM can be written as

$$\begin{aligned}\mathbf{y}_i &= \mathbf{T}_i\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\beta}_k &= \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k\mathbf{x}_i\end{aligned} \tag{5}$$

where $\boldsymbol{\varepsilon}_i$ is a $t_i \times 1$ vector of residuals such that $\boldsymbol{\varepsilon}_i \sim MVN\left(0, \sigma_k^2\mathbf{I}_{t_i}\right)$, meaning that the residuals are assumed to be constant and independent across time. Note that $\boldsymbol{\beta}_k$ does not have an i subscript nor a random effect vector $\mathbf{b}_i$, so the only source of between-person heterogeneity is through the latent classes. The goal of LCGMs is therefore to provide a semiparametric representation of the growth trajectories, meaning that there is no requirement to assume normality around the class-specific mean trajectory.[43] An advantage of the LCGM is that there are no random effects within classes, so computational difficulties are rarely encountered.[44]

Compared to GMMUs, a drawback is that the model features a simple submodel for the covariance among repeated measures.[45,46] This is intentional because the LCGM conceptualizes classes differently from GMMs. LCGMs define a class as a collection of people who follow a similar and distinct trajectory whereas GMMs define a class as a heterogeneous set of people that can be described by a single probability distribution.[43(p895)]. As a result, modeling the residuals as constant and independent over time often leads to additional classes being extracted with a LCGM relative to a corresponding GMM.[14,16] This is not to say that the LCGM is incorrect, but rather the solution represents a different definition of what constitutes a class. Though LCGMs have merit for their own theoretical considerations, they do not always align with the goal of GMMs and can be a tenuous substitution for GMMs because LCGMs often lead to different solutions and interpretations.[47,48]

2.4 Covariance Pattern Mixture Models

CPGMMs are another alternative to reducing the complexity of the GMMUs to improve convergence.[19] Covariance pattern models[49,50] are part of the family of marginal models whose goal is to estimate the marginal growth trajectory while fully modeling the covariance between repeated measures.[51] Random effect growth models have a similar goal but add an intermediate step of providing person-specific growth trajectories, which requires partitioning the marginal covariance into between-person and within-person sources. Covariance pattern models combine all sources into a single marginal covariance in order to describe the covariance between repeated measures rather than trying to explain the sources of covariation among repeated measures with random effects.

The covariance pattern growth mixture model (CPGMM) can be written as,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{T}_i \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\beta}_k &= \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \mathbf{x}_i \end{aligned} \qquad (6)$$

where $\boldsymbol{\varepsilon}_i \sim MVN\left(\mathbf{0}, \boldsymbol{\Sigma}_{ik}\left(\boldsymbol{\theta}_k\right)\right)$. There is no random effect vector in the second expression, so $\boldsymbol{\Sigma}_{ik}$ combines within- and between-person sources of variability in class *k*, whose structure is a function of class-specific parameters in $\boldsymbol{\theta}_k$. The LCGM is a special case of the CPGMM when $\boldsymbol{\theta}_k = \sigma_k^2$ but more complex structures can be used to allow for heterogeneity within classes and to better reproduce the variances and covariances of the repeated measures. The most general structure allows for the variances in each class *k* to be uncoupled from the correlations such that $\boldsymbol{\Sigma}_{ik} = \mathbf{D}_{ik}^{1/2}\mathbf{P}_{ik}\mathbf{D}_{ik}^{1/2}$ where $\mathbf{D}_{ik}$ is a diagonal matrix of time-specific variances in class *k* and $\mathbf{P}_{ik}$ is a class-specific correlation matrix.[52,53] More commonly, parsimonious structures are chosen such as compound symmetry or an autoregressive structure.

The CPGMM shares the advantages of the LCGM in that it removes random effects and the associated covariance partition that complicate estimation of GMMUs. It is more congruent with GMMs than LCGMs by more rigorously modeling the covariance structure of the repeated measures within classes. In essence, the goal is to arrive at the same marginal covariance as the GMMU but to do so without random effects in order to reduce the singularities in the likelihood surface. As a disadvantage, researchers lose the ability to obtain person-specific trajectories and the ability to differentiate between-person and within-person sources of variance. Additionally, the researcher is responsible for selecting the structure of the marginal covariance, which can be more challenging for data with many repeated measures.

Table 1 compares the key features of the four types of GMMs we have covered for modeling heterogeneity in growth trajectories.

3 Convergence Difficulties in Empirical PTSD Studies

To provide context and evidence for the nonconvergence in empirical studies and to gauge data characteristics seen in applications, we reviewed studies in the PTSD literature where GMM applications are frequent. To facilitate this review, we use the work of van de Schoot et al.[54] as a baseline, whose review screened 11,395 papers that satisfied keywords, ultimately whittling down to 34 papers containing 38 unique studies. The original goal of the review was to use these studies to form informative prior distributions for a Bayesian GMM analysis. To address the interest of the current paper, we re-reviewed these studies with a focus on the characteristics of the data and the modeling decisions used to arrive at the final model (characteristics not tracked in the original review). The modeling decisions of these studies were telling for how researchers dealt with nonconvergence. First, only 2 papers (6%) reported using a GMMU with class-specific covariances whereas 9 studies (27%) reported a GMMC. Another 9 studies (27%) did not provide enough information in the reports to determine if there were covariance structure constraints across classes. Presumably, these 9 studies used a GMMC because this is the default setting in the M*plus* software used in these studies. An additional 14 studies (41%) used a LCGM.

The prevalence of model choices that directly reflect nonconvergence issues is less surprising when looking at the attributes of the data in these studies. First, the median attrition from the first wave to the last wave was 35%, with the first and third quartiles being 22% and 51%, respectively (range: 9% to 74%). Missing data are rarely considered in GMM simulations[55] and large amounts of missing data exacerbate nonconvergence.

Second, the sample sizes found in these studies are markedly lower than the suggested values suggested by the methodological literature.[7] The median sample size was 517, with the first and third quartiles being 207 and 835, respectively (range: 70 to 16,488). In the methodological literature on GMMs, 300 is a typical lower bound for sample size in simulations with complete data.[13,56] Though simulation studies have yet to explicitly consider or quantify the impact of simultaneously having high attrition with modest sample sizes, estimation of GMMUs with class-specific covariance matrices in such contexts is unlikely to be auspicious.

The next section provides a simulation study with conditions inspired by these PTSD data to examine the performance and nonconvergence of different modeling approaches. Our data generation model builds off McNeish and Harring[19] and incorporates more realistic data characteristics such as modest samples, attrition, and poor class separation than were used in their proof-of-concept study that features more idealistic characteristics.

4 Simulation Study

4.1 Data Generation Model

The data generation model is based on the so-called “Cat’s Cradle” pattern that emerges in PTSD research.[57,58] Four classes typically emerge: one class that starts at higher values and maintains high values (the “Chronic” class), a second class that starts low and maintains low values (the “Resilient” class), a third class that starts high but decreases over time (the “Recovery” class), and a fourth class that starts low and increases over time (the “Delayed Onset” class). The Resilient group comprised 63% of the population, the Recovery class 12%, the Chronic class 19%, and the Delayed Onset class 6% to mirror empirical applications of GMMs where class proportions are disparate. The data feature 5 time-points that represent weeks after a traumatic incident. Time is coded as: 0 (baseline), 1, 10, 18, and 26, which follows the timing employed in motivating application data we use in Section 6, which measured people’s PTSD approximately every 2 months after 2 closely spaced baseline measurements.

Figure 1 shows a plot of the trajectories in each of the four classes. Table 2 shows the model equations and covariance structures that were used to generate data from these trajectories. Table 2 has two sets of covariance matrices that satisfy the manipulated simulation conditions because we also manipulate class separation (discussed shortly). The growth trajectories present in each class have both linear and quadratic components to account for nonlinearity in the longitudinal profiles present in Figure 1. The linear slope varies across individuals within classes in the data generation model but the quadratic slope variance was constrained to zero because quadratic variance is difficult to estimate due to scale differences[59] and we did not wish to favor models that do not feature random effects based on how we generated the data.

4.2 Simulation Conditions

We manipulated three conditions in our simulation design. First, we generated data from three different sample sizes: 100, 250, and 500. These conditions were selected from the distribution of the sample sizes observed in the review of the PTSD literature; the lower bound was near 100, the 25th percentile was near 250, and the median was near 500. Our focus is on nonconvergence, so our simulation conditions focus on the lower 50% of the sample size distribution where these issues are most probable.

Second, we had three conditions for attrition: 0%, 25%, and 45%. 20% attrition corresponds to the 25th percentile of reported values in the PTSD review and 45% attrition corresponds to the 75th percentile. Attrition was generated such that missing data did not rely on any covariates or the values of the missing data themselves. Instead, missing values were generated according to a random probability. To mimic the way that people drop out of studies, we assigned a probability to each generated person having 5, 4, 3, or 2 observed repeated measures. For the 20% attrition condition, 80% had 5 repeated measures, and 10% each had 3 or 2 repeated measures. In the 45% attrition condition, 55% had 5 repeated measures, 10% had 4 repeated measures, 15% had 3 repeated measures, and 20% had 2 repeated measures. Missing data were monotone.

Third, we manipulated the separation between the different latent classes. Sample size in GMMs is really about the *interaction* between sample size and class separation. If the classes are completely distinguishable, then the model will easily identify the number of classes and estimate the growth trajectory in each, even at very small sample sizes. For example, Verbeke and Molenberghs[60(p186)] provide a 2-class example of height changes in 20 schoolgirls without issues as the two classes are highly separated. On the other hand, samples of 1000 or more may be insufficient for estimation and class assignment when classes are poorly separated.[61,62]

Class separation is straightforward to define with 2-class solutions (e.g., Mahalanobis distance), but it is not straightforward to quantify separation with more than 2 classes as in the current simulation design. For more than two classes, class separation is less clear but relative entropy has been used in past studies by manipulating the population values of $\mathbf{G}_k$ and $\mathbf{R}_{ik}$.[61] Following this precedent, our low class separation condition featured relative entropy of 0.70 for the population model and our high separation condition resulted in relative entropy of 0.90. These values also reflect the PTSD literature review: of the 23 studies that reported relative entropy values for their final model, the 25th percentile was 0.72 and the 75th percentile was 0.93 (range: .49 to .97). To help visualize class separation conditions, Figure 2 shows the width of the 95% interval for the person-specific trajectories for each class in each class separation condition.

Each generated dataset was fit with four different models – (a) a GMMU with class-specific covariance matrices (the data generation model), (b) a GMMC where all covariance matrices are constrained to be equal across classes, (c) an LCGM, and (d) a CPGMM with a class-specific compound symmetric marginal covariance matrix. We used a compound symmetric marginal covariance structure because it would represent a moderate misspecification given that the population model includes random slopes. The misspecification was intentional to encode the difficulty of perfectly specifying the covariance structure with real data so that the results would more accurately reflect the performance of different approaches in applications. The mean structure was the same for each model and was properly specified. Because the population values for the quadratic slope variance were 0, the GMMU and GMMC conditions did not include a quadratic slope random effect. The covariance between the intercept and linear slope random effect was estimated in the GMMU and GMMC conditions. All conditions were fully crossed, meaning there were 18 possible combinations of sample size, missing data, and class separation. Within each of these possible condition combinations, 500 datasets were generated and analyzed with all 4 models with M*plus* Version 8.3 using robust maximum likelihood estimation. All data generation and analysis files can be found on the first author's Open Science Framework page (https://osf.io/k9y54/).

4.3 Simulation Outcomes

The first outcome is the percentage of the 500 replications that converged for each of the 4 models. The complexity of estimation with GMMs means that "nonconvergence" can be interpreted in slightly different ways. We consider a replication nonconvergent if the solution produces inadmissible estimates (e.g., negative variances, correlations greater than 1), the Hessian matrix was nonpositive definite, or if likelihood optimization could not be completed using M*plus*'s default criteria (500 EM algorithm iterations, convergence criteria of 1E-5). If a replication produced a solution with admissible solutions with one set of starting values but the same likelihood could not be reproduced for a different set of starting values, we counted the replication as convergent for the purposes of the simulation even though it may suggest a local maxima or saddle point. This employs a relatively lenient definition such that replications are counted as converged anytime optimization completed with admissible parameter estimates. The second outcome is the mean absolute bias of the class proportions, which will provide a single value to summarize how well the model is able to estimate the proportion of people in each class. This is calculated by averaging the absolute value of the difference between all of the class proportions from the data generation model and the estimated class proportions based on the most likely class membership assignment, $K^{-1}\sum_{k=1}^{K}\left|\hat{\varphi}_k - \varphi_k\right|$.

The third outcome is whether each model can identify the growth trajectories of the four different classes generated by the population model. This will be explored graphically rather than being quantified by distance from the population parameter value because we are more interested in ability of the model to qualitatively identify the different types of growth trajectories.

5 Simulation Results

5.1 Convergence

Table 3 shows the convergence results across simulation conditions. GMMU results highlight a familiar theme – even though the GMMU is the data generation model, under relatively advantageous circumstances of no attrition, well-separated classes, and a sample size of 500; convergence was only 50%. As conditions deviate further from ideal and feature smaller samples and attrition, GMMU became increasingly worse and ultimately converged in only 1% of replications for the condition with 100 people and 45% attrition. These results reflect the nonconvergence reported in the PTSD literature review, suggesting that nonconvergence would be rather dire even if researchers were using the exact data generation model.

GMMCs are often used to improve convergence when GMMUs fail. GMMC convergence rates were not 100%, but convergence notably improved and corroborated the popularity of this strategy, especially with smaller sample sizes and higher attrition. As it relates to convergence, the LCGM is also inviting with convergence at or near 100% across conditions. The CPGMM convergence rates exceeded GMMC and were only slightly lower than the LCGM, demonstrating the advantage that CPGMMs can provide class-specific covariance parameters without the computational issues of GMMUs.

Nonconvergence rarely was related to optimization issues like exceeding the number of iterations and these issues occurred in 1% or less of replications for each condition. Nearly all problems were attributable optimization terminating normally but at a solution with inadmissible values or the Hessian matrix being nonpositive definite when evaluated at the parameter estimates. The central issue estimating these models is not optimization choices like number of iterations or convergence criteria but rather that the likelihood surface is difficult to navigate and poorly behaved with singularities or multiple optima.

The ability of a model to merely converge to an admissible solution is not the ultimate goal and other relevant performance metrics need to be considered to ensure that the extracted classes correspond to the population model. The next two subsections discuss results related to estimated class proportions and estimated class trajectories.

5.2 Class Proportions

Table 4 shows mean absolute bias for each model in each of the simulation conditions. With this metric, 0 indicates that the estimated class proportions match the generated class proportions so lower values are better with 0 indicating that the class proportions exactly match the data generation model. Results show that in the 100 sample size condition, the CPGMM yielded the class proportions that most closely mirrored the generated class proportions in all conditions, often by a sizeable margin. This trend continued in all cells containing a sample size of 250, with one exception. In the 500 sample size condition, GMMC had the lowest mean absolute bias for class proportions when class separation was low, but the CPGMM generally produced the lowest mean absolute bias for class proportions when class separation was high. Overall, the CPGMM yielded the lowest mean absolute bias for class proportions in 72% of the simulation conditions.

Another notable feature of Table 4 is the consistently high mean absolute bias values across all conditions for the LCGM, suggesting that this model is not classifying people in a way that is consistent with the generated classes. Specifically, the LCGM tended to classify few people into the Resilient class (estimated proportions were about half the population proportions) and overclassified people in the Delayed Onset and Recovery classes. This echoes previous points that the LCGM is not a substitute for a GMMU (as it is sometimes used given the ease with which it converges) because the models have different goals and conceptualize classes differently.

It is also important to note that these values are only calculable for replications that converged, so selection effects are a possible explanation for inaccurate values in some conditions. For instance, class proportions for the GMMU models in the smaller sample size conditions could be inaccurate because the estimation is only stable enough to converge in replications that over-assign people to the smaller, minority classes. If this were the case, then the class proportions in Table 4 could be confounded with convergence.

5.3 Estimated Class Trajectories

Figure 3 compares the average estimated class trajectories across all converged replications (in black) to the generated class trajectories (in grey) for the 250 sample size conditions with 45% attrition. Figure 3 does not include GMMU in the comparison given that GMMU convergence was poor and that one of our interests lies in the best alternative when the desired GMMU fails (for the few GMMU replications that did converge, results closely mirrored the GMMC plots). Results for other simulation conditions follow a similar pattern but are not presented for brevity.

The low separation conditions in the top row of Figure 3 show that GMMC and LCGM do not recover the class trajectories very well. Rather than the “cat’s cradle” pattern of the generated model, both of these models essentially produced four horizontal lines and the “Recovering” and “Delayed Onset” classes are not readily identifiable as embodying the intended characteristics. The CPGMM class trajectories are more closely aligned to the trajectories of the data generation model and the key distinctions between the different classes are easy to identify. The result is amplified in the high separation condition in the bottom row of Figure 3. The CPGMM class trajectories overlap with the trajectories from the data generation model so closely that the lines are nearly indistinguishable. Meanwhile, GMMC and LCGM continue to produce essentially horizontal lines that resemble the trajectories of the data generation model much less faithfully.

5.4 Simulation Results Summary

Though drawing conclusions from simulation results is difficult as interpretations and trends become more convoluted as the myriad of crossed conditions increases, the results seem to suggest classifying the different model type based on the complexity of the estimation and the class-specificity of the covariance structure as is shown in Table 5. Below is a synopsis of this classification.

1. GMMUs with class-specific covariance matrices are often the model that researchers theoretically want to fit but this model may be too complex for data characteristics commonly encountered in PTSD studies. Even when the analysis model matched the data generation model, convergence with GMMUs was below 50% in the best-case scenario and as low as 1% with smaller samples and higher attrition.
2. Alternative methods like GMMCs, LCGMs, and CPGMMs can aid in improving convergence. However, class proportions and estimated class trajectories were less accurate for GMMCs and LCGMs because they place quixotic constraints on the covariance matrices. The LCGM also showed discrepant class proportions, likely because the model conceptualizes classes differently than methods that model for heterogeneity within classes. The CPGMM was more accurate in terms of class proportions and

estimating class trajectories across conditions, even if the covariance structure was not correct.

3. Simulation evidence showed that the CPGMM most effectively balances convergence with performance. The CPGMM narrowly trailed the convergence rate of LCGM and exceeded the convergence rate of GMMC, but the CPGMM outperformed both the LCGM and GMMC in terms of estimating class proportions and class trajectories.

Essentially, CPGMMs provide a more complete covariance structure than LCGMs, improves on GMMCs by allowing class-specific covariance structures, and encounters fewer convergence issues than GMMUs by removing within-class random effects while maintaining the ability to answer the fundamental questions asked in empirical studies (i.e., how many classes are there, what is the average trajectory in each class, what predicts class membership). Together, this suggests that the CPGMM – not LCGM or GMMC – should be the preferred alternative to emulate the intent of GMMUs in the face of convergence issues. To take this a step further, if there are no person-specific research questions, the CPGMM could be the analysis of choice from the onset rather than only providing utility after a more complex model fails to converge. The next section provides an application to show how choice of model can dramatically influence conclusions from GMMs.

6 Application

Data come from a study on PTSD in burn victims who were admitted to a burn center between 1997 and 2000.[63] Participants completed the Impact of Event Scale[64] up to 8 times in the 12 months following the incident. The first two waves were taken two and three weeks after the traumatic incident, respectively. The remaining 6 waves were collected in 8-week intervals thereafter (i.e., time is coded 0, 1, 9, 17, 25, 33, 41, 49). The data contain 301 people and there is 21% attrition.

In the PTSD literature, four latent classes are often found, so we fit a 1-class, 2-class, 3-class, 4-class, and 5-class GMMU with class-specific covariance matrices and estimated with robust maximum likelihood estimation in M*plus* Version 8.3 using 100 random sets of starting values and 10 optimization points to guard against local maxima.[31,65] As anticipated by simulation results, none of the GMMU models with multiple classes converged due to inadmissible parameter estimates. Reverting to a GMMC did not help – constraining the covariance matrices to be equal across all classes continued to yield inadmissible parameter estimates for models with between 2 and 5 classes.

To allay these convergence issues and to deal with the smaller sample size in a principled way, van de Schoot et al.[54] analyzed these data with Bayesian estimation using a Markov chain Monte Carlo algorithm by placing informative priors on the class proportions and covariance matrices based on the 34 aforementioned papers. This is a perfectly reasonable approach methodologically to accommodate issues related to convergence and smaller sample sizes. Nonetheless, a looming issue is that nearly all the studies upon which the informative priors were based employed some type of alteration or constraint to work around nonconvergence. As a result, the covariance matrices in the Bayesian analysis were constrained to be equal across classes and assumed a 4-class solution often found in the literature because it is difficult to place priors on unknown classes. So, whereas informative priors are a commonly cited method to circumvent small sample or convergence issues,[66] prior distributions are only as useful as the underlying source upon which they are based. In this case, the existing literature is plagued with nonconvergence and associated constraints, so basing informative priors on this literature implicitly encodes these issues into the analysis.

Instead, we applied a CPGMM with class-specific factor analytic marginal covariance matrices. A factor analytic structure is similar to a Cholesky factorization of an unstructured matrix but uses fewer parameters and is helpful for data with many repeated measures,[67,68] especially if the timing is uneven such that autoregressive structures are inappropriate. With the GMMU and GMMC, we tested between 1 and 5 classes. The 5-class solution extracted a class with very few people, so we do not consider it here. Additionally, one of the classes in the 4-class solution had nearly equal variances and near equal covariances across time, so we reduced the marginal covariance to a compound symmetric structure for that class only to reduce the number of parameters.

Though there are many methods by which to base class enumeration, we based our decision on the BIC, sample-size adjusted BIC (SA-BIC)[69], integrated classification likelihood Bayesian information criteria (ICL-BIC)[35,70], and classification likelihood information criterion (CLC)[35]. BIC is a commonly reported criteria for class enumeration[71] with some simulations advocating for the adjusted SA-BIC version in the presence of poorer class separation and smaller samples.[61,72] Simulations have shown BIC-ICL and CLC outperform likelihood-based statistics and information criterion by wide margins in when selecting among solutions with $K > 1$, smaller samples, and disparate class proportions.[73] Table 6 reports fit of models with 1 to 4 classes. Notably, all measures support the 3-class solution over the traditional 4-class solution that is commonly obtained when covariance constraints are applied and that has previously reported with this data when the covariance structure is constrained across classes. Because class enumeration was not included in the previous simulation, we perform a small-scale simulation with a small number of conditions to provide some context for these class enumeration results.

6.1 Small-Scale Class Enumeration Simulation

Because the relative entropy was high for all class solutions in Table 6, we generated 500 datasets from the 4-class High Separation condition from the model equations shown in Table 2 with the same class proportions as in the previous simulation (63%, 12%, 19%, 6%). We changed the number of time-points from 5 to 8 to match the PTSD data and coded time as 0, 1, 9, 17, 25, 33, 41, 49. The sample size for each generated data set was 300. Missing data were generated such that 80% of the data were complete, 10% were missing the last three measurement occasions (t = 33, 41, 49), and 10% were missing the last six measurement occasions (t = 9, 17, 25, 33, 41, 49). This was done to mirror the missing data pattern in the empirical data where the percentage of observed data at each measurement occasion was 100%, 98%, 93%, 88%, 84%, 82%, 80%, 80%.

We then fit CPGMMs with 2, 3, 4, and 5 classes to each generated dataset and compared the four fit indices reported in Section 6 (BIC, SA-BIC, ICL-BIC, and CLC). Convergence was defined identically to Section 4.3. The main interest was the proportion of replications in which the correct 4-class solution was selected by different information criteria. This will help determine whether the 3-class solution suggested by Table 6 might be a common under-extraction error for data that truly have 4 classes or whether these indices tend to extract 4 classes when they are present.

Table 7 shows the simulation results. 2-, 3-, and 4-class models converged in 100% of replications and 5-class models converged in 74% of replications. BIC and ICL-BIC selected the correct 4-class solution in 81% and 85% of replications, respectively. SA-BIC and CLC did not perform as well and only selected the correct number of classes in 38% and 28% of replications, respectively. Importantly, no replications across all information criteria selected 2 or 3 classes. So, when data are generated from the 4-class Cat's Cradle pattern in Figure 1 with N = 300, t = 8, and 20% attrition, BIC and ICL-BIC very often detect the correct 4-class solution and no information criteria suggest 3 classes when the data truly contain 4 classes.

If there were truly 4 classes in the data, the information criteria reported in Section 6 would seem to err in the direction of reporting more than 4 classes rather than fewer as reported in Table 6. From these results, we conclude that the 3-class solution is unlikely to be an under-extraction error and we proceed to interpret the 3-class solution results in the next subsection.

6.2 Results

The estimated class trajectories for the 3-class solution are shown in Figure 4 and the empirical data of people assigned to each class are shown in Figure 5. The breakdown of the classes is essentially that Class 1 shows people who show some symptoms initially but improve to show no symptoms after one year (similar to the traditional Recovering class and the higher portion of the traditional Resilient class). Class 2 is composed of people who showed no symptoms throughout the observation window and is essentially those at the floor of the scale (the most unambiguous members of the traditional Resilient class). Class 3 is people who show symptoms and continue to show symptoms after one year and includes those who responded erratically over time.

Note the widely different variance of the growth trajectories in each class. If constraining variances to be equal across classes, the solution would look quite different to satisfy the imposed homogeneity assumption. Differences in variability across classes seems to be a defining characteristic of the class solution as the classes appear to be defined more by the volatility of symptoms rather than the mean change in symptoms. The unappreciated importance of the covariance has recently been noted[75–77] and the current analysis is an example of how these considerations similarly apply to GMMs but are discarded with the common GMMC approach that applies equality constraints on covariance parameters in different classes.

The proportion of people classified into the Resilient class tends to be large in most empirical studies and the informative Bayesian analysis of this same data placed 76% of the sample into this class. However, after allowing the covariance matrices to be class-specific, the percentage of people who are classified as Resilient drops considerably to 54% (if considering Class 1 and Class 2 combined as "Resilient"). This decline in Resilient classification mirrors findings from Infurna and Luthar[17], who found that Resiliency classification dropped from 81% with a constrained covariance structure to 48% with an unconstrained covariance structure in a different dataset. The Resilient class has little variance because the class is at the floor of the scale and shows little growth, so constraining the covariance structure to be equal across classes in such context artificially expands the reach of the Resilient class, resulting in cannibalization of the next nearest class into the Resilient class in order to satisfy unsupported homogeneity assumptions.

Also relevant is the lack of the Delayed Onset class, a type of PTSD whose existence has been debated in the psychiatry literature outside of military populations[78–81] but whose evidence for existence is taken largely from GMMs.[57] It is also interesting to note that the only two applications in the PTSD studies reviewed in Section 3 with sufficiently rich data to fit a GMMU with class-specific covariance matrices both yielded *two* classes, not four classes as found in studies that used GMMCs or LCGMs.

Though the true structure of empirical data is unknown, these results are consistent with the simulation results and demonstrate how conclusions can be affected by model alterations that are applied solely with an interest in model convergence without being accompanied by a theoretical justification. Much of the PTSD literature suggests four classes, but our analysis along with other work in this area suggests that the intended version of the model researchers want to fit if not for nonconvergence issues tends to uncover fewer classes with different class proportions and different interpretations, largely because differences in variability of the class is a central feature of the class solution and covariance parameters should not be constrained.

7 Discussion

Despite the heavy emphasis that researchers place on the mean structure in GMMs, the covariance structure is equally important for ensuring that results and conclusions are accurate. Covariance structure parameters are the most difficult to estimate and the task does not become any simpler when latent classes are present. Nonconvergence with the most general GMMU is therefore pervasive and the analysis often devolves into post-hoc processes of trial-and-error to determine which constraints or alterations to the covariance structure will yield convergence. We are sympathetic to the intentions behind these efforts; however, this study extends previous work showing that this approach provides poor class assignment, poor class trajectories, and over-extracts classes. Our application echoed the narrative that has recently appeared in the literature that latent classes of PTSD trajectories might be artifacts of methodological choices rather than reflective of the nature of PTSD itself.[40]

Our results suggest that the recently proposed CPGMM is well-positioned to address some of the longstanding nonconvergence present in applications of GMMs with realistic data characteristics. GMMs following the random effect tradition stack latent classes on top of latent growth trajectories and estimation algorithms can only extract so much latent information from a set of a few observed repeated measures. Whereas the *person* is the focus of a random effect growth models, the *class* is the focus of GMMs. As evidence of this, of the 34 reviewed PTSD papers, *zero* included any mention of person-specific trajectories. The interests were universally to (a) determine how many classes exist, (b) discern the class-specific trajectories and (c) determine which covariates were related to class membership. The lack of person-specific trajectories is not an isolated feature of the PTSD literature and has been noted as a feature of GMM applications broadly.[22–24] This suggests that either the within-class random effects are afterthoughts that are included to model the within-class covariance structure or there is a lack of realization that the population-average versus person-specific distinction in longitudinal data analysis extends to GMMs. As in any other longitudinal analysis, if the goal is simply in the average trajectory, researchers need not pursue random effects and person-specific models – answers can more easily be obtained with population-average approaches like covariance pattern models.

7.1 Limitations and Future Directions

First, our study considered continuous outcomes only, but GMMs are sometimes applied to discrete outcomes.[82] With binary data, covariance patterns models are supplanted by generalized estimating equations (GEE)[18] which are better equipped to deal with the nuances of modeling discrete outcomes. There is some literature considering this idea[83,84] and further development of a GEE version of GMMs would be a clear future direction to extend the current line of reasoning to the context of discrete outcomes. Additionally, similar to GEE, the cluster-robust sandwich estimator for standard errors can also be applied to CPGMMs to reduce effects of covariance structure misspecification (the default robust estimator in M*plus* addresses deviation from normality, not covariance misspecification).

Second, our simulation results largely assumed that the correct number of classes was known. Class enumeration is an undoubtedly important area to refine further within CPGMMs because the favorable properties demonstrated in this study are useful only to the extent that the model is capable of reasonably identifying the proper number of classes. Our small-scale simulation showed that BIC and ICL-BIC coupled with CPGMMs were able to accurately identify the correct number of classes and suggested that the 3-class solution in the application was not an under-extraction error. However, a rigorous full-scale class enumeration simulation study is needed before making comprehensive conclusions about the ability of CPGMMs to extract the correct number of classes. This would include generating data from models with different numbers of classes with different class proportions and including a broader array of information criteria and fit statistics to gauge under what conditions CPGMMs can and cannot reliably extract the correct number of classes. A simulation with sufficient complexity to address these questions is deserving of a dedicated study of its own.

Third, our simulation conditions were heavily based upon data characteristics observed from the PTSD literature. GMMs are commonly applied in many areas, especially substance use research, where the characteristics may look different. To the extent that attrition, sample size, class separation, and class separation differ from the PTSD literature; convergence and performance noted in our simulation may not carryover to other adjacent areas of application.

Acknowledgements

Both authors were supported by the Institute of Educational Sciences program in statistical and research methodology, project number R305D190011. The view and opinions expressed are those of the authors and do not necessarily reflect those of IES. The data in the application were supplied by the original authors with a Creative Common BY 4.0 International license that allows use provided that proper attribution is provided to the original authors. We thank these authors for their transparency and commitment to making scientific research an open process.

Supplemental Materials

The simulation files and empirical data that support the findings of this study are openly available from the first author's Open Science Framework page, located at https://osf.io/k9y54/

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

References

1. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. John Wiley & Sons; 2006.

2. Nesselroade JR. Interindividual differences in intraindividual change. In: *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions*. American Psychological Association; 1991:92-105. doi:10.1037/10099-006

3. Ram N, Grimm K. Using simple and complex growth models to articulate developmental change: Matching theory to method. *Int J Behav Dev*. 2007;31(4):303-316. doi:10.1177/0165025407077751

4. Jung T, Wickrama K a. S. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Soc Personal Psychol Compass*. 2008;2(1):302-317. doi:10.1111/j.1751-9004.2007.00054.x

5. Verbeke G, Lesaffre E. A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *J Am Stat Assoc*. 1996;91(433):217-221. doi:10.1080/01621459.1996.10476679

6. Muthén B, Shedden K. Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics*. 1999;55(2):463-469. doi:10.1111/j.0006-341X.1999.00463.x

7. Kim S-Y. Sample Size Requirements in Single- and Multiphase Growth Mixture Models: A Monte Carlo Simulation Study. *Struct Equ Model Multidiscip J*. 2012;19(3):457-476. doi:10.1080/10705511.2012.687672

8. Martin DP, Oertzen T von. Growth Mixture Models Outperform Simpler Clustering Algorithms When Detecting Longitudinal Heterogeneity, Even With Small Sample Sizes. *Struct Equ Model Multidiscip J*. 2015;22(2):264-275. doi:10.1080/10705511.2014.936340

9. Wickrama KKAS, Lee TK, O'Neal CW, et al. *Higher-Order Growth Curves and Mixture Modeling with Mplus : A Practical Guide*. Routledge; 2016. doi:10.4324/9781315642741

10. Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychol Methods*. 1999;4(2):139-157. doi:10.1037/1082-989X.4.2.139

11. Heggeseth BC, Jewell NP. The impact of covariance misspecification in multivariate Gaussian mixtures on estimation and inference: an application to longitudinal modeling. *Stat Med*. 2013;32(16):2790-2803. doi:10.1002/sim.5729

12. Davies CE, Glonek GF, Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. *Stat Methods Med Res*. 2017;26(4):1982-1991. doi:10.1177/0962280215598806

13. Diallo TMO, Morin AJS, Lu H. Impact of Misspecifications of the Latent Variance–Covariance and Residual Matrices on the Class Enumeration Accuracy of Growth Mixture Models. *Struct Equ Model Multidiscip J*. 2016;23(4):507-531. doi:10.1080/10705511.2016.1169188

14. Kreuter F, Muthén B. Analyzing Criminal Trajectory Profiles: Bridging Multilevel and Group-based Approaches Using Growth Mixture Modeling. *J Quant Criminol*. 2008;24(1):1-31. doi:10.1007/s10940-007-9036-0

15. Shiyko MP, Ram N, Grimm KJ. An overview of growth mixture modeling: A simple nonlinear application in OpenMx. In: *Handbook of Structural Equation Modeling*. The Guilford Press; 2012:532-546.

16. Bauer DJ, Curran PJ. The Integration of Continuous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities. *Psychol Methods*. 2004;9(1):3-29. doi:10.1037/1082-989X.9.1.3

17. Infurna FJ, Luthar SS. Resilience to Major Life Stressors Is Not as Common as Thought. *Perspect Psychol Sci*. 2016;11(2):175-194. doi:10.1177/1745691615621271

18. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22. doi:10.1093/biomet/73.1.13

19. McNeish D, Harring J. Covariance pattern mixture models: Eliminating random effects to improve convergence and performance. *Behav Res Methods*. Published online September 11, 2019. doi:10.3758/s13428-019-01292-4

20. Anderlucci L, Viroli C. Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data. *Ann Appl Stat*. 2015;9(2):777-800. doi:10.1214/15-AOAS816

21. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. *Epidemiology*. 2010;21(4):467-474. Accessed April 18, 2020. https://www.jstor.org/stable/25680575

22. Cole VT, Bauer DJ. A Note on the Use of Mixture Models for Individual Prediction. *Struct Equ Model Multidiscip J*. 2016;23(4):615-631. doi:10.1080/10705511.2016.1168266

23. Sterba SK, Bauer DJ. Predictions of Individual Change Recovered With Latent Class or Random Coefficient Growth Models. *Struct Equ Model Multidiscip J*. 2014;21(3):342-360. doi:10.1080/10705511.2014.915189

24. Sterba SK, Bauer DJ. Matching method with theory in person-oriented developmental psychopathology research. *Dev Psychopathol*. 2010;22(2):239-254. doi:10.1017/S0954579410000015

25. Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990;55(1):107-122. doi:10.1007/BF02294746

26. McArdle JJ, Epstein D. Latent Growth Curves within Developmental Structural Equation Models. *Child Dev*. 1987;58(1):110-133. doi:10.2307/1130295

27. Muthén BO. Beyond Sem: General Latent Variable Modeling. *Behaviormetrika*. 2002;29(1):81-117. doi:10.2333/bhmk.29.81

28. Ram N, Grimm KJ. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *Int J Behav Dev*. 2009;33(6):565-576. doi:10.1177/0165025409343765

29. Grimm KJ, Ram N, Estabrook R. *Growth Modeling: Structural Equation and Multilevel Modeling Approaches*. Guilford Publications; 2016.

30. Kiernan K. Insights into Using the GLIMMIX Procedure to Model Categorical Outcomes with Random Effects. *SAS Inst*. Published online 2018:SAS2179-2018.

31. Liu M, Hancock GR. Unrestricted Mixture Models for Class Identification in Growth Mixture Modeling. *Educ Psychol Meas*. 2014;74(4):557-584. doi:10.1177/0013164413519798

32. Pastor DA, Gagné P. Mean and covariance structure mixture models. In: *Structural Equation Modeling: A Second Course, 2nd Ed*. Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. IAP Information Age Publishing; 2013:343-393.

33. Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. *Psychol Methods*. 2006;11(1):36-53. doi:10.1037/1082-989X.11.1.36

34. Biernacki C. Testing for a Global Maximum of the Likelihood. *J Comput Graph Stat*. 2005;14(3):657-674. doi:10.1198/106186005X59298

35. McLachlan GJ, Peel D. *Finite Mixture Models*. John Wiley & Sons; 2004.

36. Banfield JD, Raftery AE. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*. 1993;49(3):803-821. doi:10.2307/2532201

37. Infurna FJ, Grimm KJ. The Use of Growth Mixture Modeling for Studying Resilience to Major Life Stressors in Adulthood and Old Age: Lessons for Class Size and Identification and Model Selection. *J Gerontol Ser B*. 2018;73(1):148-159. doi:10.1093/geronb/gbx019

38. Bauer DJ, Curran PJ. Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychol Methods*. 2003;8(3):338-363. doi:10.1037/1082-989X.8.3.338

39. Gilthorpe MS, Dahly DL, Tu Y-K, Kubzansky LD, Goodman E. Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *J Dev Orig Health Dis*. 2014;5(3):197-205. doi:10.1017/S2040174414000130

40. Infurna FJ, Jayawickreme E. Fixing the Growth Illusion: New Directions for Research in Resilience and Posttraumatic Growth. *Curr Dir Psychol Sci*. 2019;28(2):152-158. doi:10.1177/0963721419827017

41. Kooken J, McCoach DB, Chafouleas SM. The Impact and Interpretation of Modeling Residual Noninvariance in Growth Mixture Models. *J Exp Educ*. 2019;87(2):214-237. doi:10.1080/00220973.2017.1421516

42. Nagin DS, Tremblay RE. Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychol Methods*. 2001;6(1):18-34. doi:10.1037//1082-989X.6.1.18

43. Nagin DS, Tremblay RE. Developmental Trajectory Groups: Fact or a Useful Statistical Fiction?*. *Criminology*. 2005;43(4):873-904. doi:10.1111/j.1745-9125.2005.00026.x

44. Kreuter F, Muthén BO. Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles. In: Gregory R. Hancock, Samuleson KM, eds. *Advances in Latent Variable Mixture Models*. Information Age Publishing; 2007:53-75.

45. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. John Wiley & Sons; 2012.

46. Muthén BO. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan D, ed. *Handbook of Quantitative Methodology for the Social Sciences*. ; 2004:345-368.

47. Sijbrandij JJ, Hoekstra T, Almansa J, Reijneveld SA, Bültmann U. Identification of developmental trajectory classes: Comparing three latent class methods using simulated and real data. *Adv Life Course Res*. 2019;42:100288. doi:10.1016/j.alcr.2019.04.018

48. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *J Clin Epidemiol*. 2012;65(10):1078-1087. doi:10.1016/j.jclinepi.2012.04.010

49. Jennrich RI, Schluchter MD. Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics*. 1986;42(4):805-820. doi:10.2307/2530695

50. Schluchter MD. Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Stat Med*. 1988;7(1-2):317-324. doi:10.1002/sim.4780070132

51. Lee Y, Nelder JA. Conditional and Marginal Models: Another View. *Stat Sci*. 2004;19(2):219-238. doi:10.1214/088342304000000305

52. Davidian M, Giltinan DM. Nonlinear models for repeated measurement data: An overview and update. *J Agric Biol Environ Stat*. 2003;8(4):387. doi:10.1198/1085711032697

53. Harring JR, Blozis SA. Fitting correlated residual error structures in nonlinear mixed-effects models using SAS PROC NLMIXED. *Behav Res Methods*. 2014;46(2):372-384. doi:10.3758/s13428-013-0397-z

54. van de Schoot R, Sijbrandij M, Depaoli S, Winter SD, Olff M, Loey NE van. Bayesian PTSD-Trajectory Analysis with Informed Priors Based on a Systematic Literature Search and Expert Elicitation. *Multivar Behav Res*. 2018;53(2):267-291. doi:10.1080/00273171.2017.1412293

55. Lee DY. Handling of Missing Data with Growth Mixture Models. Published online 2019. Accessed December 23, 2019. http://drum.lib.umd.edu/handle/1903/22129

56. Enders CK, Tofighi D. The Impact of Misspecifying Class-Specific Residual Variances in Growth Mixture Models. *Struct Equ Model Multidiscip J*. 2008;15(1):75-95. doi:10.1080/10705510701758281

57. Bonanno GA. Loss, Trauma, and Human Resilience: Have We Underestimated the Human Capacity to Thrive After Extremely Aversive Events? *Am Psychol*. 2004;59(1):20-28. doi:10.1037/0003-066X.59.1.20

58. Sher KJ, Jackson KM, Steinley D. Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *J Abnorm Psychol*. 2011;120(2):322-335. doi:10.1037/a0021813

59. Diallo TMO, Morin AJS, Parker PD. Statistical power of latent growth curve models to detect quadratic growth. *Behav Res Methods*. 2014;46(2):357-371. doi:10.3758/s13428-013-0395-1

60. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media; 2009.

61. Tofighi D, Enders CK. Identifying the correct number of classes in a growth mixture model. In: Hancock GR, Samuleson KM, eds. *Advances in Latent Variable Mixture Models*. Information Age Publishing; 2008:317-341.

62. Tueller S, Lubke G. Evaluation of Structural Equation Mixture Models: Parameter Estimates and Correct Class Assignment. *Struct Equ Model Multidiscip J*. 2010;17(2):165-192. doi:10.1080/10705511003659318

63. van Loey NEE, Maas CJM, Faber AW, Taal LA. Predictors of chronic posttraumatic stress symptoms following burn injury: Results of a longitudinal study. *J Trauma Stress*. 2003;16(4):361-369. doi:10.1023/A:1024465902416

64. Horowitz MJ, Wilner NR, Álvarez WM. Impact of Event Scale: a measure of subjective stress. *Psychosom Med*. 1979;41(3):209-218. doi:10.1097/00006842-197905000-00004

65. Li M, Harring JR, Macready GB. Investigating the Feasibility of Using Mplus in the Estimation of Growth Mixture Models. *J Mod Appl Stat Methods*. 2014;13(1):484-513. doi:10.22237/jmasm/1398918600

66. Muthén B, Asparouhov T. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychol Methods*. 2012;17(3):313-335. doi:10.1037/a0026802

67. Wolfinger RD. Heterogeneous Variance: Covariance Structures for Repeated Measures. *J Agric Biol Environ Stat*. 1996;1(2):205-230. doi:10.2307/1400366

68. McNeish D, Bauer DJ. Reducing Incidence of Nonpositive Definite Covariance Matrices in Mixed Effect Models. *Multivar Behav Res*. Published online 2020:advance online publication. doi:10.1080/00273171.2020.1830019

69. Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*. 1987;52(3):333-343. doi:10.1007/BF02294360

70. Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans Pattern Anal Mach Intell*. 2000;22(7):719-725. doi:10.1109/34.865189

71. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Struct Equ Model Multidiscip J*. 2007;14(4):535-569. doi:10.1080/10705510701575396

72. Yang C-C. Evaluating latent class analysis models in qualitative phenotype identification. *Comput Stat Data Anal*. 2006;50(4):1090-1104. doi:10.1016/j.csda.2004.11.004

73. Henson JM, Reise SP, Kim KH. Detecting Mixtures From Structural Model Differences Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics. *Struct Equ Model Multidiscip J*. 2007;14(2):202-226. doi:10.1080/10705510709336744

74. Pauler DK, Laird NM. A Mixture Model for Longitudinal Data with Application to Assessment of Noncompliance. *Biometrics*. 2000;56(2):464-472. doi:10.1111/j.0006-341X.2000.00464.x

75. Goldstein H, Leckie G, Charlton C, Tilling K, Browne WJ. Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. *Stat Methods Med Res*. 2018;27(11):3478-3491. doi:10.1177/0962280217706728

76. Wang L (Peggy), Hamaker E, Bergeman CS. Investigating inter-individual differences in short-term intra-individual variability. *Psychol Methods*. 2012;17(4):567-581. doi:10.1037/a0029317

77. Sternad D. It's not (only) the mean that matters: variability, noise and exploration in skill learning. *Curr Opin Behav Sci*. 2018;20:183-195. doi:10.1016/j.cobeha.2018.01.004

78. Andrews B, Brewin CR, Philpott R, Stewart L. Delayed-Onset Posttraumatic Stress Disorder: A Systematic Review of the Evidence. *Am J Psychiatry*. 2007;164(9):1319-1326. doi:10.1176/appi.ajp.2007.06091491

79. Frueh BC, Grubaugh AL, Yeager DE, Magruder KM. Delayed-onset post-traumatic stress disorder among war veterans in primary care clinics. *Br J Psychiatry*. 2009;194(6):515–520.

80. McNally RJ. Progress and Controversy in the Study of Posttraumatic Stress Disorder. *Annu Rev Psychol*. 2003;54(1):229-252. doi:10.1146/annurev.psych.54.101601.145112

81. Spitzer RL, First MB, Wakefield JC. Saving PTSD from itself in DSM-V. *J Anxiety Disord*. 2007;21(2):233-241. doi:10.1016/j.janxdis.2006.09.006

82. Proust-Lima C, Letenneur L, Jacqmin-Gadda H. A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Stat Med*. 2007;26(10):2229-2245. doi:10.1002/sim.2659

83. Rosen O, Jiang W, Tanner MA. Mixtures of marginal models. *Biometrika*. 2000;87(2):391-404. doi:10.1093/biomet/87.2.391

84. Tang X, Qu A. Mixture Modeling for Longitudinal Data. *J Comput Graph Stat*. 2016;25(4):1117-1137. doi:10.1080/10618600.2015.1092979

Table 1
Comparison of methods to model heterogeneity in growth trajectories

Model	Unconditional Marginal Covariance Structure	Within-Class Random Effects	Person-Specific Curves	Primary Advantage	Primary Disadvantage
GMMU	$\mathbf{\Sigma}_{ik} = \mathbf{T}_i \mathbf{G}_k \mathbf{T}_i^{\mathrm{T}} + \mathbf{R}_{ik}$	Yes	Yes	Complete random effect growth model in each class	Many random effects make estimation difficult
GMMC	$\mathbf{\Sigma}_{ik} = \mathbf{\Sigma}_i = \mathbf{T}_i \mathbf{G} \mathbf{T}_i^{\mathrm{T}} + \mathbf{R}_i$	Yes	Yes	Resembles GMMU but simplifies estimation	Prioritizes convergence over covariance structure flexibility
LGCM	$\mathbf{\Sigma}_{ik} = \sigma_k^2 \mathbf{I}_t$	No	No	Minimal assumptions and simplified estimation	Covariances between repeated measures not modeled[a]
CPGMM	$\mathbf{\Sigma}_{ik} = \mathbf{\Sigma}_{ik}(\boldsymbol{\theta}_k)$	No	No	Class-specific marginal covariance without random effects	Researcher must select pattern for marginal covariance structure

[a] This is a disadvantage from the viewpoint of defining classes as a set of people that can be described by a single probability distribution. From the viewpoint that a class is a collection of people who follow a similar and distinct trajectory, this would likely be consider an advantage.

Note: GMMU = Unconstrained Growth Mixture Model, GMMC = Constrained Growth Mixture Model, LCGM = Latent Class Growth Model, CPGMM = Covariance Pattern Growth Mixture Model. $\mathbf{I}_t$ is an identity matrix with dimension equal to the number of repeated measures, *t*.

Table 2
Data generation equations for population model

Class	Model Equation	Covariance Structures, Low Separation	Covariance Structures, High Separation
1 (Resilient)	$y_{ti} = \beta_{0i} + \beta_{1i}t + \varepsilon_{ti}$ $\beta_{0i} = 15.0 + b_{0i}$ $\beta_{1i} = -0.15 + b_{1i}$	$\mathbf{G} = \begin{bmatrix} 40 & \\ 0 & 0.04 \end{bmatrix}$ $\mathbf{R} = 28\mathbf{I}_5$	$\mathbf{G} = \begin{bmatrix} 25 & \\ 0 & 0.02 \end{bmatrix}$ $\mathbf{R} = 25\mathbf{I}_5$
2 (Recovering)	$y_{ti} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \varepsilon_{ti}$ $\beta_{0i} = 38.0 + b_{0i}$ $\beta_{1i} = -0.1 + b_{1i}$ $\beta_{2i} = -0.008$	$\mathbf{G} = \begin{bmatrix} 160 & \\ 0 & 0.04 \end{bmatrix}$ $\mathbf{R} = 72\mathbf{I}_5$	$\mathbf{G} = \begin{bmatrix} 20 & \\ 0 & 0.01 \end{bmatrix}$ $\mathbf{R} = 10\mathbf{I}_5$
3 (Chronic)	$y_{ti} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \varepsilon_{ti}$ $\beta_{0i} = 41.0 + b_{0i}$ $\beta_{1i} = 0.12 + b_{1i}$ $\beta_{2i} = -0.002$	$\mathbf{G} = \begin{bmatrix} 120 & \\ 0 & 0.04 \end{bmatrix}$ $\mathbf{R} = 60\mathbf{I}_5$	$\mathbf{G} = \begin{bmatrix} 15 & \\ 0 & 0.015 \end{bmatrix}$ $\mathbf{R} = 15\mathbf{I}_5$
4 (Delayed Onset)	$y_{ti} = \beta_{0i} + \beta_{1i}t + \beta_{2i}t^2 + \varepsilon_{ti}$ $\beta_{0i} = 18.0 + b_{0i}$ $\beta_{1i} = 0.2 + b_{1i}$ $\beta_{2i} = 0.015$	$\mathbf{G} = \begin{bmatrix} 72 & \\ 0 & .035 \end{bmatrix}$ $\mathbf{R} = 60\mathbf{I}_5$	$\mathbf{G} = \begin{bmatrix} 10 & \\ 0 & 0.015 \end{bmatrix}$ $\mathbf{R} = 12\mathbf{I}_5$

Note: t = 0, 1, 10, 18, 26

Table 3

Percentage of replications that converged each model across simulation conditions

		N =100				N =250				N =500			
Separation	Attrition	CPGMM	LCGM	GMMC	GMMU	CPGMM	LCGM	GMMC	GMMU	CPGMM	LCGM	GMMC	GMMU
Low													
	0%	84%	94%	70%	4%	97%	100%	81%	17%	100%	100%	68%	26%
	20%	92%	96%	62%	5%	98%	100%	82%	15%	100%	100%	66%	23%
	45%	88%	100%	54%	1%	98%	100%	78%	9%	100%	100%	72%	18%
High													
	0%	88%	100%	68%	7%	100%	100%	82%	29%	100%	100%	82%	50%
	20%	96%	100%	63%	7%	100%	100%	78%	24%	100%	100%	82%	46%
	45%	97%	100%	57%	4%	100%	100%	72%	15%	100%	100%	75%	36%

Note: CPGMM = Covariance Pattern Growth Mixture Model, LCGM = Latent Class Growth Model, GMMC = Constrained Growth Mixture Model, GMMU = Unconstrained Growth Mixture Model with class-specific covariance matrices

Table 4

Mean absolute bias of class proportions for each model across simulation conditions

		N =100				N =250				N =500			
Separation	Attrition	CPGMM	LCGM	GMMC	GMMU	CPGMM	LCGM	GMMC	GMMU	CPGMM	LCGM	GMMC	GMMU
Low													
	0%	**8.1**	18.7	10.2	11.4	7.3	20.3	**6.1**	9.4	7.3	20.6	**4.3**	7.0
	20%	**7.7**	18.9	12.1	12.3	**6.4**	20.6	6.5	10.3	6.8	20.8	**4.3**	7.6
	45%	**9.7**	18.6	12.6	12.0	**6.2**	20.4	8.2	11.2	6.1	20.7	**4.7**	9.9
High													
	0%	**3.5**	12.6	13.8	8.6	**1.7**	12.9	13.6	4.0	1.5	12.8	14.9	**0.8**
	20%	**3.0**	12.8	13.9	9.6	**1.2**	13.1	9.6	5.0	**0.9**	13.0	15.0	1.1
	45%	**4.9**	12.7	14.9	10.0	**0.9**	13.4	14.4	6.1	**1.2**	13.5	15.9	1.8

Note: Bold entries indicate the model with the class proportions that most closely represent the values from the data generation model for each combination of conditions. CPGMM = Covariance Pattern Growth Mixture Model, LCGM = Latent Class Growth Model, GMMC = Constrained Growth Mixture Model, GMMU = Unconstrained Growth Mixture Model with class-specific covariance matrices

Table 5
Summary classification of GMM methods by class-specificity of the covariance structure and the complexity of the estimation

	Low Estimation Complexity	High Estimation Complexity
High Covariance Specificity	Covariance Pattern GMM	Unconstrained GMM
Low Covariance Specificity	Latent Class Growth Model	Constrained GMM

Table 6
Comparison of fit for different numbers of classes for PTSD data

	Classes			
	1	2	3	4
Loglikelihood	-7931.8	-7492.1	-7194.2	-7239.0
Relative Entropy	---	.945	.912	.845
Parameters	19	39	59	65
BIC	15,972	15,207	14,725	14,849
SA-BIC	15,912	15,083	14,538	14,643
BIC-ICL	---	15,230	14,783	14,978
CLC	---	15,007	14,447	14,607

Note: Entropy based measures are undefined for models with one class. The number of parameters in the 4-class model are relatively close to the number of parameters in 3-class model because covariance structure of one class was reduced to a compound symmetric structure to be more parsimonious because the variances and covariances were very similar across time.

Table 7
Percentage of simulation replications suggesting each class solution with N =300, t = 8, and 20% attrition

	Classes			
Metric	2	3	4	5
BIC	0	0	81	19
SA-BIC	0	0	38	62
ICL-BIC	0	0	85	15
CLC	0	0	28	72
Convergence	100	100	100	74

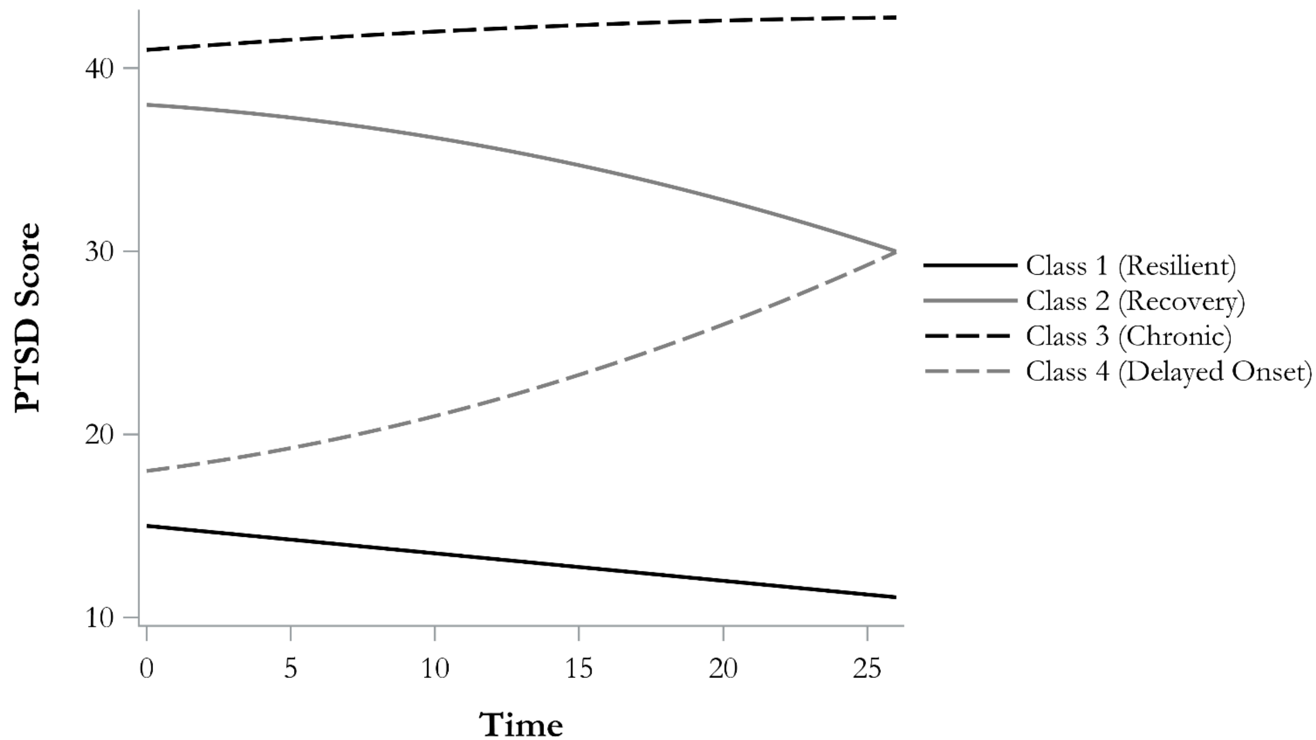

Figure 1. Generated population mean trajectories. Proportions of people were unevenly assigned to each class with 63% in Class 1, 12% in Class 2, 19% in Class 3, and 6% in Class 4.

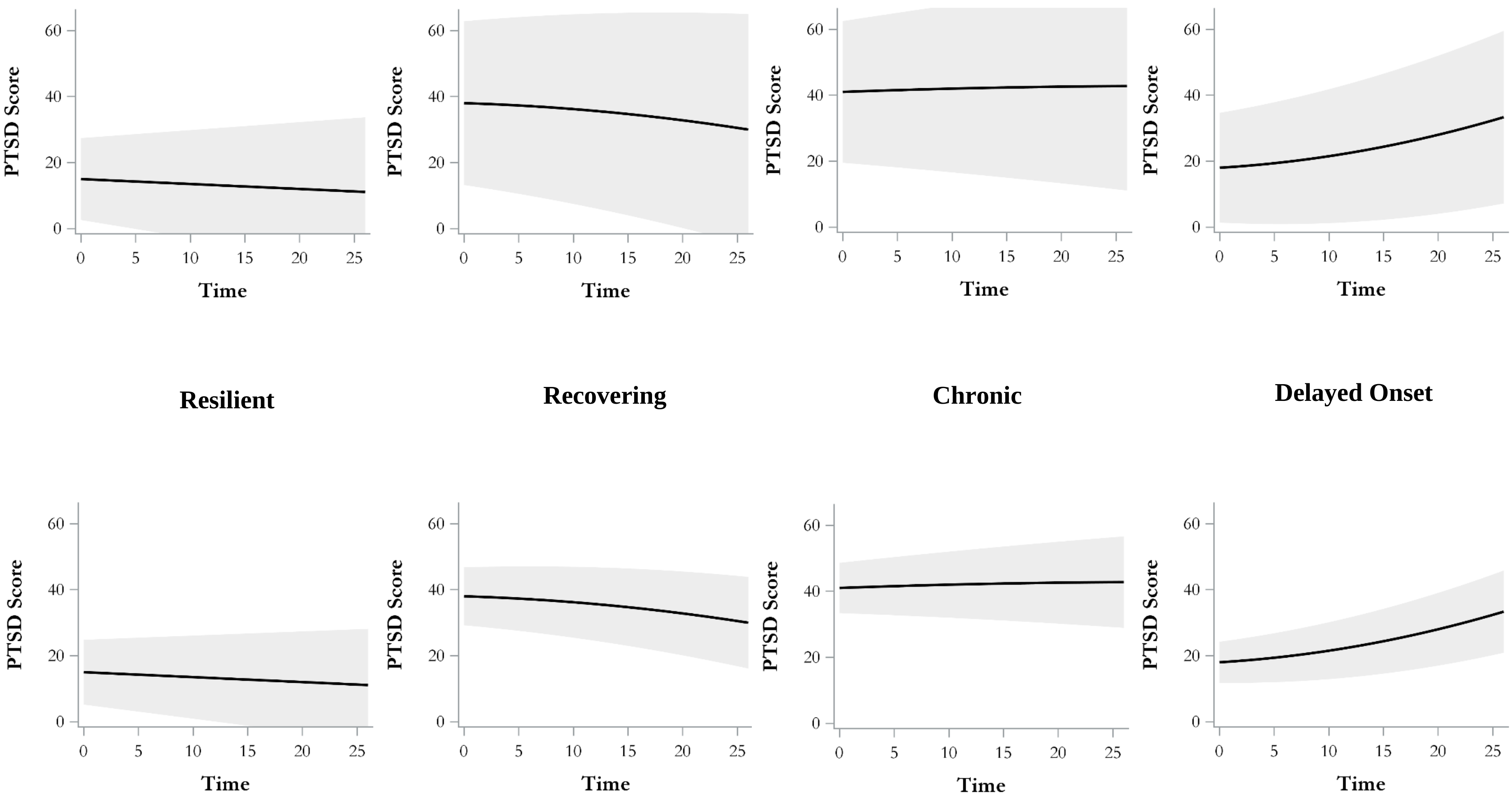

Figure 2. Class-specific growth trajectories with 95% intervals shaded in grey for the low separation condition (top row) and the high separation condition (bottom row). From left the right, the classes depicted are (a) Resilient, (b) Recovering, (c) Chronic, (d) Delayed Onset

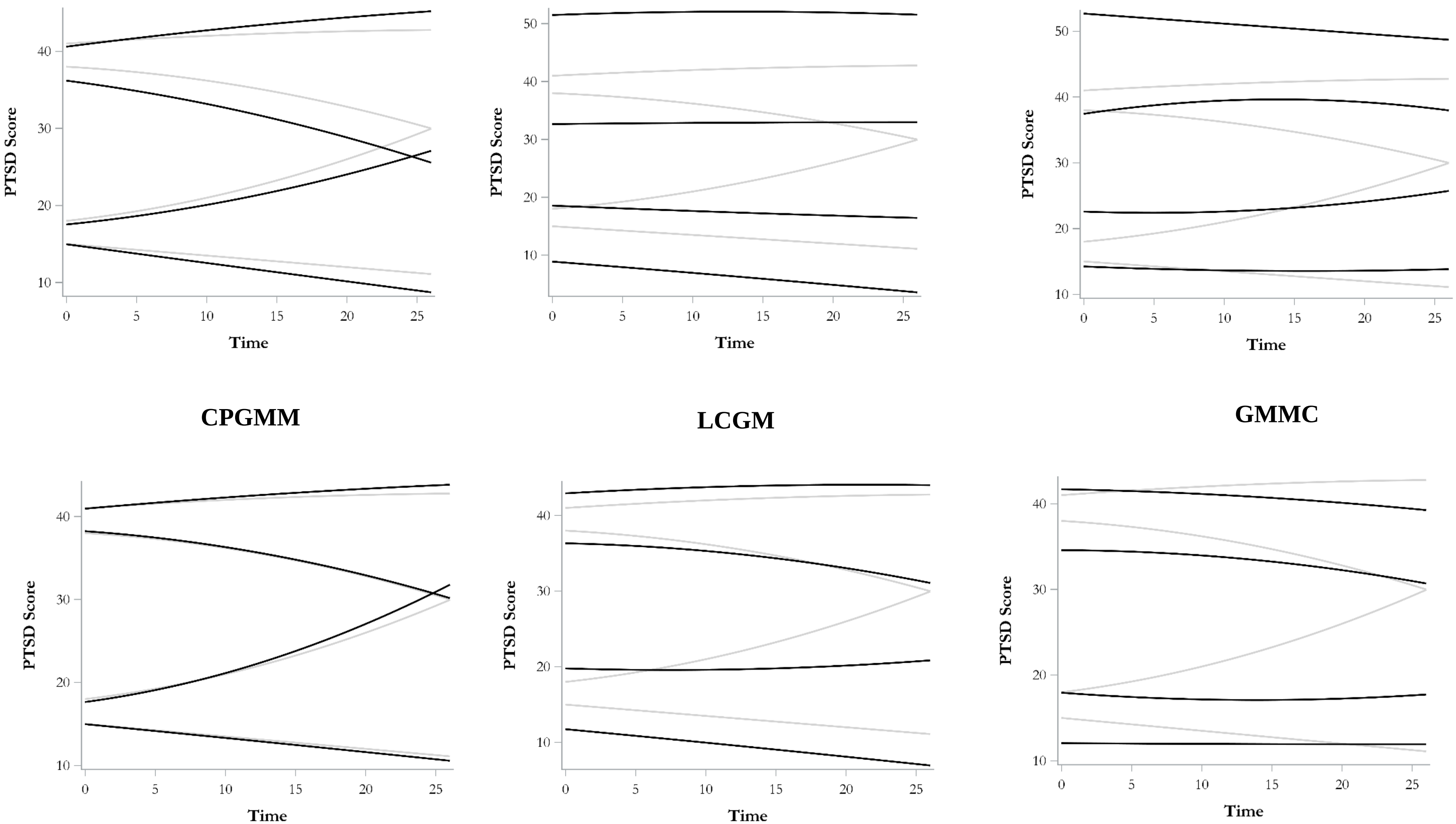

Figure 3. Comparison of population class trajectories (grey) to estimated class trajectories averaged across replications (black) for CPGMM, LCGM, and GMMC. The top row is for N = 250, 45% attrition, and low class separation. The bottom row is for N = 250, 45% attrition, and high class separation.

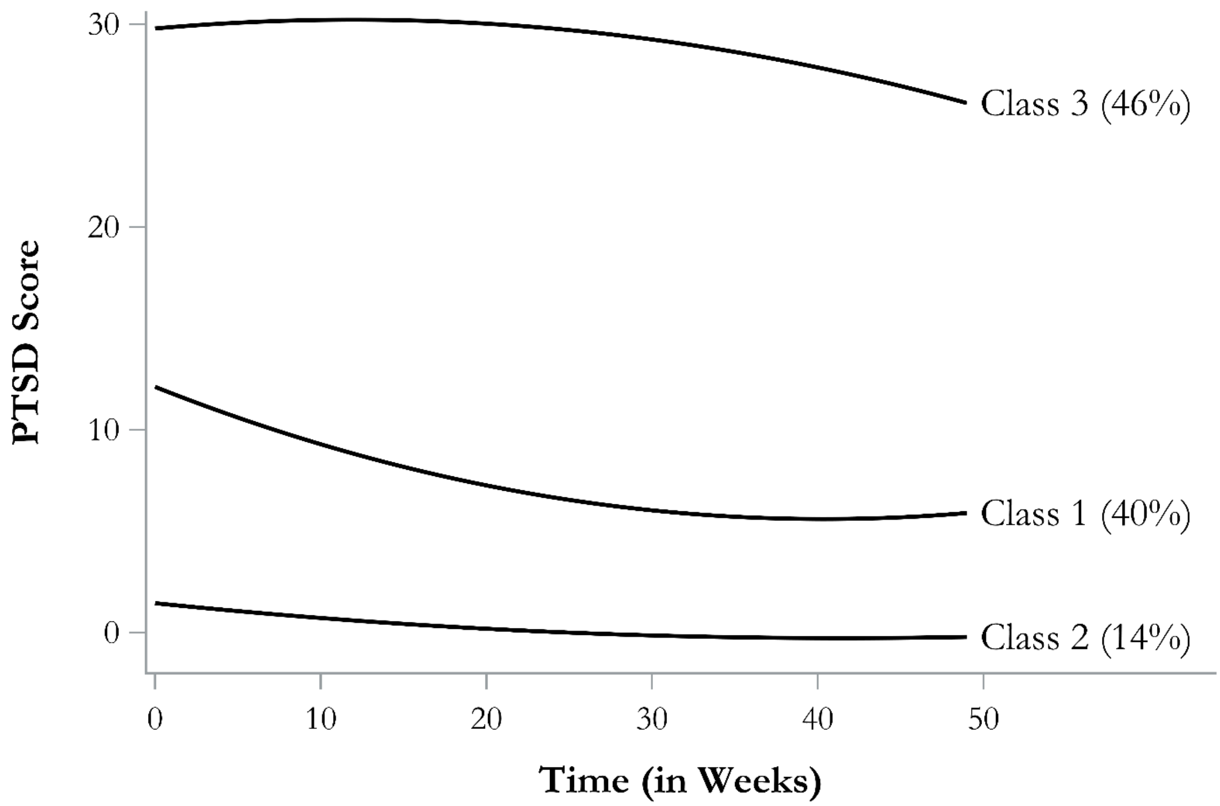

Figure 4. Class trajectories for CPGMM with class-specific factor analytic covariance using the 3-class solution.

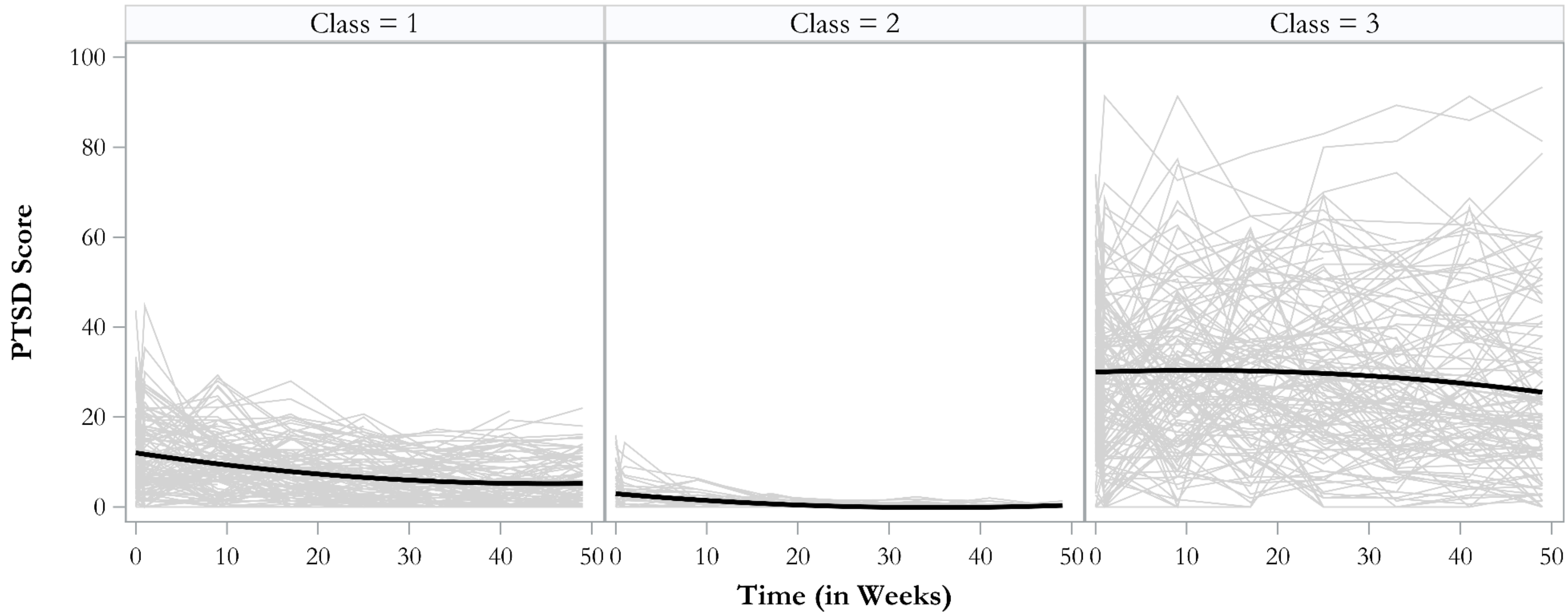

Figure 5. Plot of estimated class trajectories (in black) within the empirical data of people assigned to each class (in grey).