Fitting a Linear–Linear Piecewise Growth Mixture Model With Unknown Knots: A Comparison of Two Common Approaches to Inference

Nidhi Kohli, John Hughes, and Chun Wang
University of Minnesota
Cengiz Zopluoglu
University of Miami
Mark L. Davison
University of Minnesota

A linear–linear piecewise growth mixture model (PGMM) is appropriate for analyzing segmented (disjointed) change in individual behavior over time, where the data come from a mixture of 2 or more latent classes, and the underlying growth trajectories in the different segments of the developmental process within each latent class are linear. A PGMM allows the knot (change point), the time of transition from 1 phase (segment) to another, to be estimated (when it is not known a priori) along with the other model parameters. To assist researchers in deciding which estimation method is most advantageous for analyzing this kind of mixture data, the current research compares 2 popular approaches to inference for PGMMs: maximum likelihood (ML) via an expectation–maximization (EM) algorithm, and Markov chain Monte Carlo (MCMC) for Bayesian inference. Monte Carlo simulations were carried out to investigate and compare the ability of the 2 approaches to recover the true parameters in linear–linear PGMMs with unknown knots. The results show that MCMC for Bayesian inference outperformed ML via EM in nearly every simulation scenario. Real data examples are also presented, and the corresponding computer codes for model fitting are provided in the Appendix to aid practitioners who wish to apply this class of models.

Keywords: piecewise function, finite mixture, longitudinal data, Bayesian, maximum likelihood

Longitudinal statistical models are frequently used in behavioral and social science research to investigate how individuals’ performances, attitudes, and interests change over time, to ascertain whether growth is greater for one subpopulation than another, possibly as a result of a treatment effect (e.g., Shin, Davison, Long, Chan, & Heistad, 2013), or to determine whether the change process can be predicted from covariates (e.g., Grimm, 2008; Jordan, Hanich, & Kaplan, 2003; Miles & Miles, 1992). In particular, the latent growth curve (LGC) models that stem from a factor analysis tradition are based on the idea that the overall change process over time in observed repeated measurements is described by an underlying latent continuum (Meredith & Tisak, 1990). In this article we discuss the application of linear–linear piecewise growth mixture models (PGMM) in which the data come from a mixture of two or more latent classes, and the underlying growth trajectories in the different segments of the developmental process within each latent class are linear. Then, using Monte Carlo simulation, we compare two common approaches to parameter inference, maximum likelihood (ML) inference via an EM algorithm and MCMC for Bayesian inference, for the linear–linear PGMM.

In many psychological and educational research settings the longitudinal processes exhibit distinct phases of development in observed repeated measurements (e.g., Kreisman, 2003; Paris, 2005; Silverman, Specce, Harring, & Ritchey, 2012). For instance, in studies of early childhood intervention children may show relatively fast improvement in the postintervention follow-up period relative to the preintervention period (Kreisman, 2003). The trajectory characterizing the rate of change includes a slower acquisition phase that gives way to a later rapid acquisition phase. Another example where developmental processes are composed of distinct phases is found in the reading literature. It has been conjectured that in the beginning of second grade most students’ ability to accurately and automatically decode may increase but then changes to a different, slower rate in the middle of their third grade (Silverman et al., 2012). Because the rate of change is different in different phases, piecewise LGC models may be utilized to allow the specification of each growth phase to conform to a particular functional form of the overall change process (Chou, Yang, Pentz, & Hser, 2004; Cudeck & Harring, 2010). In this regard, the piecewise LGC modeling framework can summarize various functional forms in the different phases of development such that each phase does not have to follow the same function.
Furthermore, a review of the literature from various domains of psychology clearly shows a wide application of piecewise growth models. Researchers have used these models to examine, for instance, the development and/or decline in cognitive ability over time (e.g., Bradway & Thompson, 1962; Dominicus, Ripatti, Pedersen, & Palmgren, 2008; Hall, Lipton, Sliwinski, & Stewart, 2000; Hall, Ying, Kuo, Sliwinski, Buschke, Katz, & Lipton 2001), the onset of alcohol abuse (e.g., Li, Duncan, Duncan, & Hops, 2000; Hall, Lipton, Sliwinski, & Stewart, 2000; McCrady, Epstein, Hildebrandt, Cook, Jensen, & Hildebrandt, 2002), the acquisition of motor skills (e.g., Fox, Hershberger, & Bouchard, 1996; French, Lindenberger, & Kray, 1999), and prevention/intervention (e.g., Chou et al., 2004).

It is possible to specify different kinds of piecewise LGC models depending on the characteristics underlying the data. For example, a linear–linear piecewise LGC model could be specified when the trajectory in the first and second developmental stages is both linear. Alternatively, a quadratic–linear or exponential–linear piecewise LGC model could be specified when the trajectory in the first developmental stage has some curvature and the rate of change in the second developmental stage is constant.

One of the most interesting features of piecewise LGC models is the knot (change point), the time point at which the response function changes from one phase to another (Cudeck, 1996; Cudeck & Klebe, 2012). The knot can be known a priori or can be estimated. Figure 1 illustrates a generic piecewise LGC model that comprises two different linear growth phases. As shown, the linear pieces in Figure 1 join at the knot. This assumption could also be relaxed to allow disjointed functions in the various phases (Cudeck & Codding, 2012). Furthermore, different inference methods are available for piecewise LGC models with unknown knot locations. For instance, Wang and McArdle (2008) compared two approaches to inference for piecewise LGC models: MCMC for Bayesian inference, and frequentist inference via Taylor approximation and Gaussian quadrature. However, their model included only a single class; it is considerably more challenging to do inference when there are multiple classes. Although the linear–linear piecewise LGC model for a single class has been described in detail elsewhere (Cudeck & Harring, 2010; Cudeck & Klebe, 2002), we briefly review the model before discussing the linear–linear PGMM for two or more classes.

### Linear–Linear Piecewise LGC Model

Formulation of a linear–linear piecewise LGC model for change specifies a separate linear function for each of the two phases of development. The functional form at the $j$th time point is

$$y_{ij} = f_j + e_i$$

where

$$f_j = \begin{cases} \eta_{ij} + \eta_{2ij} t_j \leq \gamma \\ \eta_{ij} + \eta_{4ij} t_j > \gamma \end{cases}$$  \hspace{1cm} \text{(1)}$$

where $\eta_{ij}$ and $\eta_{4ij}$ refer to the intercept and the slope growth factors of the first phase, respectively; $\eta_{2ij}$ and $\eta_{4ij}$ refer to the intercept and the slope growth factors of the second phase, respectively; and $e_i$ refers to the residuals. When the growth function is polynomial of highest degree $K$, it is common practice to assume a smooth curve at the knot by imposing the assumption that the growth functions and the first $K - 1$ derivative meet at the knot. If the function is linear, this means that the linear function joins at the knot (i.e., $\eta_{1i} + \eta_{2i}\gamma = \eta_{3i} + \eta_{4i}\gamma$) in which case one of the parameters becomes redundant. Then the intercept of the second phase is $\eta_{3i} = \eta_{1i} + \eta_{2i}\gamma - \eta_{4i}\gamma$, and the function can be rewritten (Cudeck & Harring, 2010) as

$$f_j = \begin{cases} \eta_{1i} + \eta_{2ij} t_j \leq \gamma \\ \eta_{1i} + \eta_{2i}\gamma + \eta_{4i}(t_j - \gamma) t_j > \gamma \end{cases}$$  \hspace{1cm} \text{(2)}$$

The growth factors in Equation 2, $\mathbf{\eta}_i = (\eta_{1i}, \eta_{2i}, \eta_{4i})'$, are the growth characteristics of a linear function (i.e., intercept and slope) thought to be the sum of fixed and random effects, $\mathbf{\eta}_i = \mathbf{\alpha} + \mathbf{\zeta}_i$. The distributions of the random effects, $\mathbf{\zeta}_i$, and of the residuals, $e_i$, are both assumed to be multivariate normal:

$$\mathbf{\zeta}_i \sim N(\mathbf{0}, \mathbf{\Psi}) \quad e_i \sim N(0, \Theta_i).$$  \hspace{1cm} \text{(3)}$$

where $\Theta_i$ is often parameterized simply (i.e., $\Theta_i = \sigma^2_i \mathbf{I}$) when coupled with random effects, and $\mathbf{\Psi}$ denotes the covariances matrix of the random effects. Note that there is no subscript $i$ on $\gamma$ in Equations 1 and 2 because we assume that the knot location is common to all individuals. Furthermore, Equation 2 fits nicely into the LGC modeling framework. Across $n$ repeated measurements for subject $i$ ($i = 1, \ldots, N$) the function in Equation 2 can be written as a standard LGC model:

$$y_{ij} = \mathbf{A}_i(\gamma)\mathbf{\eta}_i + e_i,$$  \hspace{1cm} \text{(4)}$$

where the $(n \times p)$ factor loading matrix $\mathbf{A}_i$ is a function of both the time of measurement and the knot location.

This modeling framework provides sufficient flexibility to handle other functional forms in the different phases. For example, a quadratic–linear piecewise LGC model with unknown knot could be specified when the trajectory in the first developmental stage

![Figure 1. A line graph depicting generic linear-linear piecewise change process.](image-url)
has some curvature and the trajectory in the second developmental stage is linear. This kind of model can be specified by extending the linear–linear piecewise function (Cudeck & Klebe, 2002). For example,

$$f_j = \begin{cases} 
\eta_{i1} + \eta_{i2} t_j + \eta_{i3} t_j^2, & t_j \leq \gamma \\
\eta_{i4} + \eta_{i5} t_j, & t_j > \gamma
\end{cases}$$

(5)

where the first phase of the model corresponds to a quadratic function and the second phase of model corresponds to a linear function. Similar to the restriction imposed on linear–linear piecewise LGC model, two restrictions can be imposed on Equation 5. The first restriction is $\eta_{i1} + \eta_{i2}\gamma + \eta_{i3}\gamma^2 = \eta_{i4} + \eta_{i5}\gamma$, indicating that the two segments join at the knot. With this restriction in place, a parameter in Equation 5 can be eliminated. The second restriction is $\eta_{i2} + 2\eta_{i4}\gamma = \eta_{i5}$, which forces the first derivatives to agree at the knot. As a result of this second restriction, the linear coefficient of the first segment can be expressed in terms of the others.

An assumption inherent to the LGC model is that all individuals are drawn from the same population and share the same functional form of growth. This assumption, which applies to the piecewise LGC model as well, is not practical in situations where the data come from a mixture of two or more unobserved subpopulations (i.e., latent classes). For instance, using a sample of children with Head Start experience, Kreisman (2003) found that children with two or more years of program participation had slower achievement growth, on average, than children with only 1 year of program participation. The analysis of this type of mixture data requires the extension of the LGC model to include a categorical latent class variable. Growth mixture models (GMMs) represent a statistical method that injects latent classes into the LGC model (Muthén, 2001, 2002; Muthén & Muthén, 2000; Muthén & Shedden, 1999). More specifically, GMMs are a kind of multivariate normal mixture model that assumes the continuous observed data $y$ arise from a mixture of two or more latent subpopulation distributions, and where $y$ is assumed to be distributed normally within each latent class. GMMs allow distinctly different growth trajectories in two or more latent classes with the differences between classes represented by differences in the means and variances of growth parameter distributions across classes (Bauer & Curran, 2003). GMMs also allow for the estimation of the proportion of subjects in each latent class. Overall, the combined use of continuous (i.e., intercepts and slope growth factors) and categorical latent variables (i.e., class membership) permits individuals to vary around the mean growth curve for their particular class or subgroup such that each subgroup has its own model parameter distributions (Muthén, 2001; Muthén & Shedden, 1999). A path diagram of a conventional GMM with five equally spaced time points, a linear trajectory, continuous latent variables, and categorical latent class variable $C$, is illustrated in Figure 2.

In applying GMMs it is typically assumed that the functional form describing the overall change process is the same throughout the time periods studied in each latent class. This assumption may be unrealistic in situations where the functional form changes across intervals of time within one or more classes, in which case piecewise functions should be adopted.

### Linear–Linear PGMM

While reviewing the literature from psychology and education, we found that the previous studies that used a piecewise function to describe the underlying change process either assumed a single class and known knot location (e.g., Chou et al., 2004), or multiple classes with known knot locations (Depaoli, 2013). A more flexible model that accommodates multiple classes, piecewise functions, and unknown knot locations would be more appropriate in many applications. Furthermore, in Kohli, Harring, and Hancock (2013) the linear–linear PGMM was applied to real data using the method of maximum likelihood, but the authors neither considered alternative approaches to inference nor conducted a simulation study to compare the performance of multiple approaches. ML inference via an EM algorithm (e.g., as implemented in the Mplus software environment) and MCMC for Bayesian inference (e.g., as implemented in the JAGS software environment) are two common approaches to inference for finite mixture models. The current study helps researchers decide among these alternative approaches to inference.

Under fairly general conditions, both maximum likelihood and Bayes estimators have the same asymptotic distribution (Robert, 2001). However, little is known about the finite-sample performance of the two approaches in the setting of PGMMs. That is, there is little guidance regarding how well the two approaches perform for realistic sample sizes. We conducted this research so that we can make recommendations, and provide code, to practitioners who aim to apply PGMMs.

A brief summary of the linear–linear PGMM is provided here for the readers. Note that the simulation study investigates the performance of linear–linear PGMMs under different manipulated conditions across the two inference methods. For the real data examples, we apply both linear–linear and quadratic–linear PGMMs to two different data sets.
The main idea behind the linear–linear PGMM, a special case of the GMM and the other models discussed above, is to combine the linear–linear piecewise growth trajectory and its own time for transitioning between phases. Figure 3 illustrates a path diagram of a conventional linear–linear PGMM for two classes with unknown knots. To formulate a linear–linear PGMM, consider $n$ observed repeated measures, $j = 1, \ldots, n$, data that come from $K$ subpopulations ($k = 1, \ldots, K$), with a latent categorical variable $C$ indicating latent class membership for individual $i$, where $k$ designates each latent class and indicates that model parameter values, including the knot, may differ across classes. Assuming conditional independence and two classes, the class-specific measurement portion of the model is specified (Kohli et al., 2013) as:

$$y_{ikj} = \begin{cases} \eta_{1ikj} + \eta_{2ikj}t_j + \epsilon_{ijk}, & t_j \leq \gamma_k, \\ \eta_{1ikj} + \eta_{4ikj}t_j + \epsilon_{ijk}, & t_j > \gamma_k \end{cases}, \text{ for } i = 1, \ldots, N;$$

$$j = 1, \ldots, n \text{ and } k = 1, 2. \quad (6)$$

In Equation 6, $y_{ikj}$ is the observed response of individual $i$ at time $j$ belonging to class $k$; $t_j$ represents the time of measurement; $\eta_{1ikj}$ and $\eta_{2ikj}$ represent the intercept and the slope growth factors of the first phase, respectively; $\eta_{3ikj}$ and $\eta_{4ikj}$ represent the intercept and the slope growth factor of the second phase, respectively; $\gamma_k$ represents the location of the knot; and $\epsilon_{ijk}$ is the residual that is often assumed to be normally distributed with mean zero and covariance matrix $\Theta_{ik}$. The subscript “$i$” in $\Theta_{ik}$ indicates that the covariance matrix is subject-dependent, and thus, allows for missing data. It is often assumed that $\Theta_{ik}$ is diagonal, and that the residuals are independent between measurement occasions with constant variance across time (i.e., $\Theta_{ik} = \sigma_{i\epsilon}^2I_k$), although this structure could be replaced by any alternative structures as the data warrant. Note that the subscript $k$ in Equation 6 indicates a separate model for each latent class $k$, thus allowing for heterogeneity within the population. Furthermore, in Equation 6 there is no $i$ subscript for the knot, $\gamma_k$. This is because in this model it is assumed that the unknown location of knot is the same for all subjects within a class, but may be different across classes.

One can see from Equation 6 that there are three linear coefficients and one nonlinear knot (i.e., the knot enters the model nonlinearly), $\beta_{ik} = (\eta_{1ik}, \eta_{2ik}, \eta_{3ik}, \gamma_k)^T$. The intercept of the second phase, $\eta_{3ik}$, is not considered in the target function because it can be eliminated based on the constraint imposed by forcing the two linear segments to join at the knot (i.e., $\eta_{1ik} + \eta_{2ik}t = \eta_{3ik} + \eta_{4ik}t$). The decision as to which coefficient to eliminate is completely arbitrary, yet whatever the choice, the model incorporates this restriction. Between-subjects heterogeneity is accounted for in the class-specific structural component of the model, which is specified as:

$$\begin{pmatrix} \eta_{1ik} \\ \eta_{2ik} \\ \eta_{3ik} \end{pmatrix} = \begin{pmatrix} \alpha_{ik} \\ \alpha_{2ik} \end{pmatrix} + \begin{pmatrix} \zeta_{1ik} \\ \zeta_{2ik} \end{pmatrix}, \quad (7)$$

where $\alpha_k$ is a vector of growth factor means; and $\zeta_{ik}$ is a vector of random disturbances in the latent growth factors, $\eta_{ik}$, and such random turbulence is often assumed to be normally distributed with mean zero and a $3 \times 3$ covariance matrix $\Psi_k$. The class-specific covariance matrix $\Psi_k$ has the following form:

$$\Psi_k = \begin{pmatrix} \sigma_{\eta_{11}} & \sigma_{\eta_{12}} & \sigma_{\eta_{13}} \\ \sigma_{\eta_{21}} & \sigma_{\eta_{22}} & \sigma_{\eta_{23}} \\ \sigma_{\eta_{31}} & \sigma_{\eta_{32}} & \sigma_{\eta_{33}} \end{pmatrix}.$$

There are three underlying assumptions with linear–linear PGMM that are in common with standard factor analytic models: (a) the residuals are assumed to be uncorrelated with the continuous latent growth parameters (i.e., $\text{cov}(\epsilon_{ik}, \eta_{ik}) = 0$); (b) the residuals are also assumed to be uncorrelated with the residuals in the latent factors (i.e., $\text{cov}(\epsilon_{ik}, \zeta_{ik}) = 0$); and (c) the residuals are assumed to be uncorrelated with each other (i.e., $\text{cov}(\epsilon_{ik}, \epsilon_{i'k}) = 0$ for $i \neq i'$).

The purpose of this article is to compare two common approaches to estimation for PGMMs: ML inference via an EM algorithm (as implemented in the Mplus software environment), and MCMC for Bayesian inference (as implemented in the JAGS software environment; Plummer, 2003). We also present two real data examples showing the application of a linear–linear PGMM and a quadratic–linear PGMM, respectively.

**Method**

**ML via EM.** The parameterization of the model in Equation 6 cannot be specified directly in a SEM framework in a way that allows the estimation of $\gamma_k$. This is because the existing SEM software packages, such as Mplus 6.0 (Muthén & Muthén, 1998–2010), do not permit incorporating control statements (e.g., if-then statements) in the model setup. Thus, we need to reparameterize the model in such a way that the piecewise function can be handled in Mplus. Following Harring, Cudeck, and du Toit (2006)’s discussion, the alternative parameterization circumvents this problem by rewriting the function as a polynomial and using the nonlinear constraints feature available in many SEM software packages.

![Figure 3. A path diagram of a conventional linear–linear PGMM with unknown knots.](image-url)
More specifically, as in Harring et al. (2006), the jth row of the factor loading matrix for a linear–linear PGMM can be expressed as:

\[
[A(y)]_j = \omega_{1jk} + \omega_{2jk}t_j - \omega_{3jk}(t_j - \gamma_j)^2. \tag{8}
\]

The original regression parameters in Equation 6 have been reparameterized such that

\[
\omega_{1jk} = \frac{(\eta_{1jk} + \eta_{3jk})}{2} \quad \omega_{2jk} = \frac{(\eta_{2jk} + \eta_{4jk})}{2} \quad \omega_{3jk} = \frac{(\eta_{2jk} - \eta_{4jk})}{2}.
\]

This reparameterized linear–linear PGMM has the same number of parameters as in the original linear–linear PGMM of Equation 6, but fits within the system used by many SEM software packages. Upon convergence, the estimated parameters of the reparameterized PGMM are then transformed back to the original parameters of the linear–linear PGMM. To transform the estimated reparameterized linear coefficients back to the original model’s linear coefficients, a straightforward linear transformation based on the relationships in Equation 9 is used, whereas for transforming the estimated variance parameters, the multivariate delta method (Oehlert, 1992) can be used. Readers are referred to Appendix B of Kohli, Harring, and Hancock (2013) that shows how to transform the estimated parameters obtained from fitting the reparameterized model back to the original regression parameters.

Furthermore, the conditional expected log-likelihood function for the response data vector \(y\) and the vector of latent categorical variable \(c\) with \(K\) classes is specified as,

\[
E[\ln L(\theta \mid y, c) \mid y, \theta^0] = \sum_{i=1}^{M} \sum_{k=1}^{K} E[c_{ik} \mid y, \theta^0] \\
\times [\ln(\pi_i) + \ln[h_{ik}(y_i \mid \theta_0)]]. \tag{10}
\]

where \(\pi_i\) is the prior probability of belonging to the \(k\)th class; \(c_{ik}\) is an unobserved indicator variable whose value is equal to 1 if the \(j\)th subject belongs to the \(k\)th class and 0 otherwise; \(h_{ik}\) is the conditional log-likelihood function; and \(\theta\) is the vector of model parameters, including the mean, and the variance and covariance parameters. Using the ML via EM estimation approach, the E-step of the EM iterative procedure computes the posterior probabilities of belonging to a latent class with respect to \(\theta\) at iteration \(v\). In the M-step of the EM iterative procedure the conditional expected log-likelihood function in Equation 9 is maximized with respect to \(\theta\) to obtain the updated parameter vector \(\theta^{v+1}\).

**MCMC for Bayesian inference.** There is an increasing trend of fitting structural equation models using Bayesian methods. Wang and McArdle (2008), and Depaoli (2013) have explored the application of a Bayesian approach for estimating LGC models and subclasses of LGC models. In practice, the chief difference between the Bayesian approach (Robert, 2001) and the ML approach is that the former requires not only a likelihood \(L(\theta \mid y)\) (where \(\theta\) is used generically to denote parameters and \(y\) denotes data) but also a prior distribution \(p(\theta)\) that expresses our uncertainty regarding the parameter \(\theta\). Inference is then based on the posterior distribution \(p(\theta \mid y) \propto L(\theta \mid y)p(\theta)\). Specifically, a point estimate of \(\theta\) can be obtained as the mean of the posterior distribution (if the squared error risk function is assumed), and an \(\alpha\)-level interval for \(\theta\) can be obtained as the \((1 - \alpha)100\%\) highest posterior density (HPD) interval\(^1\) (also known as a credible interval). In most cases the posterior distribution is unwieldy or not available in closed form, in which case MCMC (Brooks, Gelman, Jones, & Meng, 2011) methods can often be used to estimate \(\pi\) and/or features of \(\pi\).

In MCMC for Bayesian inference, one constructs a Markov chain having \(\pi\) as its stationary distribution, and then simulates many iterations of the chain so that eventually subsequent points along the sample path are distributed approximately according to \(\pi\). The resulting collection of samples can then be used to estimate features of the posterior distribution, for example, the distribution itself, or the posterior means and credible intervals mentioned above. In this study, we plan to use Just Another Gibbs Sampler (JAGS) (Plummer, 2003) for conducting MCMC for Bayesian inference for PGMMs. JAGS is an extensible, cross-platform framework for analyses of Bayesian hierarchical models via MCMC.

In the context of PGMMs, an advantage of using Bayes via JAGS over ML via Mplus is that one can directly fit the model described in Equation 6 in JAGS, whereas in Mplus one has to reparameterize the original model (Equation 8). In an earlier study (Wang and McArdle, 2008) the authors found the Bayesian approach to yield more precise inference than the frequentist approach, including the first-order Taylor series expansion method and the adaptive Gaussian quadrature method, for estimating knots in a single-class piecewise LGC model. However, the authors did not consider the ML-EM approach, or a multiclass (mixture) model. Thus, whether the Bayesian approach outperforms the ML-EM approach for linear–linear PGMMs is an open question. This leads to the aim of the current study, that is, to provide concrete guidance regarding how well the two approaches perform for realistic sample sizes.

**Design of Simulation Study**

To investigate the performance of the two approaches to inference, a Monte Carlo simulation study was conducted. As mentioned earlier, one of the most interesting parameters in this model is the location of the knot. The conditions that were hypothesized to impact the estimation of the knot, along with other model parameters, included: sample size, class mixing proportion, location of the knot, the residual variance of the observed variable, and class separation (defined by Mahalanobis Distance, MD). The rationale for choosing these manipulated conditions is described in a later section. Furthermore, to evaluate parameter recovery, the linear–linear PGMM was fitted to data generated from a population model with true (known) parameters, and parameter estimates were then compared with their true values. Additionally, the effects of manipulated conditions on the percentage of properly converged replications were analyzed (where a properly converged replication is a replication for which the solution converged with

\(^1\) It is the shortest possible interval enclosing \((1 - \alpha)\%\) of the posterior mass.

\(^2\) MD = \(\sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2 + (\gamma_1 - \gamma_2)^2}\)
Data Generation

A linear–linear PGMM with two latent classes was used as a population model to generate repeated measures data conforming to nine equally spaced time points (coded 1 to 9), where the data followed a multivariate normal distribution within individual latent classes. The generated data were both balanced and complete. The R program (R Development Core Team, 2009) was the statistical package used to generate the data sets. The choice of two latent classes for simulation purposes was made to keep the scope of the study manageable. In both methodological and substantive research of piecewise growth models, the number of time points used is often six or more (see, e.g., Cudeck, 1996; Cudeck & Klebe, 2002; Harring et al., 2006), hence, the choice of nine time points seemed reasonable. Furthermore, according to Lubke and Muthén (2007), additional time points did not generally influence model performance or class assignment in the context of linear GMMs. Finally, for the purpose of this study, it was assumed that the intercept and slope growth parameters were uncorrelated, thereby simplifying the data generation model. This assumption is consistent with previous studies (e.g., Hamilton, 2009; Shin et al., 2013).

Study Design

In the simulation design, some conditions were fixed throughout all simulations, whereas other conditions were manipulated. The fixed and manipulated conditions are described below.

Fixed conditions. The primary focus of this research was to compare the performance of the two approaches to inference. The conditions that were assumed not to directly impact estimation of the knot location were fixed across all conditions. These were

1. Number of time points (coded 1 to 9),
2. Number of latent classes ($k = 1, 2$)
3. Covariance matrix of growth factors, that is,

$$
\Psi = \begin{bmatrix}
\eta_{11k} & \eta_{12k} & \eta_{14k} \\
\eta_{21k} & .4 & 0 \\
\eta_{22k} & 0 & .2 \\
\eta_{44k} & 0 & 0 & .2
\end{bmatrix}
$$

Note that the population values for the covariance matrix of growth factors were held constant across the classes. This choice was based on the suggestion made by Muthén (2001) that mixture models with large differences in the factor variances and covariances between classes were particularly sensitive to local maxima. Table 1 summarizes the population values chosen for each of the fixed conditions.

### Table 1
**Fixed and Manipulated Parameter Values**

<table>
<thead>
<tr>
<th>Fixed conditions</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intercept ($\alpha_{11}$)</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>Mean slope 1 ($\alpha_{12}$)</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Mean slope 2 ($\alpha_{14}$)</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Variance of intercept ($\sigma^2_{11}$)</td>
<td>.40</td>
<td>.40</td>
</tr>
<tr>
<td>Variance of slope 1 ($\sigma^2_{12}$)</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Variance of slope 2 ($\sigma^2_{14}$)</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Covariances among the growth factors</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2. Number of time points (1 to 9)</td>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Manipulated conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Knot location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early/late</td>
<td>3.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Middle</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>2. Class mixing proportion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal proportion</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Unequal proportion</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>3. Residual variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>High</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>4. Total sample size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 400$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Class separation (defined by MD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD $= 1.04$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intercept ($\alpha_{22}$)</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>Mean slope 1 ($\alpha_{22}$)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Mean slope 2 ($\alpha_{24}$)</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>MD $= 2.24$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean intercept ($\alpha_{12}$)</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>Mean slope 1 ($\alpha_{12}$)</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Mean slope 2 ($\alpha_{14}$)</td>
<td>1.75</td>
<td></td>
</tr>
</tbody>
</table>

2. Class mixing proportion ($\varphi_k$)

The two levels of class mixing proportion chosen, expressed as percentages, were: (50/50 and 25/75) based on Nylund, Asparouhov, and Muthén (2007). We chose these two levels because we want to see how each of the inference approaches will perform in a more challenging scenario of $\varphi_k = (.25, .75)$, especially when the sample size is small (i.e., when $N = 100$, class 1 $= 25$, and class 2 $= 75$), versus when $\varphi_k = (.50, .50)$.

3. Location of the knot ($\gamma_k$)

The two knot conditions were: early/late locations ($\gamma_k = 3, 6$), and middle locations ($\gamma_k = 4, 5$). The rationale for choosing these two levels was motivated by the fact that we wanted to see how the model performed with respect to the estimation of mean and variance parameters of the respective growth factors when the knot occurred early or late in the study and hence were widely spaced versus when the knot occurred in the middle of the study and hence were closer together.
4. Residual variance of the observed variable (σ²_{ei})

The two levels of this condition were: low residual variance (i.e., σ²_{ei} = .15) and high variance (i.e., σ²_{ei} = .50). The residual variance was varied based on the common knowledge that the amount of residual variance in the observed variable can affect the fitting of the model, thereby affecting estimation of the model parameters, including the location of the knot. In other words, when the amount of residual variance in the observed variable is small (like it would be in learning experiments where systematic change is effectively captured by the function), it should be relatively easy to fit the function, thereby making it easy to estimate the location of the knot in the fitted function. However, when the amount of variance is large (like it would be when tracking young children longitudinally where wave-to-wave variation is substantial), it may be difficult to fit the function, thereby making it relatively difficult to estimate the location of the knot. Hence, this condition is relevant for estimating the location of the knot, irrespective of the choice of inference approach. Note that the residual variance of the observed variable was held equal across classes in the data generation process. Additionally, the population values for the low and high condition for residual variance were selected keeping in mind the values of the intraclass correlation (ICC). To elaborate, the ICC represents the proportion of variance in an outcome variable explained by between subject variability, that is,

\[
\rho_{icc} = \frac{\text{Between subject variability}}{\text{Between subject variability} + \text{Within subject variability}}
\]

which translates to,

\[
\rho_{icc} = \frac{\text{Variance of intercept factor of the first phase}}{\text{Variance of intercept factor of the first phase} + \text{Residual variance}}.
\]

For the population value of variance of η_{ik} = 0.4, the ICC values range from 0.444 to 0.727.

5. Class separation (defined by MD2)

The two levels of class separation condition, a measure of distance (difference) between the two latent classes, were: 1.04 (i.e., small distance), and 2.24 (i.e., large distance) (based on Tolvanen, 2008). We chose MD as a condition because we expect the magnitude of latent distance between the two classes to affect the estimation of model parameters. That is, when the two latent classes are well-separated (i.e., the distance between them is large) it should be relatively easier to fit the model to the data and get better parameter estimates as compared with the scenario where the classes are close to each other (i.e., when the distance between them is small), and thus, not well-separated (Depaoli, 2013). Using the two levels of class separation, along with the levels of the knot location, one can generate the population values of the mean vector for the growth factors. That is,

a. When γ₁ = 3, and γ₂ = 6, respectively, and MD = 1.04:
   • The mean vector for the growth parameters in Class 1, \([a_{11}, a_{22}, a_{41}, γ₁]\) is: [2.00, 0.00, 1.25, 3.00].
   • The mean vector for the growth parameters in Class 2, \([a_{12}, a_{22}, a_{42}, γ₂]\) is: [2.25, 0.35, 1.50, 6].

b. When γ₁ = 3, and γ₂ = 6, respectively, and MD = 2.24:
   • The mean vector for the growth parameters in Class 1, \([a_{11}, a_{22}, a_{41}, γ₁]\) is: [2.00, 0.00, 1.25, 3.00].
   • The mean vector for the growth parameters in Class 2, \([a_{12}, a_{22}, a_{42}, γ₂]\) is: [3.00, 0.50, 1.75, 6].

c. When γ₁ = 4, and γ₂ = 5, respectively, and MD = 1.04:
   • The mean vector for the growth parameters in Class 1, \([a_{11}, a_{22}, a_{41}, γ₁]\) is: [2.00, 0.00, 1.25, 4.00].
   • The mean vector for the growth parameters in Class 2, \([a_{12}, a_{22}, a_{42}, γ₂]\) is: [2.25, 0.35, 1.50, 5].

d. When γ₁ = 4, and γ₂ = 5, respectively, and MD = 2.24:
   • The mean vector for the growth parameters in Class 1, \([a_{11}, a_{22}, a_{41}, γ₁]\) is: [2.00, 0.00, 1.25, 4.00].
   • Mean vector for the growth parameters in Class 2, \([a_{12}, a_{22}, a_{42}, γ₂]\) is: [3.00, 0.50, 1.75, 5].

The combination of manipulated conditions (2 × 2 × 2 × 2) resulted in a Monte Carlo simulation with 32 cells for each estimation procedure. For each condition, 100 replications were generated (Wang & McArdle, 2008).

Fitting the Model

The parameters of the two-class linear–linear PGMM were estimated using Mplus 6.0 and JAGS 3.3.0.

Mplus. The default estimator for mixture analyses using Mplus 6.0 is MLR (maximum likelihood with robust sandwich-type SE estimates) via the EM algorithm. The default number of random starting values used by Mplus 6.0 (i.e., 10 sets of random starting values with two final iterations using the best likelihood values from the original set of 10) was increased to 100 sets of random starting values with five final iterations for each set to investigate local solutions more thoroughly. According to the research conducted by Lubke and Muthén (2007), the complexity of the model with respect to the factor structure, or the number of observed variables within each class, did not influence model performance. However, when estimating mixture models, in general, using ML via EM estimation, failure to converge to a stable solution within a given number of iterations or converging to a local maximum of the likelihood were common problems (Bauer & Curran, 2003; Hipp & Bauer, 2006; Muthén, 2001).

While estimating linear–linear PGMMs with two latent classes under different manipulated conditions in Mplus 6.0, all the model parameters were allowed to be freely estimated across classes; that is, the mean and the variance of the intercept and slope parameters of both phases, the covariances among the growth factors (intercepts and slopes), the location of the knot, and the residual variances.

Just Another Gibbs Sampler. We chose noninformative prior distributions for all parameters, as follows. Note that these prior distributions have been widely used (Congdon, 2001; Fruhwirth-Schnatter, 2006; Wang & McArdle, 2008). If the user possesses reliable prior information about parameters based on sound theory or practitioner experience, more informative prior distributions can be used for those parameters. Informative prior distributions should be used with care; however, informative priors can lead to better performance, but they can also lead to erroneous
inference, especially for smaller datasets (because a prior exerts a stronger influence on the posterior when the sample size is small).

The prior distribution for the fixed effects $\alpha_i$ was spherical normal with mean zero and common variance 1,000,000 (this large variance ensures that posterior inference is insensitive to the prior). The prior distribution for the residual precision $1/\tau^2$ was $\gamma$ with parameters 0.001 and 0.001. The prior distribution for the precision matrix $\Psi_i$ was a Wishart distribution with a $3 \times 3$ identity matrix as the scale matrix and degrees of freedom equal to 3. The prior distribution for the knot location $\gamma_i$ was a uniform distribution in the interval [1, 9]. Finally, the prior distribution for the mixing proportions was Dirichlet with equal concentration parameters for all components.

We allowed JAGS to choose starting values, in which case the initial value for a given parameter is a “typical” value for that parameter’s prior distribution. The meaning of typical depends on the prior distribution, but the initial value is usually the prior’s mean, median, or mode.

To assess convergence of the Markov chain, we used fixed-width output analysis (Flegal, Haran, & Jones, 2008). In fixed-width analysis, one chooses a small threshold and halts the simulation when all Monte Carlo standard errors are no larger than the threshold. In other words, one declares the Markov chain to have converged when the variability of the MCMC estimator is sufficiently small. We computed Monte Carlo $SE$s using the consistent batch means method (Haran & Hughes, 2012), and found that 20,000 samples were generally sufficient to yield $SE$s on the order of 0.001, which indicates a close approximation to the posterior distribution.

**Label Switching**

One challenge that is unique to estimating mixture models is label switching, that is, inconsistent assignment of class labels. Label switching occurs because the likelihood is invariant to permutation of the class labels. For a Bayesian procedure, this leads to marginal posterior distributions that are identical for all mixture components, which renders parameter estimates meaningless (Jasra, Holmes, & Stephens, 2005). We addressed this issue by applying the (re)labeling by normal likelihood (NORMLH) algorithm of Yao and Lindsay (2009) to the posterior samples before computing posterior estimates. The NORMLH algorithm performs well when the correctly labeled posterior samples are approximately normally distributed, which we verified (for the scenarios considered in this article) by conducting a pilot simulation study.

After the parameter estimates were obtained, we applied a post hoc class assignment algorithm developed by Tueller, Drotar, and Lubke (2011). That is, after estimating parameters for a simulated dataset in JAGS, we checked whether the labels were switched using the column maxima switched label detection algorithm described in Tueller et al. (2011). When label switching was detected for a replication, final parameter estimates were repositioned accordingly before computing the outcome measures. We used the same procedure to address label switching for Mplus.

**Outcome Measures**

Upon convergence, the estimated parameters of the reparameterized model obtained from Mplus were transformed back to the original parameters of the linear–linear PGMMs with two latent classes. Because JAGS can fit the original model in Equation 6, and because a pilot simulation study produced comparable results for the two parameterizations, we applied our Bayesian procedure using the original parameterization only.

Strictly speaking, one need not evaluate a Bayesian procedure according to frequentist criteria (Robert, 2001). Frequentist evaluation of Bayesian procedures is not uncommon, however, and some recent work suggests that perhaps one should consider the “long-run” performance of Bayesian procedures (Fraser, 2011). In any case, we judge the performance of our procedure against frequentist standards (e.g., bias, coverage rates) because doing so provides us with a convenient means of comparing ML via EM and Bayesian inference via MCMC.

To evaluate the performance of the two-class linear–linear PGMMs under different manipulated conditions, the following outcome measures were used: coverage rate, average parameter bias, and root mean square error (RMSE).

In a frequentist setting, one desires that $(1 - \alpha)100$ of 100 confidence intervals for a given parameter contain the true value of that parameter, that is, one desires a coverage rate of $(1 - \alpha)100\%$. In a Bayesian setting, the credible interval is analogous to the confidence interval, but a credible interval has a different interpretation. Specifically, a $(1 - \alpha)100\%$ credible interval contains the true value of the parameter with probability $1 - \alpha$. Strictly speaking, these are incommensurable points of view, and so, in the interest of comparing the ML and Bayesian procedures, we judged our credible intervals according to the frequentist criterion, that is, we estimated the coverage rates of our credible intervals.

Average parameter bias was defined as the average difference between the $i$th estimated parameter value and the corresponding population true value, that is,

$$\text{Average Bias} : \frac{\sum_{i=1}^{N^*} (\hat{\theta}_i - \theta_{true})}{N^*}, \text{ } N^*: \text{number of converged replications.} \quad (11)$$

Note that average bias was computed only for successful replications (i.e., a replication for which the solution converged with no parameter estimates outside the possible range for the parameter) in each cell. In our experience, researchers are likely to reject any unconverged solution, and so we computed bias only for replications likely to be reported by a researcher. Additionally, RMSE corresponding to each of the estimated parameters for each replication in each cell was computed as:

$$\text{RMSE}: \sqrt{\frac{\sum_{i=1}^{N^*} (\hat{\theta}_i - \theta_{true})^2}{N^*}}, \text{ } N^*: \text{number of converged replications.} \quad (12)$$
The following section describes the results obtained from the simulation study.

**Results of Simulation Study**

**Model Convergence Rate**

In the context of the Bayesian procedure we use the term “converge” in two ways. Some datasets cause JAGS to encounter numerical difficulties at the outset, in which case the sampling run terminates. This is what we mean when we say that convergence failed for a given dataset. If the computation remains stable, JAGS will return a sample path from the Markov chain. That sample path can then be used to assess convergence to the chain’s stationary distribution (in a Bayesian setting, the posterior distribution). In our studies, all sample paths used for inference yielded Monte Carlo SEs on the order of 0.001. In this sense, all of our chains converged.

Table 2 shows the model convergence rate across the different manipulated conditions for the two estimation approaches. The rate of converged replications for the two-class linear-linear PGMMs across all the conditions (total number of simulation conditions were 32) was found to be between 48% and 91% for the ML approach, and 10% to 100% for the Bayesian approach. Although the lower bound of the range for the Bayesian approach is much smaller than the lower bound of the range for ML approach, the Bayesian procedure overall succeeded much more often with regards to the model convergence rate. To elaborate, for the Bayesian approach out of a total of 32 conditions, 17 conditions reported 100% convergence rate, and 11 conditions reported convergence rate of 85% to 99%. The same is not true for the ML approach. None of the conditions under ML approach had 100% convergence rate. Out of a total of 32 conditions, only 4 conditions had convergence rate of 85% to 99%, 21 conditions had convergence rate between 70% to 84%, and 7 conditions had convergence rate less than 70%. Thus, it is evident that in general the Bayesian inference did a much better job with regards to model convergence than the ML inference.

Additionally, we further examined the four conditions for which the Bayesian procedure reported poor convergence rates and found that failure can be attributed to small sample size ($N/H_{1100}^{100}$), high residual variance ($\epsilon_{2}/H_{9268}^{1.50}$), and the knot locations for the two classes being close to each other (i.e., $\gamma_{1} = 4$, and $\gamma_{2} = 5$). Furthermore, it seems on the surface that the convergence rates for these four conditions were high for the ML approach, but upon closer inspection we found that it did not necessarily translate to better parameter estimation. That is, the bias and RMSE for these two conditions under the ML approach were unusually large.

<table>
<thead>
<tr>
<th>Mixing proportions</th>
<th>Sample size</th>
<th>Residual variance</th>
<th>Knot locations</th>
<th>MD</th>
<th>ML</th>
<th>Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>77%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>88%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>82%</td>
<td>85%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>73%</td>
<td>98%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>86%</td>
<td>97%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>72%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>91%</td>
<td>16%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>100</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>91%</td>
<td>62%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>74%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>84%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>72%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>81%</td>
<td>100%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>76%</td>
<td>99%</td>
</tr>
<tr>
<td>0.50-0.50</td>
<td>400</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>66%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>74%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>77%</td>
<td>88%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>76%</td>
<td>99%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>76%</td>
<td>93%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>53%</td>
<td>94%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>48%</td>
<td>98%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>76%</td>
<td>10%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>100</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>73%</td>
<td>50%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>68%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.15</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>73%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.15</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>64%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>1.04</td>
<td>71%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.50</td>
<td>[3, 6]</td>
<td>2.24</td>
<td>67%</td>
<td>100%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>1.04</td>
<td>70%</td>
<td>93%</td>
</tr>
<tr>
<td>0.25-0.75</td>
<td>400</td>
<td>0.50</td>
<td>[4, 5]</td>
<td>2.24</td>
<td>63%</td>
<td>98%</td>
</tr>
</tbody>
</table>
indicating poor estimation of model parameters. Overall, the Bayesian procedure provided a higher convergence rate on average than the ML procedure.

**Coverage Rate**

Table 3 shows the average coverage rates. Among the 16 parameters in the model, Bayesian procedure for the most part provided coverage rates close to the desired 95%. ML procedure, on the other hand, did not have overall good coverage rates across all the model parameters. For instance, using ML procedure the average coverage rates for the estimates of variance parameters for slope in the first and the second phase, respectively, were only 63.9% to 79% (for Class 1 and Class 2, respectively), and 81.2% to 70.3% (for Class 1 and Class 2, respectively). In summary, we can conclude that the average coverage rates for the ML procedure were outdone by the Bayesian approach.

**Average Parameter Bias**

Table 4 presents the average bias for each parameter across all conditions for both estimation procedures. As shown in Table 4, the size of the average bias for the Bayesian inference approach was within the range of 0.002 to 0.053, and for the ML inference approach the range was within 0.001 to 0.057, indicating that both procedures yielded parameter estimates that were close to the true values. Thus, there were no noticeable differences between the two estimation procedures with respect to average parameter bias.

**Root Mean Square Error**

The average RMSE for each parameter across all conditions for both estimation procedures is shown in Table 5. Based on the examination of Table 5 it seems that the size of the average RMSE is considerably smaller for the estimates obtained from the Bayesian approach. The average RMSE for the Bayesian approach is within the range of 0.042 to 0.179 whereas the range of average RMSE for the ML approach is 0.037 to 0.382. In addition, the Bayesian approach seems to estimate the variance–covariance matrix of growth factors much more precisely than the ML approach. Thus, the Bayesian approach did a better job with regard to the precision of parameter estimates as compared with the ML approach for most of the conditions examined in this study. Overall, we can conclude that MCMC for Bayesian inference outperformed ML via EM on the outcome measures—coverage rate, coverage, and RMSE.

**Real Data Examples of PGMMs**

**Application of Linear–Linear PGMM: Analysis of ECLS-K Data**

To demonstrate how researchers can fit linear–linear PGMM in JAGS (we have only provided results for the Bayesian approach because it performed better than the ML via EM estimation approach), a random subsample (n = 400) was extracted from the data on mathematics achievement scores (Math IRT Scaled Scores) from the Early Childhood Longitudinal Study–Kindergarten Cohort (ECLS-K; total n = 15,305). ECLS-K is a nationally representative longitudinal cohort sample of U.S. students who entered kindergarten in the fall of 1998. In the ECLS-K study, data were collected during the fall and spring of kindergarten and first grade, respectively, and spring of 3rd, 5th, and 8th grades, respectively (time points coded as: 0, 0.5, 1, 1.5, 3.5, 5.5, and 8.5, respectively) and included direct assessment of child functioning, parent interviews, teacher questionnaires, and administrator surveys. In the original data set, 51% of students were male, and 49% of students were female. Furthermore, about 57% of students were White, 13% were Black, 18% were Hispanics, 7% were Asian, and 5% were other race. One of the interesting research questions to investigate is whether there are more than one subpopulation (subgroup) of students in the data set, and whether there are differences in the development of mathematics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>0.990</td>
<td>0.994</td>
<td>0.945</td>
<td>0.933</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>0.866</td>
<td>0.943</td>
<td>0.956</td>
<td>0.941</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>0.940</td>
<td>0.923</td>
<td>0.935</td>
<td>0.939</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.827</td>
<td>0.863</td>
<td>0.794</td>
<td>0.937</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.926</td>
<td>0.960</td>
<td>0.962</td>
<td>0.913</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.639</td>
<td>0.790</td>
<td>0.969</td>
<td>0.966</td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>0.812</td>
<td>0.703</td>
<td>0.966</td>
<td>0.962</td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>0.859</td>
<td>0.896</td>
<td>0.686</td>
<td>0.894</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>0.057</td>
<td>0.001</td>
<td>0.049</td>
<td>-0.005</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-0.034</td>
<td>-0.027</td>
<td>-0.111</td>
<td>-0.010</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>0.025</td>
<td>0.005</td>
<td>0.022</td>
<td>0.002</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.021</td>
<td>-0.070</td>
<td>0.027</td>
<td>-0.005</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>-0.022</td>
<td>0.004</td>
<td>0.052</td>
<td>0.052</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.008</td>
<td>-0.007</td>
<td>0.053</td>
<td>0.011</td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>-0.014</td>
<td>0.014</td>
<td>0.026</td>
<td>0.016</td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>0.008</td>
<td>0.004</td>
<td>0.029</td>
<td>0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>0.299</td>
<td>0.174</td>
<td>0.179</td>
<td>0.116</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>0.259</td>
<td>0.149</td>
<td>0.105</td>
<td>0.063</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>0.127</td>
<td>0.133</td>
<td>0.077</td>
<td>0.065</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.382</td>
<td>0.319</td>
<td>0.144</td>
<td>0.091</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>0.373</td>
<td>0.209</td>
<td>0.166</td>
<td>0.165</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.157</td>
<td>0.063</td>
<td>0.084</td>
<td>0.035</td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>0.058</td>
<td>0.223</td>
<td>0.053</td>
<td>0.042</td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>0.052</td>
<td>0.037</td>
<td>0.057</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table 3

**Average Model Coverage Rate Across All Conditions**

Table 4

**Average Parameter Bias Across All Conditions**

Table 5

**Average RMSE Across All Conditions**

Table 6
achievement trajectories over time across the different subgroups of students.

Results—Bayesian Inference

We fitted three models to the ECLS-K data: a single-class piecewise LGC model and two- and three-class PGMMs. The priors used for the analysis were similar to the ones used in the simulation study. In all three cases the Markov chain converged in less than 20,000 iterations (Monte Carlo SEs on the order of 0.001). We used the Bayesian information criterion (BIC; Schwarz, 1978) for model selection. In general BIC has demonstrated great promise in the SEM literature as a number of researchers have explored its use, both for different types of latent variable models under varying conditions of model complexity and misspecification (see, e.g., Cudeck & Browne, 1983; Steiger & Lind, 1980), and to other conventional fit measures (see, e.g., Haughton, Oud, & Jansen, 1997). Among alternative model selection criteria for GMMs, such as AIC and Adjusted AIC, BIC excelled at choosing the model closest to the data-generating model without overfitting (Nylund et al., 2007). BIC penalizes overparameterized models and adjusts for sample size. Additionally, there is a large body of literature on Bayesian model comparison, yet BIC remains compelling because it is intuitive and easy to compute and has attractive theoretical properties (that arise from the fact that BIC is approximately equal to the logarithm of the Bayes factor, i.e., the posterior odds of the null hypothesis when the prior probability of the null is one half; Kass & Raftery, 1995). The three fits led to BIC values of 16,909, 16,723, and 16,755, respectively. Because a smaller value of BIC indicates a “better” model, and because a change in BIC larger than 10 indicates strong evidence for the model with lower BIC, these fits led unequivocally to selection of the two-class PGMM.

The results for the two-class fit are given in Table 6 and Figure 4. Table 6 shows the parameter estimates. Figure 4 shows the data along with plots of the fitted trajectories for the two classes. The plot show eminently plausible fit and compelling differences between class growth trajectories.

On average, students in Class 2 have higher levels of achievement early on. Class 2 also had a larger average slope for the first interval. However, the two groups differed little in their mean slopes after the knots. For both classes, growth slowed substantially after the knot, but the transition to the slower growth rate occurred earlier for Class 2 than Class 1, as indicated by their knot location estimates. As can be seen in Figure 4, the average trajectory for Class 2 begins at a higher achievement level, increases more rapidly over the first phase of growth, and remains higher than that for Class 1 over the remainder of the time periods. Our results suggest that the gap between the classes may widen over time, although the widening is slight for the time points included in the study. For Class 2, there is substantially more within-class variation in the intercepts. The residual variance is a bit larger in Class 2. Last, the estimated proportion of individuals in Class 1 was 48% and in Class 2 was 52%.

Application of Quadratic–Linear PGMM: Analysis of Quantitative Measures for a Task Completion Experiment

To demonstrate how researchers can fit quadratic–linear PGMM in JAGS, Real Data (n = 228) from a learning experiment were used. One of the goals of the study was to measure quantitative skill acquisition from the performance on a procedural task over the span of 12 blocks of 32 trials each (a total of 384 trials). For the experiment, participants were asked to learn a set of declarative

---

Table 6: Parameter Estimates of Linear–Linear PGMM of Random Subsample From ECLS-K Data

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>CLASS 1</th>
<th>CLASS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$</td>
<td>23.02</td>
<td>32.10</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>18.35</td>
<td>24.62</td>
</tr>
<tr>
<td>$a_{21}$</td>
<td>5.59</td>
<td>5.47</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>4.57</td>
<td>4.04</td>
</tr>
<tr>
<td>$a_{31}$</td>
<td>1.93</td>
<td>3.18</td>
</tr>
<tr>
<td>$a_{32}$</td>
<td>2.04</td>
<td>1.65</td>
</tr>
<tr>
<td>$a_{41}$</td>
<td>1.63</td>
<td>2.87</td>
</tr>
<tr>
<td>$a_{42}$</td>
<td>2.61</td>
<td></td>
</tr>
</tbody>
</table>
guidelines for evaluating attributes of visual stimuli presented in a series. The researcher then aggregated response time (i.e., median time to response) and accuracy score (i.e., average score) across trials within each block. As evident in many learning experiments, the response time of participants during the early trials is often followed by a leveling off in response time during later trials. In addition, the decline in response time in the initial trials is a bit more rounded (and not so abrupt) as the decline in the later trials, and so a quadratic–linear piecewise function seems appropriate for modeling the change in response time over trials.

Results—Bayesian Inference

Similar to the last example, we fitted three models to the Task Completion Experiment data: a single-class quadratic–linear piecewise LGC model and two- and three-class quadratic–linear PGMMs. The BIC values for the three models were 4,837, 4,219, and 4,233, respectively. Based on these BIC values, the two-class PGMM was found to be the best fitting model out of the three. The results for the two-class fit are given in Table 7 and Figure 5. Table 7 shows the parameter estimates. Figure 5 shows the data along with plots of the fitted trajectories for the two classes. The plot shows clear differences between class growth trajectories. On average, the location of knot, indicative of the transition from rapid quantitative skills acquisition phase to the leveling off phase, for participants in Class 1 (trial block, $\gamma_1 = 3.58$) occurs later than for participants in Class 2 (trial block, $\gamma_2 = 2.93$). Furthermore, on average, for Class 1 the change in the quadratic slope in the first segment for each trial block (quadratic coefficient is defined as, “half the amount by which the linear slope is expected to change per unit of time,” Preacher & Hancock, in press) was 0.88 score points (on the raw scale for response time) until the middle of trial block three and four, and thereafter, the average linear decline in response time of participants is at the rate of 0.39 score points for each trial block. Similarly, on average, for Class 2 participants the change in the quadratic slope in the first segment was 1.34 score points until close to trial block three, and thereafter, the average linear decline in response time was 0.19 score points for each trial block. For Class 1, there is substantially more variation around the intercept growth factor ($\sigma_{i0|1} = 2.76$) than in Class 2 ($\sigma_{i0|2} = 1.75$). Class 1 also has more within-subject variation ($\sigma_{i1} = 1.26$) as compared with Class 2 ($\sigma_{i2} = 0.59$). Last, the estimated proportion of individuals in Class 1 was 46.5% and in Class 2 was 53.5%.

Figure 5. Two-class quadratic–linear PGMM fit to Task Completion Experiment Data: For a given time point, the data for that time point are plotted as black dots (Class 1) or gray dots (Class 2). The dots are staggered to the left and right of the time point so that the data for the two classes can be distinguished easily. The solid lines show the fitted trajectories for Class 1 and Class 2.

Table 7

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{i1}$</td>
<td>13.87</td>
<td>9.18</td>
</tr>
<tr>
<td>$\alpha_{i2}$</td>
<td>0.88</td>
<td>1.34</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.39</td>
<td>-0.19</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>3.58</td>
<td>2.93</td>
</tr>
<tr>
<td>$\sigma_{i0</td>
<td>1}$</td>
<td>2.76</td>
</tr>
<tr>
<td>$\sigma_{i0</td>
<td>2}$</td>
<td>0.90</td>
</tr>
<tr>
<td>$\sigma_{i1}$</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>$\sigma_{i2}$</td>
<td>1.26</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Discussion

The linear–linear PGMM is a novel and flexible model useful for applications in psychological, educational, and developmental research. As described in Kohli et al. (2013, p. 950),

The advantage of using the piecewise linear–linear LGM for studying the pattern of change is that it allows researchers to specify each developmental phase to conform to a particular form of the overall change process within each latent class. Moreover, the coefficients obtained from fitting this model have meaningful interpretation in the context of the research questions of a study. This is not the case when interpreting the coefficients obtained from fitting polynomial functions, such as the quadratic and cubic.

Despite the difficulty of interpreting coefficients in polynomial models, not all within-phase growth curves will be linear as shown in one of the real data examples—Task Completion Experiment data.

There are two notable features of a PGMM model. First, it posits that growth proceeds in distinct phases separated by knots. Phases of growth or decline have been observed in various domains, such as the development of cognitive skills (Bradway & Thompson, 1962; Hall, Lipton, Sliwinski, & Stewart, 2000; Hall et al., 2001), and the onset of alcohol abuse (e.g., Li et al., 2001; McCrady et al., 2009) The second feature of the model is distinct classes that vary in their initial starting values, rates of growth within a phase, and...
the timing of their transitions between phases. Qualitatively distinct classes are likely associated with qualitatively distinct causes. For instance, one can imagine two different therapy types (drug vs. behavioral) that vary in the timing of rapid improvement onset, rates of improvement within phases, and seriousness of symptoms before the intervention. As another example, language spoken at home (English vs. non-English) may influence the timing of development in phonemic awareness of English language phonemes. Characteristics such as gender may influence the timing and rates of development in different developmental phases. In some cases, the qualitatively distinct causes may correspond to variables that are known and observable (gender, home language), but in others the classes may not be directly observable or they may be observable, but unknown in which case the underlying classes are latent classes that must be uncovered through data analysis. In some cases, the latent classes may overlap with observable groups but not be identical to those groups.

In fitting a growth mixture model, the researcher faces several issues. First, in many cases the number of latent classes is likely unknown and the researcher may want to fit models with differing numbers of classes as we did with the ECLS-K data. The final model can be selected based on fit. We used the BIC because there is large body of literature showing that BIC is a sensitive and high-performing index for latent variables with varying levels of model complexity, but other fit statistics are available (e.g., AIC, Adjusted AIC). Second, a researcher must select a sample size. With our data, neither algorithm performed well with 100 respondents with respect to convergence rate or estimation of model parameters. Given our results, samples as small as 100 cannot be recommended unless, the researcher has some confidence that the residual error is small. Informative priors, when theoretically justifiable, may provide more accurate estimates with smaller samples, but an incorrect choice of prior can bias results. Note in our simulation study, we intentionally chose an extremely uninformative prior (with the variance of the prior equal to 1,000,000) to demonstrate that the Bayesian inference is still valid without the aid from the prior. Researchers, however, can choose less stringent uninformative priors in their data analysis. Additionally, future research may consider replacing the Wishart prior, which has well-understood weaknesses (Barnard, McCulloch, & Meng, 2000). There are many other alternatives available (e.g., separate γ priors, or nonconjugate priors such as truncated normal priors, or separate uniform priors for each diagonal element in the precision matrix). Third, one must sample time points so as to bracket the transition point(s) and so as to provide sufficient numbers of time points within each phase to permit estimation of the coefficients within the phase. Here we were able to adequately recover means and variances of linear parameters with as few as three time points in an interval, but more complex models with more parameters within an interval, will likely require more time points within an interval. If the knots for the latent classes are at similar time points, estimation problems may arise.

In summary, the current study compared two common approaches to statistical inference—maximum likelihood via an EM algorithm and MCMC for Bayesian inference—for finite mixture models under a variety of manipulated conditions that were thought to be important to substantive researchers. Additionally, we provided two real data examples that illustrated how the practitioners can utilize linear–linear PGMM, and quadratic–linear PGMM. Nonetheless, as with any research study, the current study has some limitations. For example, in the real data examples of PGMMs we only used BIC as a model selection criterion when comparing the alternative models (i.e., one-, two-, and three-class PGMMs), or in the design of the simulation study we only considered 100 replications per condition.

PGMMs are a useful class of models that allow researchers to examine the patterns of change over time where the growth trajectory comprises of distinct phases of development and thereby allowing researchers to address interesting substantive hypothesis. In this article we found, based on the results from the Monte Carlo simulation study, that in the most challenging conditions—small sample size (n = 100), large residual variance, and close knot location across the two classes—neither approach performed well with respect to convergence rate and average parameter bias. In the other conditions, however, we found that MCMC for Bayesian inference performs better than ML via EM for linear–linear PGMMs. Hence, the Bayesian procedure should be preferred in practice when the sample is small to moderate in size, residual variance is small to moderate, and the knot location for the two classes are well separated. To facilitate adoption of the Bayesian approach by practitioners, we have provided our JAGS code in the appendix. The Mplus code is also provided in the Appendix as a reference.

We chose to focus our attention on the linear–linear PGMM with one fixed knot per class because this model is intuitive as well as useful. Certain generalizations of the linear–linear PGMM may also prove useful to practitioners; our JAGS code can be extended or enhanced to accommodate, for example, multiple knots per class, random knots, or more complicated piecewise functions.

References


Cudeck, R., & Harring, J. R. (2010). Developing a random coefficient


Steiger, J. H., & Lind, J. C. (May, 1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society; Iowa City, IA.

Appendix

Model Estimation Code

JAGS Code for Fitting a Linear–Linear PGMM With Unknown Knots

```jags
# model specification
model {
  # Iterate over the n subjects.
  for (i in 1:n) {
    # Iterate over the m observation time points to define the
    # model for subject i at each time. The response has
    # mean mu[i, t] and precision tau[k]. where k is the
    # class of subject i.
    for (t in 1:m) {
      y[i, t] dnorm(mu[i, t], tau[class[i]])
      mu[i, t] <- beta[i, 1] + beta[i, 2] * min(t, knot[class[i]]) + beta[i, 3] * max(0, t - knot[class[i]])
    }
    # Now, beta[i, ] is multinormal with mean alpha[k, ] and
    # precision matrix Omega[k, , ].
    beta[i, 1:3] dmnorm(alpha[class[i], 1:3], Omega[class[i], 1:3, 1:3])
    # Put a categorical prior on the K classes. The
    # probabilities are pi[1:K]
    class[i] dcat(pi[1:K])
  }
  # Put a Dirichlet prior on pi. gamma is a vector of K ls.
  pi[1:K] ddirch(gamma)
  # Now set the rest of the priors.
}
```

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
for (k in 1:K)
{
  tau[k] ~ dgamma(0.001, 0.001)
  sigma[k] <- 1/sqrt(tau[k])

  # The prior for the knots is uniform on 1:m.
  knot[k] ~ dunif(1, m)
  # The prior for alpha[k, ] is zero-mean, spherical
  # normal with prior variance 1,000,000.
  Alpha[k, 1:3] ~ dmnorm(alpha0, S)
  # The prior for Omega[k, ] is Wishart with Omega0 = I.
  Omega[k, 1:3, 1:3] ~ dwish(Omega0, 3)
}

# data definitions
n <- 400
m <- 9
K <- 2
alpha0 <- rep(0, 3)
Omega0 <- structure(c(1,0,0,0,1,0,0,0,1), .Dim = c(3, 3))
S <- structure(c(0.000001,0,0,0,0.000001,0,0,0,0.000001), .Dim = c(3, 3))

# Note that fitting of the quadratic-linear PGMM only requires the following change to the above
# JAGS code.
mu[i, t] <- beta[i, 1] + beta[i, 2]*times[t] + beta[i, 3]*min(0, times[t] - knot[class[i]])^2

Mplus Input Code for Fitting Linear-Linear PGMM Unknown Knots

TITLE: Linear-Linear PGMM;
DATA: FILE IS math.dat;
VARIABLE: NAMES ARE t1-t7;
  CLASSES = c(2);
ANALYSIS: TYPE = MIXTURE;
  ALGORITHM = EM;
  STARTS = 500 20;
  STITERATIONS = 50;
  ITERATIONS = 5000;
  SDITERATIONS = 250;
  MITERATIONS = 1000;

MODEL:

%OVERALL%
  w1 BY t1-t7@1;
  w2 BY t1@0 t2@0.5 t3@1 t4@1.5 t5@3.5 t6@5.5 t7@8.5;
  w3 BY t1*(p1);
  w3 BY t2-t7*(p2-p7);
  w1*(v1);
  w2*(v2);
  w3*(v3);
  w1 WITH w2@0;
  w1 WITH w3@0;
  w2 WITH w3@0;
  [t1-t7@0];
  t1-t7*1(i);

(Appendix continues)
MODEL CONSTRAINT:
NEW(gam1*3.6, gam2*3.6);

v1 > 0;
v2 > 0;
v3 > 0;
v4 > 0;
v5 > 0;
v6 > 0:
p1 = -(sqrt((0-gam1)^2));
p2 = -(sqrt((0.5-gam1)^2));
p3 = -(sqrt((1-gam1)^2));
p4 = -(sqrt((1.5-gam1)^2));
p5 = -(sqrt((3.5-gam1)^2));
p6 = -(sqrt((5.5-gam1)^2));
p7 = -(sqrt((8.5-gam1)^2));
q1 = -(sqrt((0-gam2)^2));
q2 = -(sqrt((0.5-gam2)^2));
q3 = -(sqrt((1-gam2)^2));
q4 = -(sqrt((1.5-gam2)^2));
q5 = -(sqrt((3.5-gam2)^2));
q6 = -(sqrt((5.5-gam2)^2));
q7 = -(sqrt((8.5-gam2)^2));

OUTPUT: TECH1 TECH8;
SAVEDATA:
FILE = 'math_genCM.dat';
FORMAT IS F8.2;
SAVE = CPROBABILITIES;