DETECTING UNKNOWN TURNING POINTS USING PIECEWISE GROWTH MIXTURE MODELS: A COMPARISON OF ML/EM VERSUS BAYESIAN ESTIMATION

A Dissertation

by

LING NING

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wen Luo
Committee Members,	Victor Willson
	Oiman Kwok
	Myeungsun Yoon
Head of Department,	Victor Willson

August 2015

Major Subject: Educational Psychology

ABSTRACT

Piecewise growth mixture modeling (PGMM) can be used to investigate growth and change of subpopulations consisting of distinct developmental phases (Muthén, 2008). The major difficulty in specifying a PGMM is how to optimally locate a turning point (or transition point, or knot). Recently, Kohli, Harring, and Hancock (2013) proposed a version of a two-stage (or two-piece) PGMM that allows free estimation of a turning point. The procedure offers more advantages over the practice of determining a turning point a priori. Yet, many questions regarding the performance of the procedure remain to be answered. The dissertation conducted comprehensive Monte Carlo simulation studies to investigate and compare the performance of the proposed procedure under two dominant estimation methods: Bayesian estimation framework via the Markov Chain Monte Carlo (MCMC) algorithm and the Maximum likelihood estimation via the expectation maximization (EM) algorithm (ML/EM).

The dissertation consisted of two studies. Study One used a two-piece PGMM as the population model to generate data and compared the performance of a PGMM with unknown turning points under both estimation methods with regard to parameter recovery, and individual classification accuracy in different experimental conditions. It was found that individual classification accuracy under both estimation methods varied depending on the separation of the latent classes and the number of time points. Overall, the Bayesian method with informative priors gave a more accurate individual classification than the same method with other prior specifications and the ML/EM method. As for the recovery of true parameter, with exception to the Bayesian method

with informative priors, which was found to be able to recover the fixed effects in both classes to an acceptable degree, the same method with other prior specifications and the ML/EM method had difficulty in recovering the fixed effects, particularly, the second slope of class two. Both estimation methods were also found to have difficulty in estimating random effects.

Study Two explored and compared the accuracy of commonly used enumeration indexes with PGMMs in identifying the correct number of latent classes under both Bayesian and ML/EM estimation methods. The findings showed that, for the ML/EM method, the ICL-BIC was the most useful index in correct model identification across all the data conditions under examination. For the Bayesian method, the WAIC2, DIC3,and BIC could be used with confidence in class enumeration with PGMMs.

The findings can help applied researchers determine under what conditions a PGMM performs better fitted by one estimator than the other. The findings also provide reasonable ways for using model fit statistics in combination to decide an optimal PGMM analyzed by either the ML/EM or the Bayesian method.

ACKNOWLEDGEMENTS

I am indebted to many people who have helped make this dissertation happen. But first of all, I would like to acknowledge my academic advisor, Dr. Wen Luo, for all her time, support, patience, and research mentorship. Your integrity and discipline has won my deepest respect and admiration. You have been an exemplary role model to me in my academic journey. I thank you for introducing me into the amazing world of statistics and for your professional guidance and supervision, which has laid a solid foundation for my future growth. I am grateful for the opportunities to teach, research, publish, attend workshops, and to travel to national and international conferences. Thank you for always being supportive during the past five years, especially through my ups and downs. Without you, this dissertation would not have been possible. I am truly blessed to have you in my life.

I would also like to thank my committee members, Dr. Willson, Dr. Kwok, and Dr.Yoon, for their guidance and support throughout the course of this research. I am especially grateful to my committee members and Dr. Hall for being supportive and accommodating my family situation.

My heartfelt thanks also go to Dr. Bo Zhang, Dr. Cindy Walker, and Dr. Razia Azen at University of Wisconsin-Milwaukee, for their support, encouragement, and guidance during my study there.

I also want to thank Dr. Lili Zhao at University of Michigan at Ann Arbor for checking my Bayesian models and the R codes for the dissertation. Thanks also go to

my colleagues and the department faculty and staff for making my time in Aggieland pleasant and memorable.

Special thanks are also due to the sisterhood and brotherhood at the Grace Valley Chinese Church. Your encouragement and prayers have kept my faith strong and have spurred me “on toward love and good deeds” (Hebrews 10:24).

I would like to express my sincere thankfulness to my dear mother and my deceased father, who passed away this May, for their devoted love and selfless sacrifices for me and my two brothers. Many thanks to my brothers for their support and for being truly brothers when needed.

Last but not least, I would like to express my deepest gratitude to my husband, for his love, support, and patience over the last five years. Thank you, my cute little son, for being a well-behaved child while I was in school.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Change is a constant. Measuring individual development change over time has long fascinated empirical researchers. Since the 1980s, researchers have been able to study change well with appropriate statistical models, one of which is known as the Latent Growth Model (LGM) in the Structural Equation Modeling (SEM) framework (Preacher, 2008; Singer & Willett, 2003). Up to date, the majority of applications of the models in longitudinal data analysis have been limited to the assumption that change follows a simple linear trend. However, when longitudinal data are collected over a long period of time, they often do not follow a linear trend, and a simple smooth polynomial functional form may not adequately describe the data that consist of different growth phases. A more flexible approach to model the nonlinear form of growth is the Piecewise Latent Growth Curve Model (PLGCM). This approach breaks up the curvilinear growth trend into separate linear segments or pieces of different slopes, which are again tied together by knots (or turning points or change points). The flexibility of PLGCM allows the formulation of different functional forms for the different phases of growth such that each phase does not have to conform to the same function (Harring, Cudeck, & du Toit, 2006; Khoo, West, Wu, & Kwok, 2006; Kohli et al., 2013; Kwok, Luo, & West, 2010). The approach is particularly appealing when researchers are interested in comparing growth rates for two or more periods based on a substantial interest, such as a marked effect of schooling on children's scholastic attainments before and after secondary school (Chou, Yang, Pentz, & Hser, 2004; Rutter, 1996).

One of the assumptions underlying PLGCMs is that the population being studied is homogeneous in terms of the growth process, and all individuals in the data experience the same differential growth process tied together by the same turning points. However, if the data consist of heterogeneous growth processes, the homogeneous assumption may be concealing important features of inter-individual and intra-individual variations, which produces modeling results that are unreliable (Lu & Huang, 2014).

A more realistic alternative is the Piecewise Growth Mixture Modeling (PGMM) technique. It groups individuals into a finite number of growth trajectories of different growth processes based on observed data patterns. Compared to conventional latent growth curve models, the technique is more flexible: it can not only detect distinctive turning points at which differential growth rates manifest and explore the heterogeneity in the population's growth trajectories, but also examine latent trajectory-specific variance components (Li, Duncan, Duncan, & Hops, 2001; Muthén, 2001). The flexibility and dynamism of the method has attracted enthusiastic applications in social and behavioral sciences (Galatzer-Levy, Bonanno, & Mancini, 2010; Heybroek, 2011; Johnsson, Leifman, & Berglund, 2008; Li et al., 2001; Mora et al., 2009; Uher et al., 2010; Wu, Zumbo, & Siegel, 2011).

The major difficulty in specifying a PGMM concerns an optimal identification of a turning point, designated by a fixed or free time point where the growth rate changes from one linear slope to the other. A brief literature review showed that empirical studies using PGMMs relied exclusively on theoretical considerations to specify a priori a turning point. However, such situations are not often attained.

Recently, a statistical procedure proposed by Kohli, Harring, and Hancock (2013) allows free estimation of a turning point using a two-piece PGMM. Using statistical procedures to detect turning points has several advantages. First, it does not limit the trajectory shape to the substantive hypothesis that two developmental phases are connected by a priori determined fixed time point. Furthermore, allowing free estimation of turning points and time specific factor loadings can sometimes entail a more optimal functional form in the description of observed data patterns over time (Kwok, Luo, & West, 2010; Wood & Jackson, 2013).

Despite the advantages of the newly proposed PGMM procedure, many issues have not yet been resolved. One of the issues involves the choice of an appropriate parameter estimation method for mixture models. The most commonly used estimator has been the Maximum Likelihood estimation method via the EM algorithm (ML/EM). Yet, ML/EM is also known for difficulties in finding global maximum, as well as for slow convergence, parameter inconsistency, and very large sample size needed in order to apply the asymptotic theory particularly to mixture models (Depaoli, 2013; Frühwirth-Schnatter, 2006; G. McLachlan & Peel, 2004). In contrast, a Bayesian method allows one to search through the parameter space and to have a full posterior distribution of a parameter that is free from local maximum and parameter uncertainty. Another advantage associated with using the Bayesian estimation framework to analyze PGMMs is that the measurement occasions are allowed to vary across individuals; if using the ML/EM estimator under the SEM framework, the measurement occasions have to be fixed equal for all individuals, which is not realistic in actuality. Moreover, turning

points analyzed by Bayesian methods can be random within class, which is not possible if using ML/EM methods in the SEM framework. Such properties, along with many others, have made Bayesian methods an attractive alternative to ML/EM in fitting mixture models (Richardson & Green, 1997).

Nevertheless, many questions regarding the performance of the two predominant estimation methods for fitting PGMMs still remain to be examined. Is the performance of ML/EM similar or dissimilar to the Bayesian method for estimating a PGMM? Specifically, how do ML/EM and the Bayesian method perform in terms of class parameter recovery and classification accuracy? Furthermore, how is the respective performance of ML/EM and the Bayesian estimation method in identifying the correct number of growth trajectories with a PGMM? Finally, in each of the three considerations, under what conditions would they perform similar or dissimilar to each other? The purpose of the study is to shed some light on the above questions.

In summary, the purposes of the dissertation are two-fold. First, the study investigated and compared the performance of a PGMM with unknown turning points under ML/EM and Bayesian estimation frameworks with regard to parameter recovery and classification accuracy. Second, the performance of a PGMM in class enumeration using ML/EM and Bayesian estimations were examined and compared to explore under what conditions one estimator outperforms the other for fitting a PGMM.

The dissertation proposal is organized as follows. Chapter I introduces the background and the purpose of the study. Chapter II presents a comparative study of the performance between ML/EM and Bayesian estimation methods in detecting turning

points, recovering population parameters and individual classification accuracy with PGMMs. Chapter III presents a comparative study of the performance of PGMMs between ML/EM and Bayesian estimation methods in growth trajectories enumeration accuracy. Chapter IV concludes with a summary of the findings and implications of the findings for applied research using PGMMs of unknown turning points.

CHAPTER II

LITERATURE REVIEW

2.1 Piecewise Growth Mixture Models with Unknown Turning Points

A PGMM is essentially a type of multivariate normal mixture model. One underlying assumption is that the repeated measures of the outcome y are a finite mixture of two or more latent subpopulations. In each latent subpopulation, individuals' mean growth trajectory exhibits distinct developmental stages before and after a turning point. Different functional forms can be formulated for the growth stages before and after the turning point. To keep it simple, we considered a two-stage piecewise linear growth mixture model.

To formulate a two-stage piecewise linear growth mixture model with one unknown turning point, suppose that the sample data consists of *K* subpopulations with *k* indexing subpopulations ((*k* = 1… *K*). The Level 1 (repeated measures) model for Trajectory *k* is specified as

$$y_{ijk} = \begin{cases} l_{1k}(t) : a_{1ik} + b_{1ik}(t_{ij}) + \varepsilon_{ijk} & t_{ij} \le \gamma_k \\ l_{2k}(t) : a_{2ik} + b_{2ik}(t_{ij}) + \varepsilon_{ijk} & t_{ij} > \gamma_k \end{cases}, \qquad (1)$$

where y_{ijk} is the response at the *j*th measurement for the *i*th individual in Trajectory *k*. a_{1ik} and b_{1ik} are the intercept and the slope growth factors of the first phase, and a_{2ik} and b_{2ik} denote the corresponding growth factors of the second phase after the turning point. γ_k is the location of the unknown turning point in Trajectory *k* marking the shift from one growth phase to the other. The location of γ_k is fixed within class but varies

across classes. ε_{ijk} is the level-1 residual for individual *i* at measurement *j* in Trajectory *k* [$\varepsilon_{ijk} \sim N(0, \boldsymbol{\sigma}^2_{\varepsilon k})$].

The trajectory is assumed to be continuous and exhibit no jump from stage one to stage two, therefore, the two stages for $l_{1k}(t)$ and $l_{2k}(t)$ are connected at the turning point. That is, when $\mathrm{t_{ij}} = \gamma_{\mathrm{k}}$, $\mathrm{a_{1ik}} + \mathrm{b_{1ik}}(\gamma_{\mathrm{k}}) = \mathrm{a_{2ik}} + \mathrm{b_{2ik}}(\gamma_{\mathrm{k}})$, which gives $\mathrm{a_{2ik}} = \mathrm{a_{1ik}} + \gamma_{\mathrm{k}}(\mathrm{b_{1ik}} - \mathrm{b_{2ik}})$. Thus Model (1) that has five parameters is reduced to a four-parameter model

$$\mathrm{y_{ijk}} = \begin{cases} \mathrm{l_{1k}(t) : a_{1ik} + b_{1ik}(t_{ij}) + \varepsilon_{ijk}} & \mathrm{t_{ij} \le \gamma_k}, \\ \mathrm{l_{2k}(t) : a_{1ik} + b_{1ik}\gamma_k + b_{2ik}(t_{ij} - \gamma_k) + \varepsilon_{ijk}} & \mathrm{t_{ij} > \gamma_k}. \end{cases} \tag{2}$$

The Level-2 (between-subject) model for Trajectory *k* is specified as

$$\begin{cases} \mathrm{a_{1ik}} = \mu_{\mathrm{a1k}} + \zeta_{\mathrm{a1ik}} \\ \mathrm{b_{1ik}} = \mu_{\mathrm{b1k}} + \zeta_{\mathrm{b1ik}} \\ \mathrm{b_{2ik}} = \mu_{\mathrm{b2k}} + \zeta_{\mathrm{b2ik}} \end{cases}, \tag{3}$$

$$\text{with} \quad \begin{bmatrix} \zeta_{\mathrm{a1ik}} \\ \zeta_{\mathrm{b1ik}} \\ \zeta_{\mathrm{b2ik}} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{\pi 00k} & \tau_{\pi 01k} & \tau_{\pi 02k} \\ \tau_{\pi 10k} & \tau_{\pi 11k} & \tau_{\pi 12k} \\ \tau_{\pi 20k} & \tau_{\pi 21k} & \tau_{\pi 22k} \end{bmatrix} \right), \tag{4}$$

where μ_{a1k}, μ_{b1k}, and μ_{b2k} are growth factor means and ζ_{a1k}, ζ_{b1k}, and ζ_{b2k} are random disturbances in their respective growth factors for Class *k*. The random disturbances are assumed to be uncorrelated with the Level 1 residuals. The Level 1 residuals and the Level 2 disturbances are also assumed to be uncorrelated with the latent growth factors.

However the parameterization of Model (2) cannot be specified and estimated in regular Structural Equation Modeling (SEM) programs. Harring, Cudeck, & du Toi (2006) and Kohli et al. (2013) suggest a re-parameterization to combine the two linear trajectories in Model (2) into one equation

$$y_{ijk} = \lambda_{1ik} + \lambda_{2ik} t_{ij} + \lambda_{3ik} \sqrt{(t_{ij} - \gamma_k)^2} + \varepsilon_{ijk} \ , \qquad (5)$$

Model (2) and (5) are identical given: $\lambda_{1ik} = (a_{1ik} + a_{2ik})/2$, $\lambda_{2ik} = (b_{1ik} + b_{2ik})/2$, and $\lambda_{3ik} = (b_{2ik} - b_{1ik})/2$. As an example, suppose we have 6 measurement waves with $t = (0,1,2,3,4,5)'$ and $\gamma_k = 3.3$ for the *k*th trajectory. Equation (5) can be written in the matrix form as follows

$$\begin{pmatrix} y_{i1k} \\ y_{i2k} \\ y_{i3k} \\ y_{i4k} \\ y_{i5k} \\ y_{i6k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \gamma_k \\ 1 & 1 & \sqrt{(1-\gamma_k)^2} \\ 1 & 2 & \sqrt{(2-\gamma_k)^2} \\ 1 & 3 & \sqrt{(3-\gamma_k)^2} \\ 1 & 4 & \sqrt{(4-\gamma_k)^2} \\ 1 & 5 & \sqrt{(5-\gamma_k)^2} \end{pmatrix} \begin{pmatrix} \lambda_{1ik} \\ \lambda_{2ik} \\ \lambda_{3ik} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1ik} \\ \varepsilon_{2ik} \\ \varepsilon_{3ik} \\ \varepsilon_{4ik} \\ \varepsilon_{5ik} \\ \varepsilon_{6ik} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 3.3 \\ 1 & 1 & 2.3 \\ 1 & 2 & 1.3 \\ 1 & 3 & 0.3 \\ 1 & 4 & 0.7 \\ 1 & 5 & 1.7 \end{pmatrix} \begin{pmatrix} \lambda_{1ik} \\ \lambda_{2ik} \\ \lambda_{3ik} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1ik} \\ \varepsilon_{2ik} \\ \varepsilon_{3ik} \\ \varepsilon_{4ik} \\ \varepsilon_{5ik} \\ \varepsilon_{6ik} \end{pmatrix} . \qquad (6)$$

2.2 Maximum-Likelihood Estimation of PGMMs via EM Algorithm in SEM

By specifying the design matrix in Model (5) appropriately using the polynomial constraint, the functional form can be directly implemented and estimated using regular SEM programs.

The EM algorithm can be applied to finding maximum likelihood (ML) parameter estimates for a PGMM assuming $\mathbf{z}_{ik}$, the conditional probability of individual *i* classified into trajectory *k*, "missing" but independently and identically distributed

according to a multinomial distribution. The complete-data log likelihood is formulated as follows

$$\log L(\boldsymbol{\psi} \mid \mathbf{y}, \mathbf{z}) = \sum_{k=1}^{K} \sum_{i=1}^{N} z_{ik} \log(\pi_k f(y_{ij} \mid \boldsymbol{\psi}_k)), \quad (7)$$

where $f(\mathbf{y}_{ij} \mid \boldsymbol{\psi}_k)$ is the density of an observation $\mathbf{y}_{ij}$ from Trajectory *k*, $\boldsymbol{\psi}_k$ is the set of model parameters, π_k is the probability of an individual being classified into Class *k* ($\pi_k > 0$ and $\sum_{k=1}^{K} \pi_k = 1$), and *K* denotes the number of growth trajectories.

The EM algorithm iterates between an E (for expectation) step and an M (for maximization) step. In the E-step, the quantity of $\mathbf{z}_{ik}$ and other parameter estimates can be obtained based on the previous iteration. In the M-step, the complete data log likelihood, with its parameter replaced by the current conditional expectation estimates, is then maximized and updated. The E- and M- steps are alternated repeatedly until the difference of the likelihood values is small enough to obtain "convergence" of a sequence of likelihood values (McLachlan & Peel, 2004; Redner & Walker, 1984).

The EM algorithm is straightforward and has become a general approach to maximum likelihood fitting of mixture models. However, it is not without limitations. Frühwirth-Schnatter (2006) summarized the limitations as follows: first, unless the latent trajectories in the growth mixtures are well separated, the sample size is large, or the estimation starts with reasonable values, the algorithm can be very slow in convergence. Second, the EM algorithm tends to give a local maximum, and it is difficult to identify and avoid local maximum when maximizing the log likelihood function. Third, the

standard errors of maximum likelihood estimates of a growth mixture model could be very difficult to obtain when using the EM algorithm. Fourth, a very large sample size is usually needed in order to apply the asymptotic theory of maximum likelihood, particularly to mixture models.

The limitations with the ML/EM estimator may not be easily bypassed in actual research settings. It is certainly not possible for applied researchers to have an optimal research situation in which sample size or the magnitude of latent class separation is very large. Nor is it possible for an applied researcher to be able to specify good starting values to begin with. A viable alternative for fitting a PGMM is the Bayesian estimation method.

2.3 Bayesian Estimation of PGMMs

In the Bayesian perspective, all unknown model parameter values are treated as random variables and the observed data are treated as fixed. The ultimate goal of Bayesian methods is to infer a posterior distribution of the parameters $\Pr(\boldsymbol{\psi} \mid \mathbf{y})$, which is the product of the mixture likelihood function $\Pr(\mathbf{y} \mid \boldsymbol{\psi})$ and prior distributions $\Pr(\boldsymbol{\psi})$. When a posterior distribution is difficult to integrate explicitly, Markov Chain Monte Carlo (MCMC) methods can be used to summarize the posterior distribution. The fundamental idea of the MCMC method is that we assemble a large number of samples from a posterior distribution $\Pr(\boldsymbol{\psi} \mid \mathbf{y})$, and then applies discrete formulas to these samples to summarize and obtain the mean and variance of the posterior distribution (Gelman et al., 2013; Gilks, Richardson, & Spiegelhalter, 1996; Gill, 2014).

One of the major distinctions between the ML/EM and the Bayesian methods is the specification of prior distributions in the Bayesian framework, which allows the inclusion of additional information of certainty into the estimation process. The specification of prior distributions falls in a continuum ranging from completely uninformative to very informative. An uninformative (or diffuse or objective) prior contains vague or very general information about a model parameter. The use of uninformative priors in Bayesian analysis yields results which are generally believed to be not too different from maximum likelihood estimation methods, as the uninformative priors have little effect on the likelihood function in the posterior distribution. However, all priors are informative in some way and there is no truly uninformative prior (Irony & Singpurwalla, 1997). Above all, the assumption of the functional form of a prior distribution for a model parameter already imposes information. Natarajan and McCulloch (1998) conducted a study on the impact that uninformative priors have on posterior distribution and recommended to be cautious about the results obtained using uninformative priors as they can lead to inaccurate posterior estimates particularly when sample size is small.

Somewhere on the continuum from an uninformative prior to an informative prior lies a weakly informative prior. The goal of using a weakly informative prior is to incorporate some sensible yet limited prior information into posterior estimation to regularize and stabilize the results in compliance with our knowledge (Richardson & Green, 1997). Summary statistics obtained using weakly informative priors for the

model parameters posterior are also believed to correspond closely to maximum likelihood estimates.

A prior that contains precise scientific information about a model parameter value is often referred as an informative prior. Informative priors are not simply ready to be used as they are based on expert opinion or results from previous data. However, the inclusion of information from expert belief or from previous models can increase the precision of the posterior distribution. Depaoli's (2013) study examined the impact of different prior distributions on mixture class recovery in the context of growth mixture modeling and found that only "accurate" informative priors based on population values produces optimal parameter recovery, and the performance of weakly informative priors varied depending on sample size.

2.4 ML-Based Class Enumeration Indices

The accuracy of the number of classes extracted during estimation is of prime importance as it can affect the features of the latent growth trajectories which, in turn, can affect within-class parameter estimates and statistical inference. Generally, under the ML/EM framework, four major categories of model selection statistics are available for identifying the optimal number of growth trajectories: (a) information-based criteria (IC) statistics, (b) nested model likelihood ratio test derivative, (c) the entropy statistic and entropy-penalty based indexes, (d) goodness of fit tests (Bauer & Curran, 2003b; Liu & Hancock, 2014; Peugh & Fan, 2012; Peugh & Fan, 2015; Tofighi & Enders, 2008).

2.4.1 Information-Based Criteria (IC) Statistics

2.4.1.1 Akaike's Information Criterion (AIC)

Akaike (1974) developed an information criterion as a measure to identify an optimal and parsimonious model from competing models. The equation of AIC is formulated as below

$$\text{AIC} = -2 \log L + 2p, \tag{8}$$

where log L is the maximum log likelihood and p is the number of free parameters in a PGMM with K growth trajectories. AIC was found to have the tendency to over-extract the number of classes in the mixture modeling context (Celeux & Soromenho, 1996; Hurvich & Tsai, 1989; G. McLachlan & Peel, 2004; Soromenho, 1994).

2.4.1.2 Bayesian Information Criterion (BIC)

Schwarz (1978) proposed an alternative approach to AIC within the Bayesian framework

$$\text{BIC} = -2 \log L + p \log N, \tag{9}$$

where log L of the Bayes estimator is equivalent to the maximum likelihood estimator given a large sample, and the penalty term of p log N penalizes extra parameterization. BIC differs from AIC by the multiplication of log N /2 to the second term of AIC. The larger N is, the more remarkable difference these two measures would show. BIC tends to pick up models with fewer parameters than AIC.

2.4.1.3 Consistent AIC (CAIC)

Bozdogan (1987) provided an extension to AIC to make it asymptotically consistent and more stringent on penalty for extra parameterization.

$$CAIC = -2 \log L + p (\log N + 1), \quad (10)$$

where the penalty term, p (log N +1), is designed to depend on the sample size, which makes AIC asymptotically consistent and at the same time increases penalty for over-parameterization. Given a more stringent penalty term, CAIC tend to pick up even simpler models than AIC and BIC.

2.4.1.4 Sample Adjusted BIC (SABIC)

A similar adjustment to BIC was first proposed by Rissanen (1978) to rectify the coincidental nature of the penalty term in BIC:

$$SABIC = -2 \log L + p \log [(N+2) /24] \quad (11)$$

The second term in SABIC differs from the corresponding term in AIC, BIC, and CAIC, and gives an even heavier penalty for extra parameterization.

2.4.1.5 DBIC

Draper (1995) used Laplace approximations to compute Bayes factors for model selection and the simpler approximation is:

$$DBIC = -2 \log L + p (\log N - \log 2\pi) \quad (12)$$

Controversies arose regarding the term $-p \log 2\pi$. Draper (1995) held that compared with BIC, the inclusion of $-p \log 2\pi$ would improve the accuracy in model selection; however, others believed that the term hurts the accuracy of the approximation (Kass & Wasserman, 1995; Raftery, 1995).

2.4.1.6 Other Information Criteria

Hannan and Quinn (1979) proposed an HQ index which multiplies a term, log (log N), to the penalty term in AIC to obtain strong consistency. Hurvich & Tsai (1989)

proposed an index AICC which adds one term, ((2[p+1][p+2])/(N-p-2)), to AIC. They claimed that AICC can rectify the over fitting problem of AIC when the sample size is small. It is noted that the term, ((2[p+1][p+2])/(N-p-2)), can also be added to most of the IC statistics and the Entropy-based criteria discussed below to create their corresponding sample size adjusted versions (Peugh & Fan, 2012; Peugh & Fan, 2015; Sclove, 1987). Finally, Andrews and Currim's (2003) proposed another variant of AIC (i.e., AIC-3), which replaced the penalty term in AIC with 3 multiplied by the number of free parameters.

2.4.2 Nested Model Likelihood Ratio Test Derivative

When determining the correct number of classes in a growth mixture model, we are inclined to use the traditional likelihood ratio test -2 (log L_{k-1} – log L_k) to test the competing models with k latent trajectories versus a nested model with k-1 trajectories. However, with mixture models, the likelihood ratio statistic does not follow the usual asymptotic χ^2 distribution with degrees of freedom equal to the difference between the number of parameters under the competing model and the nested model. This is because when the parameters for one or more growth trajectories are fixed to be zero, the parameters (i.e., the mixing proportions) under the null hypothesis are set to be zero, at the boundary of the parameter space, resulting in non-identifiable parameters under the null model (McLachlan & Peel, 2004). There have been many adjusted versions of the likelihood ratio statistic, among which, the LMR LRT test and the BLRT test are two widely used procedures.

2.4.2 .1 The LMR LRT Test

Lo, Mendell, and Rubin (2001) extended the test procedure originally developed in the context of regression by Vuong (1989) to normal mixture modeling. According to Lo, Mendell, and Rubin (2001), the LMR LRT test is able to approximate the likelihood ratio test procedure of *k*-1 growth trajectories normal mixture versus a *k* growth trajectories normal mixture as shown in Vuong's (1989) study. Under general regularity conditions, the likelihood ratio statistic of the H_0 and H_1 has asymptotic property and distributed as weighted sum of independent chi-square values with one degree of freedom.

2.4.2 .2 Bootstrapping LRT (BLRT)

McLachlan (1987) showed that approximate *p*-values of the likelihood ratio test could be obtained via a parametric bootstrap method. The following procedures are used to conduct the BLRT test:

(i) Use the real data to estimate the null PGMM with *k*-1 latent trajectories and the alternative PGMM with k latent trajectories and compute the 2 times log likelihood difference (-2 (log L_{k-1} – log L_k)).

(ii) Use the parameter estimates for the null PGMM model obtained in step (i) to generate data and analyze the generated data again by the null PGMM model and the alternative PGMM model and calculate the 2 times log likelihood difference. This step is repeated for a sufficiently large number of times to generate the bootstrap sampling distribution of the 2 times log likelihood difference.

(iii) Estimate the *p*-value of the observed value of 2 times log likelihood difference obtained in step (i) by comparing it with the bootstrap sampling distribution created in step (ii).

2.4.3 Entropy-based Criteria

As standard regularity of conditions for the likelihood ratio (LR) test does not hold in the growth mixture context, the ICs that rely on the usual asymptotic theory of the likelihood ratio test are problematic in practical situations (McLachlan & Peel, 2004). Entropy-based criteria were proposed to avoid the disadvantages associated with information-based criteria statistics. In mixture modeling, the primary interest is in cluster analysis, where choosing a correct number of clusters is essential. Entropy-based criterion takes into account of the clustering purpose, and is able to choose one mixture model over the other in terms of model-based classification accuracy (Biernacki & Govaert, 1997; Biernacki, Celeux, & Govaert, 2000; Celeux & Soromenho, 1996). Below is a review of the commonly used entropy-based criteria.

2.4.3.1 Normalized Entropy Criterion (NEC)

Entropy (EN) is a measure of the performance of a PGMM in providing accurate classification, with larger values indicating poor classification accuracy. Entropy is calculated by:

$$\mathrm{EN}(\hat{\mathrm{t}}) = -\sum_{\mathrm{i}=1}^{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{K}}\hat{\mathrm{t}}_{\mathrm{ik}}\log\hat{\mathrm{t}}_{\mathrm{ik}}\ , \qquad (13)$$

where $\hat{\mathrm{t}}_{\mathrm{ik}}$ is the estimated conditional probability that an individual *i* is in group *k*. Note that entropy labeled in Mplus Version 7.2 output is an alternative rescaled entropy

measure proposed by Ramaswamy, DeSarbo, Reibstein, and Robinson (1993). The rescaled entropy ranges from 0 to 1 with 1 indicating perfect classification (Bauer & Curran, 2003a; B. Muthén & Muthén, 2000). Celeux & Soromenho (1996) proposed the normalized entropy criterion (NEC) to make entropy more effective in assessing the number of clusters in a mixture model:

$$\text{NEC} = \frac{\text{EN}(\hat{t})}{\log \text{L} - \log(1)}, \tag{14}$$

where log L is the log-likelihood of the mixture model with K components, and log (1) is the log-likelihood for a single class model. Later, Bieernacki, et al. (2000) suggested setting the NEC to 1 for the one-class model for comparisons between one-class models and two-class models as the NEC is undefined when K=1.

2.4.3.2 Classification Likelihood Criterion (CLC)

It is observed that the estimated mixture log likelihood can be decomposed into two components (Hathaway, 1986; G. McLachlan & Peel, 2004)

$$\text{Log } L = \log L_c + \text{EN}(\hat{t}), \tag{15}$$

where log L_c is the classification likelihood (or the complete-data likelihood) which equals log L given a perfect classification of observations, while the entropy ($EN(\hat{t})$) captures the uncertainty of classification. The classification likelihood information criterion (CLC) is derived based on the relation between mixture log likelihood and classification likelihood in Equation (8), which is calculated by:

$$\text{CLC} = -2 \log L + 2\,\text{EN}(\hat{t}) \tag{16}$$

In CLC, the estimated entropy is used to penalize model complexity.

2.4.3.3 Integrated Completed Likelihood Criterion and BIC (ICL-BIC)

AIC, BIC, and other ICs as well, favor models that can more accurately reproduce the observed data, yet fail to take into account the performance of a PGMM in terms of growth trajectories classification accuracy. Biernacki, et al. (2000) proposed a BIC-like approximation procedure to approximate integrated completed likelihood (ICL), hoping to correct for the tendency of ICL to overestimate the correct number of latent trajectories. The ICL-BIC is formulated as

$$\text{ICL-BIC}=-2\log L + p\log N + 2\,\text{EN}(\hat{t}). \quad (17)$$

ICL-BIC becomes equivalent to BIC if without the entropy term $2\ \text{EN}(\hat{t})$, yet adding the term would take into account a mixture models' accuracy in partitioning trajectories. If without the penalty term $p \log N$, the procedure would be reduced to CLC; the addition of the penalty term to CLC enables the completed log-likelihood to penalize over parameterization.

For the information criteria-based statistics and the entropy-based indices, the smaller value indicates better model fit. Therefore, when estimating the number of trajectories, researchers need to fit several competing PGMMs (i.e., one-class, two-class, three-class PGMMs, etc.), and the model that gives the minimum value on those indices is the "best fitting" PGMM.

2.5 Bayesian-Based Enumeration Indices

Bayesian methods for determining the number of latent growth trajectories with GMM have been widely used as effective alternatives to ML/EM methods particularly in situations when we do not have a general theoretical knowledge for the number of

growth trajectories (Lindsay, 1995; Steele & Raftery, 2009) or when the estimates of parameters are not asymptotically consistent, leading to inappropriate information-based criteria statistics under the ML/EM estimator (Park & Lord, 2009).

Within the Bayesian framework, there are two main approaches that could be used to identify the correct number of growth trajectories with a PGMM: the first type of approaches assumes that the number of classes is an unknown variable and could be estimated using methods such as Dirichlet process mixtures (Escobar & West, 1995), distributional distances (Mengersen & Robert, 1993), reversible jump MCMC (Richardson & Green, 1997), and Birth-and-Death MCMC (Stephens, 2000). The Bayesian paradigm is "particularly suited to mixture analysis especially with an unknown number of components" (Richardson & Green, 1997, p.732). However, these methods come with a heavy cost of intensive computation and the sensitivity of the posterior distribution to the prior selection for the unknown number of components (Park, Zhang, & Lord, 2010).

In this study, we opt for the second type of approaches using various Bayesian model selection criteria including information-based criteria (AIC, BIC), a variation of deviance information criterion (DIC3; Celeux, Forbes, Robert, & Titterington, 2006), a Watanabe-Akaike information criterion (WAIC2; Watanabe, 2010), a log-pseudo marginal likelihood (LPML; Geisser & Eddy, 1979) and pseudo-Bayes Factor (PsBF; Geisser & Eddy, 1979).The advantage of information-based criteria is that they are easy to calculate, and with exception to DIC, none of them depends on prior information. Moreover, from a Bayesian perspective, the posterior model probability should be the

tool for model comparison; therefore, the information criteria do not need a formal Bayesian model justification in their use for model comparison (Koop, 2003).

2.5.1 AIC and BIC

We have presented AIC and BIC in the ML framework in the above. In the Bayesian framework these two information-based criteria use the Laplace approximation to obtain the integrated likelihood assuming ignorable priors. Keribin (2000) showed that BIC was reliable in identifying the correct number of components in a mixture model. Steele & Raftery (2009) found that BIC was highly accurate in detecting the correct number of components in normal mixture models.

2.5.2 Deviance Information Criterion (DIC3)

The Deviance Information Criterion, or DIC, was proposed by Spiedelhalter et al. (2002) based on Bayesian measures of model fit and model complexity, with an intention to lighten the computational demand involved in calculating Bayes factors (McGrory & Titterington, 2007; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). DIC is defined as:

$$\text{DIC} = \overline{D(\theta)} + P_D = 2\overline{D(\theta)} - D(\hat{\theta}) \ , \qquad (18)$$

where $P_D = \overline{D(\theta)} - D(\hat{\theta})$, and $\overline{D(\theta)}$ is the mean of -2 log L over the posterior distribution, and $D(\hat{\theta})$ is the -2 log L of the posterior model. P_D is in fact an estimate of the effective number of parameters in the Bayesian model, a penalty term for over parameterization (Johnson, 2004; McGrory & Titterington, 2007; Spiegelhalter et al., 2002).

DIC3 (Celeux, Forbes, Robert, & Titterington, 2006) is a variation of DIC, which replaces $D(\hat{\theta})$ with the estimated density -2 log $\hat{f}(\mathbf{y})$ approximated in each MCMC evaluation. DIC3 is defined as:

$$\text{DIC3} = \overline{D(\theta)} + 2 \log \hat{f}(\mathbf{y}) \tag{19}$$

In this study we use DIC3 rather than the standard DIC because studies have shown that DIC3 can better account for the number of parameters in mixture models and takes care of the endemic issues of non-identifiability and label switching in mixture models (Celeux et al., 2006).

2.5.3 Watanabe-Akaike information criterion (WAIC)

Watanabe (2010) introduced WAIC to approximate Bayesian cross-validation. The computation involves first calculating the log predictive density, and then correcting the log likelihood with a penalty term for model overfitting (Gelman, et al. 2013). There have been two proposals for the penalty terms to adjust for model overfitting (i.e., P_WAIC1 and P_WAIC2). Watanabe (2010) and Gelman, et al. (2013) gave detailed explanations of the formulation of P_WAIC1. The present study focused on WAIC2 defined as:

$$\text{WAIC2} = -2 (\log L + \text{P\_WAIC2}) \tag{20}$$

where P_WAIC2 is the penalty term defined as

$$\text{P\_WAIC2} = \sum_{i=1}^{n} \text{var}_{\text{post}} (\log L) . \tag{21}$$

P_WAIC2 is obtained by summing the posterior variance of individual log likelihood over *n* individuals. The multiplicative factor of -2 can rescale WAIC2 to be on

the deviance scale and thus make it comparable to AIC, BIC, and DIC3. As a fully Bayesian estimate, WAIC is more desirable than AIC, BIC, and DIC as it averages across the posterior, while the latter three evaluate model performance based on the plug-in predictive density. Furthermore, it has shown that WAIC2 works particularly well with mixture and hierarchical models (Gelman, et al. 2013).

2.5.4 Log Pesudo Marginal Likelihood (LPML)

As a summary statistics of Conditional Predictive Ordinate (CPO), LPML (Geisser & Eddy, 1979) has been found to be a useful measure for model comparison in the Bayesian context. LPML is defined as

$$\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i) \ , \qquad (22)$$

where $\text{CPO}_i = p(y_i \mid y_{[i]}) = \int p(y_i \mid \theta, y_{[i]})\, p(\theta \mid y_{[i]}) d\theta$, estimates the cross-validation predictive densities, with y_i denoting the i^{th} observation, $y_{[i]}$ denoting the set of observations excluding y_i and θ denote all parameters under the model.

2.5.5 Pseudo-Bayes factor (PsBF)

The method of Bayes factors is commonly used within the Bayesian framework for model comparisons. However, Bayes factor has some drawbacks. It is difficult to compute and interpret Bayes factor when priors are diffuse. In addition, the Bayes factors suggested by Lewis and Raftery (1997) are based on the unrealistic assumption that one of the compared models is the true model (Johnson, 2004). Pseudo-Bayes factor (PsBF) is a more flexible alternative to Bayes factor. PsBF can be estimated via a predictive likelihood approach which assumes that only part of the observations is used

in estimating a model. The Pseudo Bayes factor is defined as the ratio of products of "leaving one out" cross-validation predictive densities based on CPO under Model 1 and Model 2:

$$\mathrm{PsBF} = \frac{\prod_{i=1}^{N} \mathrm{CPO}_i(\mathrm{ModelA})}{\prod_{i=1}^{N} \mathrm{CPO}_i(\mathrm{ModelB})} \quad . \tag{23}$$

PsBF can also be computed using LPML as PsBF = EXP ($LPML_1$-$LPML_2$).

2.6 A Review of GMM's Class Enumeration Performance

Researchers have investigated the performances of many of the model selection indices in terms of the class enumeration accuracy in Growth Mixture Models. The current study will benefit from a systematic review of those studies. The study of Tofighi and Enders (2008) compared the performances of the following indices: BIC, SABIC, AIC, CAIC, SACAIC, the LMR LRT nested model test, and the goodness of fit statistics of the Multivariate Skewness Test (MST) and the Multivariate Kurtosis Test (MKT). A population GMM with *k*=3 heterogeneous growth trajectories were used in the study. Across all the independent experimental conditions of sample size, class proportion, class separation, and inclusion and exclusion of covariates, SABIC and the LMR LRT test were the "most accurate" indices (Tofighi & Enders, 2008, p. 383). The study also found that the inclusion of covariates had detrimental effect on correct trajectory identification rate across all the enumerations measures.

The work of Peugh and Fan (2012) examined the performance of class enumeration measures when the population is either homogeneous (*k*=1 linear growth trajectory) or heterogeneous (*k*=3 linear growth trajectories). They found that entropy-

based indexes such as CLC, ICL-BIC, SAICL-BIC correctly identify the homogeneous growth trajectory regardless of sample size while the performance of other information criteria (ICs), such as CAIC, BIC, and DBIC varied depending on sample size. When identifying heterogeneous growth trajectories, the performance of enumeration indexes varied across the design factors and none of the enumeration indices performed well particularly when the proportions of different trajectories were unequal.

Peugh & Fan (2015) extended their study in 2012 to test Muthén's (2003) hypothesis that the performance of latent trajectory identification measures could be improved by including time-invariant (antecedent) and time-varying (concurrent) covariates, and by regressing a distal (consequent) outcome on the extracted growth mixture classes. The study used a population GMM that had three latent trajectories that differed in both intercepts and slopes. They found that when sample size is small (N ~300-500), even with a high class separation, the effects of including antecedent, concurrent, and consequent covariates were negligible for the majority of the enumeration indices. However, when sample size is large (N =3000) and class separation is high, particularly under equal class proportions, the inclusion of antecedent, concurrent, and consequent covariates largely improved the performance of DBIC, SACAIC, HQ, SABIC, SAICL-BIC (correct class enumeration rate > 97%). However, some enumeration measures (e.g., SADBIC, NHQ, and BLRT) saw a slight decrease in correct model identification rate. Since in applied research, the scenarios that give improved performance of enumeration indices are highly unlikely, the authors restated

the importance for the use of both statistical and substantive checking when population heterogeneity is suspected.

Another most recent study by Liu and Hancock (2014) evaluated the "two-step" approach proposed by Bauer & Curren (2004) in the context of GMM. According to Bauer & Curren (2004), misspecifications of the structural model could lead to over-extraction of latent classes and they suggested following a "two-step" rule to identify a correct number of classes. First, we fit an unrestricted (or saturated) model to the data to determine the optimal number of latent classes, as a saturated model could prevent the extraction of spurious latent classes due to misspecifications of the structural model. Then we estimate the sample means and covariance using the optimal number of classes identified in Step 1. Liu and Hancock (2014) examined the performance of unrestricted multivariate normal mixture models (UMMs) versus the linear GMMs in the accuracy of class enumeration using sample data generated from k =2 GMM. In their findings, the indices that work well for UMM also works well with linear GMM; some indices indicate that UMM does not work well in class identification probably due to its over-parameterization. Overall, they recommended using BIC and DBIC in the linear GMM settings and the DBIC across different types of mixture models. Besides, CAIC, SACAIC, BIC, SABIC, DBIC, LMR, and BLRT are sensitive to sample size and perform better with the increase of sample size.

In summary, due to the different models and conditions examined in previous simulation studies, no consensus could be reached regarding the performance of class enumeration indices in GMM with the ML/EM estimator. The ML/EM estimator has

been the dominant estimation method in growth mixture models due to their efficiency in computation time. However, it's been unanimously acknowledged that ML/EM methods have severe problems such as convergence to local maxima, and the tendency to overfitting especially when data are sparse or noisy (Depaoli, 2013; Neelon, Swamy, Burgette, & Miranda, 2011).

As for class enumeration with GMM in the Bayesian framework, no study has been done to evaluate the class enumeration performance of commonly used Bayesian-based model selection indices. In particular, there has been no simulation study to investigate the class enumeration performance of the model selection indices in PGMMs with unknown turning points and to compare their performances based on the ML/EM and the Bayesian estimations. The current research conducted a comprehensive Monte Carlo simulation study to fill the gap and to advance the knowledge of PGMMs under the two dominant estimation methods in different data scenarios.

CHAPTER III

STUDY ONE: DETECTING UNKNOWN TURNING POINTS USING PIECEWISE GROWTH MIXTURE MODELS: A COMPARISON OF ML/EM VERSUS BAYESIAN ESTIMATION

3.1 Overview

Piecewise growth mixture modeling (PGMM) is a longitudinal data analytic technique that can be used to investigate the growth trajectories consisting of distinct developmental phases of several unknown heterogeneous subpopulations (Muthén, 2008). Oftentimes, a growth trait measured over a sufficiently long interval of time does not follow a linear trend or a simple polynomial functional form because the growth trend exhibits differential developmental stages. PGMM is an extension of the conventional growth mixture modeling (GMM) techniques to account for the nonlinear processes in developmental changes. Like GMM, PGMM explores the heterogeneity in the population's growth trajectories and examines trajectory-specific variance components. More importantly, PGMM allows the identification of distinctive turning points (or change points or knots), the presence of which marks the change of one growth rate to the other and thus introduces an inherently nonlinear growth processes in the functional form (Harring, Cudeck, & du Toit, 2006; Kwok, Luo, & West, 2010).

The flexibility and dynamism of PGMM has attracted enthusiastic applications in social and behavioral science. For example, Li et al. (2001) found development process in alcohol use among the youths consisting of two distinct growth processes with one subpopulation exhibiting an increase in alcohol use only after middle school and the other subpopulation showing a continued, linear growth throughout both middle and

high school years. Kim and Kim's (2012) study identified five growth trajectories of individuals' smoking craving level before and after the quit date using PGMMs.

The major difficulty in specifying a PGMM concerns how to optimally determine the turning point. Sometimes, graphical exploration of individual and average growth patterns over time could be helpful in suggesting a turning point and differences in patterns of change (Weiss, 2005). A brief review of the literature showed that the majority of applied studies using PGMMs tend to rely on theoretical considerations to specify a priori the location of a turning point. For example, Uher et al. (2010) used PGMM to model the individual variability in clinical response over antidepressant treatment and categorized the individuals into two groups: one with overall gradual improvement and the other with rapid improvement at the initial stage followed by a more gradual improvement. McAuley et al. (2011) used PGMM to examine the differential effects of randomized controlled exercise trial on self-efficacy and identified three growth trajectories, each exhibiting distinctive growth rate before and after intervention. In these two studies, the turning point was set at the time of intervention. However such considerations may not be always reasonable, because the turning point may occur after the intervention due to delay in response to intervention.

There have been quite a few statistical procedures proposed to determine unknown turning points. For example, Kwok et al's (2010) study proposed to use modification index to detect the turning point in the linear latent growth modeling framework. They argued that the turning point can be optimally located using modification indexes because the fixed loadings on the slope factor has to be relaxed to

account for the change in growth rate. Other approaches have been proposed in the context of mixed effect models in which the turning point is a parameter in statistical models to be estimated and is allowed to have inter-individual and intra-individual variability (Dominicus, Ripatti, Pedersen, & Palmgren, 2006; Wang & McArdle, 2008). Recently, a two-stage piecewise growth mixture model with an unknown turning point was proposed by Kohli, Harring, & Hancock (2013) in the framework of Structural Equation Modeling (SEM).

Generally speaking, using statistical procedures to estimate turning points is advantageous over the complete reliance on theoretical considerations to specify a priori turning points. First, it does not limit the shape of the trajectories to the assumption that the developmental phases are connected by a priori determined fixed time points. Second, it allows the formulation of different functional forms for the different phases of growth such that each phase does not have to conform to the same function (Harring et al., 2006; Khoo, West, Wu, & Kwok, 2006; Kohli, Harring, & Hancock, 2013; Kwok, Luo, & West, 2010). Furthermore, allowing free estimation of turning points and time specific factor loadings can oftentimes entail a more optimal functional form in characterizing observed data patterns over time (Kwok, Luo, & West, 2010; Wood & Jackson, 2013).

Despite the advantages of the newly proposed PGMM procedure by Kohli et al. (2013) and the growing interest in using PGMMs to describe differential growth phases in developmental changes, many issues have not yet been resolved. One of the issues involves the choice of an appropriate estimation method for PGMMs. So far, the most

commonly used estimator has been the Maximum Likelihood estimation method via the EM algorithm (ML/EM). On the other hand, the Bayesian estimation framework via the Markov Chain Monte Carlo (MCMC) algorithm has been argued to be a better alternative to ML/EM for fitting mixture models (Richardson & Green, 1997). However, many questions regarding the performance of the ML/EM vs. Bayesian estimation in analyzing PGMMs with unknown turning points remain to be examined. The purpose of this research was to conduct a comprehensive Monte Carlo study to compare the performance of ML/EM and Bayesian estimation for estimating unknown turning points and other growth factors with PGMMs. Specifically, parameter recovery and classification accuracy would be examined. Moreover, the study would also explore under what circumstances the use of one estimator in analyzing a PGMM would outperform the other.

In order to carry out a Bayesian analysis, we need to specify a prior distribution for every unknown parameter in model (2). Prior probability distribution, often simply called prior, describes what is known a priori about a model parameter value to be estimated.

3.2 Methods

The population model used in this study was a 2-class PGMM, where each class exhibits a two-stage linear-linear growth trend. The number of classes was chosen based on past empirical studies using PGMM (e.g., Li, et al., 2001; Li, et al., 2001; Kohli, et al., 2013; Zhao & Banerjee, 2012). Data with two known classes under two-level model

were first generated and then analyzed using Bayesian estimation method and the ML/EM estimation method.

In the above model (2), within Trajectory *k*, a total of 11 parameters need to be specified: four fixed effect coefficients (i.e., μ_a, μ_{b1}, μ_{b2}, and γ) and seven variances and covariance of random effects (i.e., σ^2, $\tau_{\pi 00}$, $\tau_{\pi 10}$, $\tau_{\pi 20}$, $\tau_{\pi 11}$, $\tau_{\pi 21}$, $\tau_{\pi 22}$). Table 1 displays the population parameters for the average growth models and the covariance components of random effects for the two trajectories to be generated.

The average growth models were specified based on the parameter estimates in Kohli, et al.'s (2013) study of verbal skill acquisition data set. Specifically, the following model parameters in their study were chosen: the means of the intercept growth factor and the slope factors in the first and second phase for Class 1 and the means of the slope factors in the first and second phase for Class 2. The mean of the intercept growth factor in the first phase for Class 2 was calculated based on the magnitude of within-class variance parameters determined by the degree of separation. For simplicity, the error variance for each time point was set to follow a standard normal distribution.

3.2.1 Design Factors

Although no PGMM simulation have been published to date, some simulation studies related to GMM have shown consistent findings about what factors could impact class parameter recovery and classification accuracy (Chen, Kwok, Luo, & Willson, 2010; Enders & Tofighi, 2008; Henson, Reise, & Kim, 2007; Tofighi & Enders, 2008). Based upon the previous findings, five design factors were considered in the study, including (a) the degree of trajectory separation, (b) the number of repeated measures,

(c) mixing percentages of latent classes, (d) sample size, (e) specification of prior distributions.

Table 1

Population Parameters for the Two Classes

	Class 1	Class 2
a	25	22.5
b1	-3.78	-4
b2	-0.24	-0.18
γ	4.3	2.4[1]
σ^2	1.0	1.0

High Separation

$$T_\pi = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .1 & & \\ .025 & .05 & \\ 0 & .0175 & .05 \end{bmatrix}$$

or

Low Separation

$$T_\pi = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .7 & & \\ .175 & .35 & \\ 0 & .1225 & .35 \end{bmatrix}$$

[1] When the number of time points change to 10, the turning point was set to be 7.2 for Class one, and 4.1 for Class two to create equal proportion of distance between the initial status and the turning point for both levels of time points.

3.2.1.1 Degree of Class Separation

We varied the degree of class separation by manipulating the within class variance components. The bigger the variation of the within class variance components are, the less separated and thus more overlapping the two classes are. We chose a small size of variances and covariance matrix of the random effects for the high class separation condition according to Raudenbush and Liu's (2001) criteria, which was set as follows:

$$T_{\pi} = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .1 & & \\ .025 & .05 & \\ 0 & .0175 & .05 \end{bmatrix}$$

For the low separation condition, we increased the variance and covariance as follows.

$$T_{\pi} = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .7 & & \\ .175 & .35 & \\ 0 & .1225 & .35 \end{bmatrix}$$

In the high separation condition, given = .1 and , the difference between the intercept for Class 1 and that for Class 2 was approximately 2.4 standard deviation units. In the "low separation" condition, the intercept for Class 1 and Class 2 differed by approximately 1.9 standard deviation units. They were similar to the high and low separation conditions specified in Tofighi and Enders' (2008) study.

In both high and low separation conditions, $\tau_{\pi 11}$ and $\tau_{\pi 22}$ were set to be half of the size of $\tau_{\pi 00}$, because the variation of the intercept has generally been larger than the variation of the growth trends in longitudinal data. The size of the covariance, $\tau_{\pi 10}$ and $\tau_{\pi 21}$ was set to be .025 and .0175, .175 and .1225 in high and low separation conditions so that there was a moderate correlation (i.e., ρ=0.35) between the intercept and growth

factors in the first phase and between the growth factors in the first and second phases as well.

3.2.1.2 Sample Size

The sample size factor values were chosen based on the conditions used in the past simulation studies (Chen et al., 2010; Enders & Tofighi, 2008; Henson et al., 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008) and a careful review of substantive GMM and PGMM applications. As a result of a literature search in Web of Sciences (from 2000-2013) for studies applying GMM and PGMM in different substantive areas, a total of 346 studies were found. A random sample of 100 out of the 346 studies were reviewed and the studies varied dramatically with respect to sample size. We chose the sample size to be 300 and 1000, representing approximately the 25^{th} and 75^{th} percentiles of the sample size distribution in the 100 studies that we reviewed.

3.2.1.3 Number of Repeated Measures

It is more typical for a two-piece linear growth model to include more than two time points in each piece, with 5 points in total making it possible to estimate a full growth model (fully random) (Bollen & Curran, 2006). The use of fewer than 5 time points restricts the possibility to estimate a complete set of random effects (i.e., variances and covariance for the intercept and slope factors). Hence we chose 6 waves of repeated measures as one of the conditions. We chose 10 waves of repeated measures as the other condition since the mean number of waves among the empirical GMM studies that have more than 6 measurement waves is about 10.

3.2.1.4 Mixing Percentage of the Two Classes

Following the simulation studies in Chen et al. (2010), Nylund et al. (2007), and Liu and Hancock (2014), we considered two sets of mixing percentages of the two classes. The mixing percentages in the balanced condition were 50% and 50% for the two classes. In the unbalanced situation, we considered: 75% vs. 25% for the fast-decreasing group and the slow-decreasing group respectively. The mixing percentages in the unbalanced situation correspond to the mean mixing percentages in the empirical studies using PGMM (Zhao & Banerjee, 2012; Li et al., 2001; Kohli, et al., 2013).

3.2.1.5 Prior Specifications for PGMMs

To estimate a Bayesian PGMM, each parameter has to be assigned a prior distribution. The unknown parameters in a PGMM include the mixture proportions (π_k), mean vector of the growth factors and the turning point $\boldsymbol{\eta}_k = (\mu_{a_k}, \mu_{b1_k}, \mu_{b_{2k}}, \mu_{\gamma_k})'$, the variance and covariance vector $\boldsymbol{\varphi} = (\sigma^2_{a_k}, \sigma^2_{b_{1k}}, \sigma^2_{b_{2k}}, \sigma_{a_k b_{1k}}, \sigma_{a_k b_{2k}}, \sigma_{b_{1k} b_{2k}})'$, and the variance of the residuals ($\boldsymbol{\Omega}_{e_k}$). Non-informative Jeffrey's prior was used for $\pi(\pi \sim \text{Beta}(0.5,0.5))$. The prior distributions for the random effects variances were given the Inverse Wishart (IW) distribution [i.e., $\sigma^2_{a_k}, \sigma^2_{b_{1k}}, \sigma^2_{b_{2k}}, \sigma_{a_k b_{1k}}, \sigma_{a_k b_{2k}}, \sigma_{b_{1k} b_{2k}}$] ~*IW* (3, 10) and the level-1 error variance were given the inverse gamma distributions with both the shape and scale parameters being known [i.e., $\sigma^2_{e_{ijk}}$ ~*IG* (0.001, 1000)]. For the prior distributions for the means of the growth factors and the turning point, we used $\mu_{a_k} \sim \text{N}(\mu_a, \sigma^2_a)$, $\mu_{b1_k} \sim \text{N}(\mu_{b_1}, \sigma^2_{b_1})$, $\mu_{b_{2k}} \sim \text{N}(\mu_{b_2}, \sigma^2_{b_2})$, and $\mu_{\gamma_k} \sim \text{N}(\mu_r, \sigma^2_\gamma)$ in which the hyperparameters (i.e., μ_a, σ^2_a, μ_{b_1}, $\sigma^2_{b_1}$, μ_{b_2}, $\sigma^2_{b_2}$) were manipulated in our simulation design to

examine whether priors have impact on the performance of Bayesian methods in estimating PGMMs with unknown turning points.

Following Depaoli's (2013) study, three classes of the prior distributions for the means of the growth factors were considered: weakly informative priors, data driven priors, and accurate informative priors. For a normally distributed growth parameter with certain mean and variance hyperparameters, the sizes of the hyperparameters determine the degree of certainty a prior has on the posterior estimate of the parameter value. For informative prior, the mean hyperparameter was set to be the corresponding growth parameter population value, while the variance hyperparameter was set to be 5% of the growth parameter population value. For example, since the population value of the intercept of Class 1 was 25, the mean hyperparameter was set to be 25 and the variance hyperparameter was calculated to be 1.25 (25*5%). So the informative prior for the intercept is *N* (25, 1.25). Weakly informative priors were specified by taking the population value as the mean hyperparameter but 50% of the population value as the variance hyperparameter [i.e., *N*(25, 12.5)]. The weakly informative and informative priors for the rest growth parameters were determined in a similar way. As for the data driven priors, the ML/EM parameter estimates of each growth factor (i.e., the mean and variance) in one replication for each data scenario was obtained as the mean hyperparameter and the variance hyperparameter for the growth factors.

Combining all the design factors, the simulation used a 2 (magnitude of the T_π matrix: small or medium) × 2 (number of sample size: 300 or 1000) × 2 (number of repeated measures: 6 or 10) × 2 (mixing percentages: 50%:50% or 75%:25%) factorial

design to generate data. A total of 200 replications were generated for each condition using R (2013), yielding a total of 3,200 (16 conditions $*$ 200 data sets) data sets. Each replication was fitted respectively using the ML/EM method under Mplus Version 7.2 (Muthén & Muthén, 2012) and the Bayesian method under JAGS (Plummer, Stukalov, & Denwood, 2015). In addition, within the Bayesian estimation framework, we compared and measured the impact of three levels of prior distribution on the posterior estimates.

3.2.2 Outcomes

3.2.2.1 Hit Rate

Hit rate is the percentage of correctly classified individuals in the two classes. Both estimation methods were evaluated by comparing their respective hit rate averaged across all replications for each condition.

3.2.2.2 Standardized Bias

The turning point, the mean growth parameter and their corresponding random effect parameter estimates from both the ML/EM and the Bayesian methods were summarized across all replications for each condition. The standardized biases of the estimates [i.e., $B(\hat{\theta}) = (\hat{\theta} - \theta) / S(\hat{\theta})$] [2] were calculated. The mean of the standardized bias is equivalent to a Cohen's d, which measures the standardized distance between the estimate and the parameter. Based on the guidelines for Cohen's *d*, the value of less than .4 is considered acceptable.

[2] Where $\hat{\theta}$ is the parameter estimate, θ the population parameter value, and $S(\hat{\theta})$ the standard deviation of the estimates across 200 replications.

3.2.2.3 Coverage

The coverage of a confidence interval is the proportion of replications whose 95% confidence interval includes the specified true parameters. According to Burton, Altman, Royston, and Holder (2006), the coverage rate of a 95% confidence interval should be approximately equal to .95, with a margin of error of .03[3]. Put it differently, there should be between 92% and 98% of the replications whose confidence intervals cover the true parameter. A coverage rate greater than 98% indicates more replications fail to detect significant results thus leading to decreased power or inflated Type II error rate. A coverage rate lower than 92% indicates more replications have incorrectly identified significant estimates, which resulted in inflated Type I error rate. The confidence interval was only computed for fixed effects.

3.2.3 Analysis

Analysis of variance (ANOVA) was used to partition the total variation in the standardized bias, coverage, and hit rate to determine the effects of the design factors on these three outcome measures. In each estimation method, individual parameter estimate were analyzed with a separate factorial ANOVA. Given that the purpose of using ANOVA in the present study was descriptive rather than inferential, the *p* value of the F-test was not reported. Instead, the eta-squared (η^2) effect size[4] was computed and reported as a measure of practical significance. Effects were considered substantial with the semi-partial eta-squared greater than 0.1.

---

[3] Using the formula provided by Burton, Altman, Royston, and Holder (2006), margin of error = $1.96\sqrt{p(1-p)/B}$, where p is the nominal coverage rate (p=.95) and B is the number of replications (B=200).

[4] The eta-squared effect size is computed by $\eta^2 = SS_{effect} / SS_{total}$.

3.3 Results

3.3.1 Convergence

Convergence has always been a challenging issue with finite mixture models (Chen, 1995; G. McLachlan & Peel, 2004). We examined each replication for convergence problems to determine if it should be included in further analyses. According to Chen et al. (2010), replications that have fewer than 2 individuals in any of the latent classes should be treated as improper results and be excluded from further analysis. We did not find such a problem in our replications, and across all replications under both ML/EM and Bayesian methods, each class had been assigned an adequate number of individuals. Another concern over the convergence issues with GMMs estimated by the ML/EM method was replications that gave negative error variance (i.e., *Heywood cases*) (Liu & Hancock, 2014; Tofighi & Enders, 2008; Tolvanen, 2007). In our study, 14% of the overall replications were found to bear *negative error variance estimates* across all replications. However, the estimates were found to be not statistically significantly different from zero, and were most likely induced from sampling fluctuations (Gerbing & Anderson, 1987; Dillon, Kumar & Mulani, 1987). Therefore, we considered it appropriate to include those replications for further analysis. For the Bayesian approach, the Geweke and the Heidelberg-Welch diagnostics were used to examine convergence problems. A replication was considered converged if the posterior distribution of each parameter passed both convergence tests. For the correctly specified two-class models, the non-convergence rate is around 25%. A high convergence rate up to 91 % was found in the condition of small sample size (N=300),

balanced class proportion, and 6 time points. Holding other factors constant, the change from balanced class proportions to unbalanced ones incurred higher rate non-convergence.

3.3.2 Hit Rate

The results showed that for the ML/EM estimation, the percentage of correct classification of individuals ranged from 84% to 94%, while the range for the Bayesian estimation was from 86% to 97%. Under the Bayesian method, there was a very high hit rate with informative priors (from 90% to 100%), followed by the data driven priors (from 85% to 97%), and the weakly informative priors (from 81% to 95%).

The ANOVA results indicated that under the ML/EM estimation, the design factors of class separation and sample size were found to have substantial effect on the hit rate; while for the Bayesian estimation, the class separation, sample size, and prior specification were found to have substantial effect. Table 2 showed mean hit rate broken down by the three factors. As expected, for both estimation methods, the hit rate was higher under the high separation condition than the low separation condition. Specifically, for the ML/EM estimation, as the class separation changed from low to high, the hit rate increased from 86% to 93%, and for the Bayesian estimation, the hit rate increased from 89% to 93%. Also, for both estimators, the hit rate increased with the increase in sample size.

3.3.3 Standardized Bias

Figure 1 and Figure 2 showed the mean standardized biases of the parameter estimates under both ML/EM and Bayesian approaches across all the conditions along

with the Monte Carlo 95% confidence intervals of the means. The lower and the upper limit of the 95% confidence interval were set at the 2.5th and 97.5th percentiles of the sampling distributions of the mean standardized bias of the parameter estimates from all conditions considered.

3.3.3.1 Standardized Bias of Fixed Effect Estimates

Overall, under both approaches, the estimates were considered acceptable with negligible biases for the intercept (a) and the first slope (b1) of the both classes. The mean standardized bias for the turning points (γ) in class one and class two was 0.16 and -0.34 respectively under the ML/EM approach, and the values were 0.09 and -0.23 under the Bayesian approach. Large biases were found with estimates of the second slope (b2), and the ML/EM approach recovered the parameter with a smaller bias (-0.28) than its counterpart, the Bayesian approach (-0.45). ANOVA showed that under the ML/EM estimation, class separation had significant impact on the estimates of the turning point for class-two (η^2 =0.45) and the second slope in each latent growth curve (η^2 ranges from 0.51 to 0.53); while under the Bayesian method, class separation and the specification of the prior distributions showed significant impact on the second slope for each class (ranges from 0.46 to 0.49).

3.3.3.2 Standardized Bias of Random Effect Estimates

Under both estimation methods, the random effect estimates were highly biased. The level of bias was particularly problematic for the variance estimate of the intercept, with the ML/EM method producing a much higher degree of bias than the Bayesian method. Under both estimation methods, with exception to the variance estimate of the

intercept for each class, the estimated random effects were negatively biased across all conditions.

ANOVA results showed that under the ML/EM approach, the number of time points had significant impact on all random effect estimates (η^2 ranges from 0.31 to 0.52), and class separation was found to have substantial impact on the estimates of the first and second slope for class two (η^2 =0.4 for both). Under the Bayesian method, except for the first slope, the estimation of the random effects varied significantly depending on the prior distribution specifications (η^2 ranges from 0.11 to 0.56), class separation (η^2 ranges from 0.09 to 0.43), and sample size (η^2 ranges from 0 to 0.52).

3.3.3.3 Standardized Bias of Residual Variance Estimates

The mean standardized bias of the residual variance estimates for the ML/EM method is 0.16 while the corresponding value for the Bayesian method is 0.7. For the ML/EM approach, the number of time points (η^2 =0.39) and class separation (η^2 =0.44) were found to account substantially for the variance in the bias of the residual. Under the Bayesian method, the class separation (η^2 = 0.35) and sample size (η^2 = 0.43) were the most influential factors.

Table 2

Mean Hit Rate for Bayesian and ML/EM Estimations

Impact Factors			Hit Rate			
Mixing %	Separation	Sample Size	ML/EM	Bayesian (Informative Prior)	Bayesian (Data Driven Prior)	Bayesian (Weakly Informative Prior)
50%:50%	High	300	91%	91%	85%	83%
		1000	93%	99%	96%	95%
	Low	300	84%	90%	85%	81%
		1000	87%	95%	92%	92%
75%:25%	High	300	92%	95%	90%	85%
		1000	94%	100%	97%	94%
	Low	300	85%	92%	90%	82%
		1000	88%	96%	93%	91%

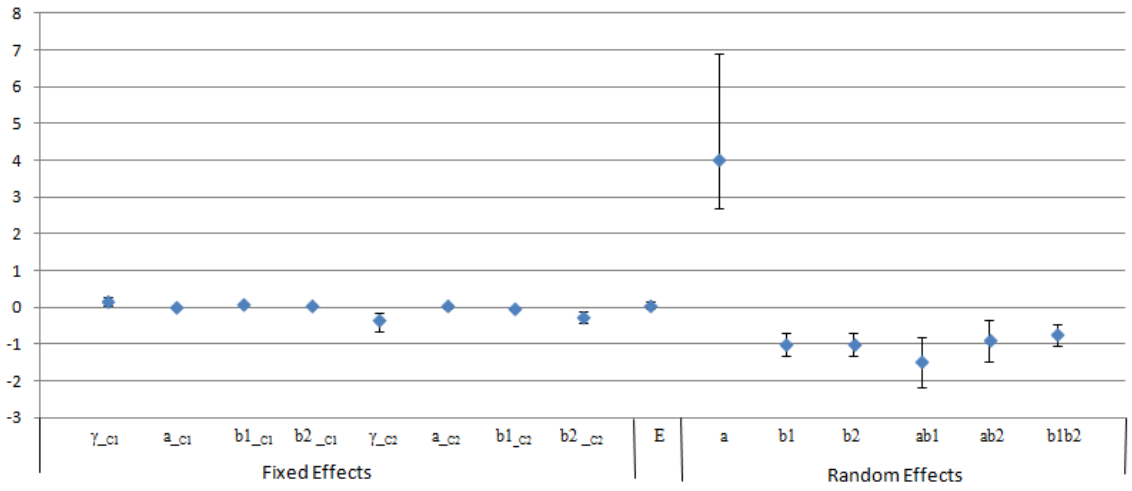

Figure 1 Monte Carlo simulation 95% confidence intervals under the ML/EM method

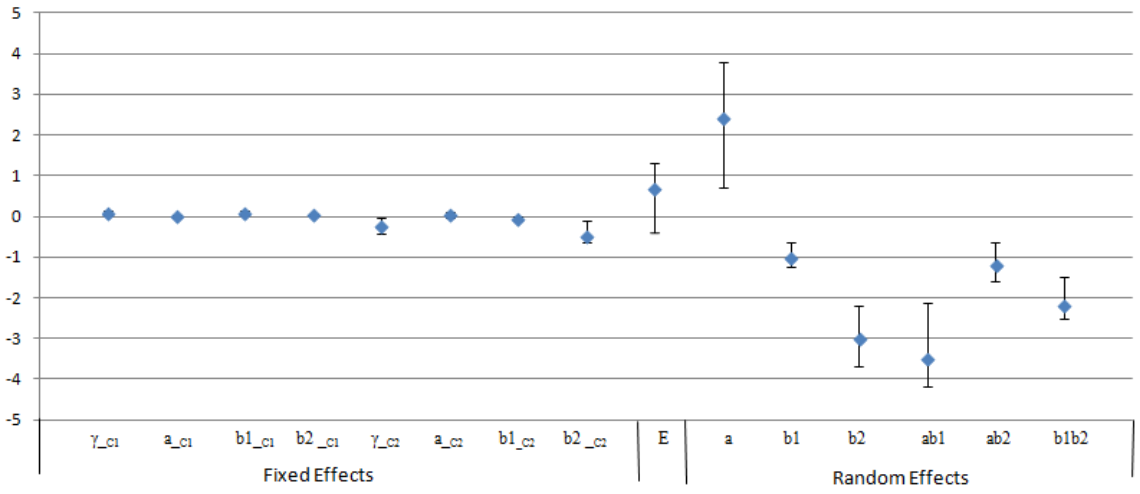

Figure 2 Monte Carlo simulation 95% confidence intervals under the Bayesian method

3.3.4 Coverage

As shown in Figure 3, under both estimation methods, for Class 1, coverage rates lied within the acceptable range between 92% and 98% across all conditions. For Class

2, the coverage rates of the turning point (γ) under both estimations and the slope of the second piece (b2) under Bayesian estimation showed a slight deviation from the acceptable range. Overall, compared to the Bayesian method, the ML/EM estimation gave higher coverage rates for parameter estimates of both classes. ANOVA results showed that for the ML/EM estimation, no design factor was found to have a significant impact on the coverage rate.

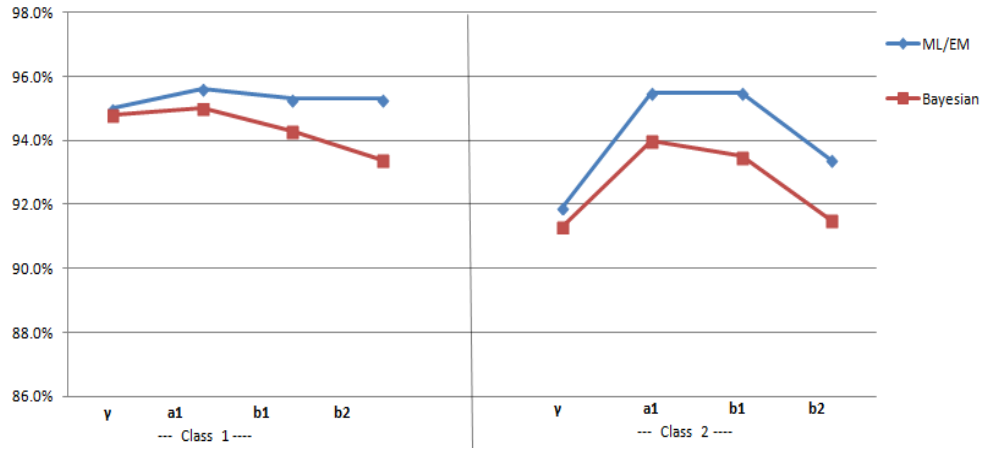

Figure 3 Mean coverage rates of fixed effect estimates for both estimations across all conditions

For the Bayesian approach, the prior specifications had critical impact on the coverage rate, particularly with respect to the estimation of the turning point (η^2 =0.31) and the second slope (η^2 =0.28) for class 2. Figure 4 presented a comparison of the mean coverage rates of the fixed effect estimated under the ML/EM method and the Bayesian method with informative priors. It could be noted in Figure 4 that with informative priors, the coverage rate of the parameter estimates in each class lied within

an acceptable range for the Bayesian method and the coverage rates of parameter estimates in class 2, particularly with the turning point and the second slope, was better than the ML/EM method.

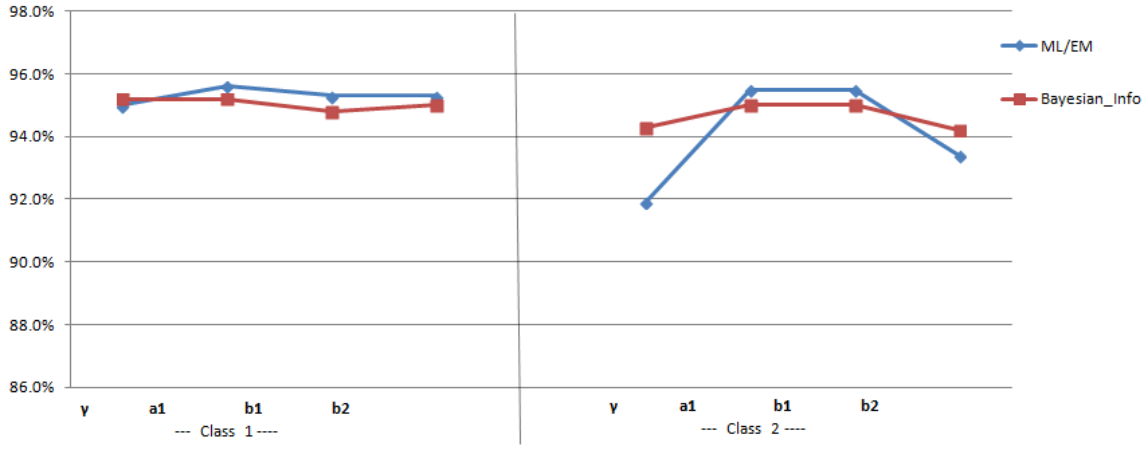

Figure 4 Mean coverage rates of fixed effect estimates for both the ML/EM and Bayesian with informative priors for the mean growth factors

When using data driven priors, both estimation methods produced acceptable coverage rate of parameter estimates in Class 1. However, as shown in Figure 5, under the Bayesian method, the coverage rate of the slope of the second piece (b2) in class 2 was below the acceptable range. The same pattern was observed in Figure 6, which showed the comparison of the coverage rates between the ML/EM method and the Bayesian method with weakly informative priors. The coverage rates of the second piece slope (b2) in class 1 and the turning point (γ) and b2 in class 2 lied far below the acceptable rage under the Bayesian method with weakly informative priors.

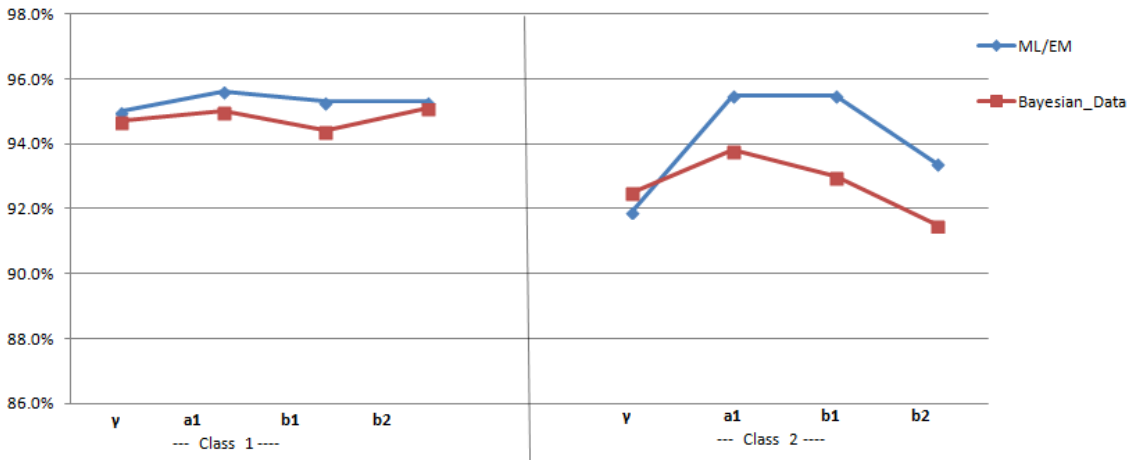

Figure 5 Mean coverage rates of fixed effect estimates for both the ML/EM and Bayesian estimation with data driven priors for the mean growth factors

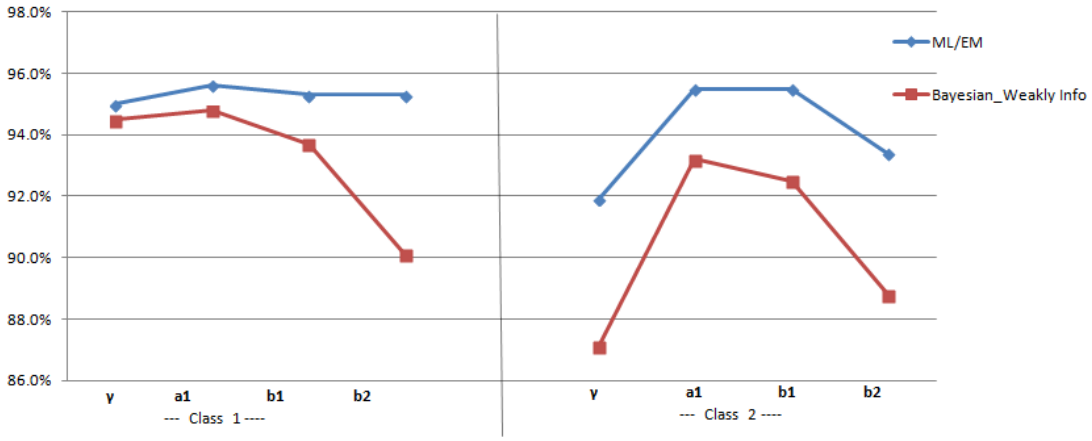

Figure 6 Mean coverage rates of fixed effect estimates for both the ML/EM and Bayesian estimation weakly informative priors for the mean growth factors

3.4 Discussion

This study examined and compared the performance of the ML/EM method and the Bayesian method in estimating PGMM with unknown turning point in terms of the accuracy of individuals 'class assignment and parameter estimates. Four design factors were considered, that is, number of time points, sample size, class proportions, and class

separations. For the Bayesian method, different specifications of prior distributions for the mean growth factors were considered.

3.4.1 Parameter Estimates

Both estimation methods have almost no difficulty in recovering the true parameters of class one. For the true parameters of class two, however, only the Bayesian approach with informative priors can recover them to an acceptable degree. The ML/EM method and the Bayesian approach with data driven priors can give considerably good estimates of the turning point, the intercept, and the first slope for class two, but fail to do so with regard to the true parameter of the second slope for the second class. Comparatively speaking, the Bayesian approach with weakly informative priors performs the worst in recovering the true parameters of class two.

Random effect estimation presents a challenge for both the Bayesian and the ML/EM approaches, although a specification of informative prior for a Bayesian PGMM could help alleviate the problem to some degree. Both approaches tend to overestimate the variances of the random effects, with exception to the intercept, which was largely underestimated. The results are in line with the findings in Depaoli's (2013) study. In her study, data were generated using a 3-class model, with the third class model adding a quadratic term over and above the linear trajectories in the first two classes. The results of her study showed that the estimated variances of the random effects associated with the slope of the linear trajectory in the second-class and the slope of the quadratic term in third class showed large bias in both ML/EM method and Bayesian estimation with different levels of prior specifications. The biases in the random effects could have led to

biased estimates in the standard errors of the fixed effects, which in turn lead to the undercoverage or overcoverage of their confidence intervals.

3.4.2 Impacts of the Design Factors

The impact of the design factors on the model performance varied depending on which estimation method to use. Overall, for both estimation methods, class separation determined the accuracy in individuals' class assignment, the level of biases, and the degree of undercoverage. As expected, well defined classes of high separation make it easier to recover the true parameters and to accurately assign individuals to the correct class.

Sample size and the number of time points also have great impacts on the outcomes. The large number of time points (i.e., 10 in our study) combined with a small sample size (N=300) hampered the recovery of the random effects under both estimation methods. The problem became more serious with the Bayesian method, as the priors for the random effects were non-informative. The large variability associated with non-informative priors in combination with a small sample size and a large number of time points could lead to random effect estimates way off their true values.

Our findings regarding the Bayesian estimation are consistent with the extant literature (Kass & Wasserman, 1996; Depaoli, 2013). It has been noted that priors, particularly with noninformative priors, can have a substantial impact on parameter estimates, especially so when the sample size is small (Lambert, et al., 2005). Small sample size, coupled with poor class separation will magnify the impact of the inaccurate prior specifications on parameter estimation, resulting in less accuracy in

individual assignment and true parameter recovery. The results of our study again highlighted the importance of including accurate prior information in modeling analysis of a turning point PGMM. An inclusion of information that summarizes the previous studies into the Bayesian model specification could weaken the detrimental effect of small sample size, and thus lead to a comparatively increased power to produce more accurate parameter estimates (Gelman, Carlin, Stern, & Rubin, 2004; Depaoli, 2013).

3.4.3 Implications

For empirical researchers whose research interests lie in the identification of distinctive turning points that differentiate one change process from the other, particularly associated with the intervention effect, a turning point PGMM is of particular importance as it allows for an accurate estimation of the turning point. A complete reliance on theory to determine a turning point a priori when specifying a PGMM could lead to a mis-presentation of the growth trait in the data. However, when using the Bayesian approach to estimate a turning point in PGMM, the accuracy of the growth trajectory estimates (including turning points and growth factors), and the latent class assignment depend on the level of accuracy in the information that we could provide to the Bayesian model estimation. For the ML/EM approach, valid and good estimates could be produced given that the classes are well separated, plus an adequate number of time points and sample size. For example, in our study, the data scenario that is a combination of high class separation, 6 time points, and 1000 individuals recovered the trajectory shapes reasonably well.

CHAPTER IV

STUDY TWO: CLASS IDENTIFICATION EFFICACY IN PIECEWISE GROWTH MIXTURE MODELS WITH UNKNOWN TURNING POINTS: A COMPARISON OF ML/EM VERSUS BAYESIAN ESTIMATION

4.1 Overview

Piecewise Growth Mixture Modeling (PGMM) is an extension of conventional Growth Mixture Models (GMM) to accommodate the developmental change in longitudinal data that exhibits differential growth phases. On top of the capabilities to estimate the heterogeneity of growth trajectories in the population and the trajectory-specific variance components, a PGMM allows for the incorporation of turning points (or change points or knots) in the functional form, the existence of which indicates a shift in the development from one growth phase to another. Longitudinal data collected over a long interval of time often exhibit nonlinear trend. The presence of a turning point renders a PGMM an inherently nonlinear function that can more precisely describe the change patterns in data across time intervals (Harring, Cudeck, & du Toit, 2006; Kwok, Luo, & West, 2010), than conventional GMM which is restricted to the linearity assumption.

The specification of traditional piecewise growth functions usually relied on the researchers' substantive knowledge to determine a priori a turning point that connects different growth phases. For example, Hardy & Thiels' (2009) study determined a turning point to be at the time point where treatment sessions ended and post-treatment sessions began, hypothesizing that individuals under assessment exhibited different growth rate during treatment than post treatment. McAuley et al. (2011) used PGMM to

examine the differential effects of randomized controlled exercise trial on self-efficacy and identified three growth trajectories, each exhibiting distinctive growth rate before and after intervention. In these two studies, the turning points were set at the time of intervention. However, such consideration may not always be reasonable, because the turning point may occur after the intervention due to delay in response to intervention. Moreover, no attempt was found in applied studies to test the location of the turning point hypothesized based on theoretical considerations. A misspecification of a turning point may give a suboptimal functional presentation of the observed data patterns over the course of time, leading to misleading modeling results and statistical inferences of growth traits.

A more realistic alternative is to estimate unknown turning points based on data. There have been several statistical procedures available for such realization. For example, Kwok et al. (2010) proposed to use modification index to detect the turning point in the linear latent growth modeling framework. Dominicus, Ripatti, Pedersen, & Palmgren (2006) and Wang & McArdle (2008) presented statistical modeling approaches in the context of mixed effect models in which the turning point is a parameter to be estimated and is allowed to have inter-individual variability. Recently, a two-stage (or two-piece) piecewise growth mixture model with one unknown change point was proposed by Kohli, Harring, & Hancock (2013) in the framework of Structural Equation Modeling (SEM).The newly proposed PGMM with unknown turning points is more advantageous than conventional PGMMs as it offers applied researchers an opportunity to determine a turning point empirically as well as theoretically.

Furthermore, to freely estimate a turning point and the factor loadings of the growth factors can oftentimes give a more optimal functional form describing the observed data patterns over time (Kwok, Luo, & West, 2010; Wood & Jackson, 2013).

However, many questions regarding the performance characteristics of a PGMM with unknown turning points still remain to be examined. One of the questions of primary interest is the performance of PGMMs in determining the correct number of growth trajectories. The aim of the current study was to examine the accuracy of commonly used enumeration indexes with PGMMs in growth trajectories enumeration. In particular, comparisons were made between the two dominant estimation methods: the Bayesian estimation framework via the Markov Chain Monte Carlo (MCMC) algorithm and the Maximum likelihood estimation via the expectation maximization (EM) algorithm (ML/EM) in the SEM framework.

4.2 Methods

A 2-class PGMM [i.e., model (2)] was used as the population model in the study, with both classes exhibiting a two-piece linear-by-linear growth curve. The number of classes was decided based on past empirical studies using PGMM (Li, et al., 2001; Kohli, et al., 2013; Zhao & Banerjee, 2012). Table 1 presents the population parameters for the average growth models and the covariance components of random effects for the two trajectories.

4.2.1 Design Factors

Previous simulation studies on Growth Mixture Models have shown consistent findings upon what factors could impact class enumeration (Q. Chen, Kwok, Luo, &

Willson, 2010; Enders & Tofighi, 2008; Henson, Reise, & Kim, 2007; Tofighi & Enders, 2008). Based upon the previous findings, five design factors were considered in the study, including (a) the degree of trajectory separation, (b) the number of repeated measures, (c) mixing percentages of latent classes, (d) sample size, (e) specification of prior distribution.

4.2.1.1 Degree of Class Separation

We varied the degree of class separation by manipulating the within class variance components. The larger the within class variance is, the smaller the degree of class separation is. According to Raudenbush and Liu's (2001) criteria, a small size of variances and covariance matrix of the random effects is set to be as follows for the high separation condition:

$$T_\pi = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .1 & & \\ .025 & .05 & \\ 0 & .0175 & .05 \end{bmatrix}$$

Given $\tau_{\pi 00}$ = .1 and $\sigma^2 = 1$, the difference between the intercept for Class 1 and that for Class 2 was approximately 2.4 standard deviation units. This was considered as the high separation condition according to Tofighi and Enders' (2008) study. In the "low separation" condition, the within-class variance components were specified as below:

$$T_\pi = \begin{bmatrix} \tau_{\pi 00} & & \\ \tau_{\pi 10} & \tau_{\pi 11} & \\ \tau_{\pi 20} & \tau_{\pi 21} & \tau_{\pi 22} \end{bmatrix} = \begin{bmatrix} .7 & & \\ .175 & .35 & \\ 0 & .1225 & .35 \end{bmatrix}$$

which results in approximately 1.9 standard deviation units difference between the intercept for Class 1 and Class 2. In both high and low separation conditions, $\tau_{\pi 11}$ and $\tau_{\pi 22}$ were set to be half of the size of $\tau_{\pi 00}$, because the variation of the intercept has

generally been larger than the variation of the growth trends in longitudinal data. The size of the covariance, $\tau_{\pi 10}$ and $\tau_{\pi 21}$ was set to be .025 and .0175 in the high separation condition, and .175 and .1225 in the low separation condition so that there was a moderate correlation (ρ=0.35) between the intercept and growth factors in the first phase and between the growth factors in the first and second phases as well.

4.2.1.2 Sample Size

The sample sizes were chosen based on the conditions used in the past simulation studies (Q. Chen et al., 2010; Enders & Tofighi, 2008; Henson et al., 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008) and a review of substantive GMM and PGMM applications. A total of 346 studies were found from a literature search in PsycINFO (from 2000-2013) for studies applying GMM and PGMM in different substantive areas. A review of a random sample of 100 out of the 346 studies showed that the studies varied dramatically with respect to sample size. We chose the sample size values to be 300 and 1000, representing approximately the 25th and 75th percentiles of the sample size distribution in the 100 studies that we reviewed.

4.2.1.3 Number of Repeated Measures

It is more typical for a two-piece linear growth model to include more than two time points in each piece, with 5 points in total enabling a full growth model (fully random) to be estimated (Bollen & Curran, 2006). The use of fewer time points restricts the ability to estimate the full set of random effects (i.e., variances and covariance for intercepts and slopes). Hence we chose 6 waves of repeated measures as the small number. We chose 10 waves of repeated measures as the medium number since the

mean number of waves among the empirical GMM studies that have more than 6 measurement waves is about 10.

4.2.1.4 Mixing Percentage of the Two Classes

Following the simulation studies in Chen, et al. (2010), Nylund et al. (2007), and Liu & Hancock (2014), we considered two sets of mixing percentages of the two classes. The mixing percentages in the balanced condition were 50% and 50% for the two classes. In the unbalanced situation, we considered 75% vs. 25% for the two groups respectively. The mixing percentages in the unbalanced situation correspond to the mean mixing percentages in the empirical studies using PGMM (Zhao & Banerjee, 2012; Li, et al. 2001; Kohli, et al. 2013).

4.2.1.5 Specifications of Prior Distributions

The impact of three classes of priors on Bayesian posterior estimates were considered in the study: weakly informative priors, informative priors, and data driven priors with the priors specified based upon estimates from the ML/EM estimation. The three classes of priors were specified following Depaoli's (2013) study. For each of the normally distributed growth parameters (i.e., $\mu_{a_k} \sim N(\mu_a, \sigma_a^2)$, $\mu_{b1_k} \sim N(\mu_{b_1}, \sigma_{b_1}^2)$, $\mu_{b_{2k}} \sim N(\mu_{b_2}, \sigma_{b_2}^2)$, and $\mu_{\gamma_k} \sim N(\mu_r, \sigma_\gamma^2)$), there are mean and variance hyperparameters. For example, μ_a and σ_a^2 are the respective mean and variance hyperparameters for parameter *a* (i.e., the intercept mean factor). The hyperparameters determine the degree of certainty a prior has on the posterior estimate of the parameter value. To specify an informative prior, the mean hyperparameter was set to be the corresponding growth parameter population value, while the variance hyperparameter was set to be at 5% of

the growth parameter population value. For example, the mean hyperparameter of the intercept of Class 2 was set to be 25 and the variance hyperparameter was set to be 1.25 (25*5%). Therefore, the informative prior for the intercept in Class 2 is *N* (25, 1.25).

Weakly informative priors were specified by taking the population value as the mean hyperparameter but 50% of the population value as the variance hyperparameter [i.e., *N* (25, 12.5)]. The weakly informative and informative priors for the rest growth parameters were determined in a similar way. As for the data driven priors, the ML/EM parameter estimate of each growth factor (i.e., the mean and variance) in all replications for each data scenario were obtained as the mean hyperparameter and the variance hyperparameter for the growth factors.

The above were the specifications of prior distributions for the mean growth factors in the correctly specified 2-class model. For the mis-specified under extracted 1-class model, the priors were specified based on the average of the population values of the mean growth factors of the two classes. For example, to specify an informative prior to the intercept of the mis-specified 1-class model, the mean hyperparameter was the average of the intercept of class one (i.e., 22.5) and the intercept of class two (i.e., 25), which is 23.78; and the variance hyperparameter was set be 1.19 (23.78*5%). Therefore, the informative prior for the intercept in mis-specified 1-class model is *N* (23.78, 1.19). Follow suit, we would have the informative priors and the weakly informative priors for the mean growth factors in the mis-specified 1-class model. The data driven priors were based on the ML/EM parameter estimates of the 1-class model.

For the mis-specified over extracted 3-class model, the priors for the mean growth factors of first two classes were the same as how the priors were set up for the mean growth factors of the two classes in the true 2-class model. The priors for the mean growth factors of the third class were specified based on the average of the population values of the mean growth factors of the two classes in the true 2-class model, the same as how prior distributions for the mean growth factors in the under extracted mis-specified 1-class model were specified.

Combining all design factors, the simulation used a 2 (magnitude of the matrix: small or medium) ×2 (number of sample size: 300 or 1000) × 2 (number of repeated measures: 6 or 10) × 2 (mixing percentages: 50%: 50% or 75%: 25%) factorial design to generate data. A total of 200 replications were generated for each condition using R (2013), yielding a total of 3,200 data sets(16 conditions 200 data sets). For each replication, three different models (i.e., one-, two-, & three-class models) were fitted using ML/EM algorithm under Mplus Version 7 (Muthén & Muthén, 2012) and Bayesian estimation under JAGS (Plummer, Stukalov, & Denwood, 2015) respectively. Also within the Bayesian estimation, three levels of prior specifications were compared for the respective impact on the accuracy of class enumeration.

4.2.2 Outcomes and Analysis

The primary outcome of interest in the study is the percentage of replications that accurately retrieved the correct number of trajectories based on a particular enumeration index. The average percentages of one-class, two-class, and three-class models identified by a particular enumeration index were summarized. Based on the information, we can

tell whether any indices are more liable to under- or over-extract the number of trajectories. Analysis of Variance (ANOVA) was conducted to investigate the impact of the design factors on the class enumeration accuracy of the index.

4.3 Results

4.3.1 Convergence

Convergence has always been a challenging issue with finite mixture models (Chen, 1995; G. McLachlan & Peel, 2004). The convergence problem under ML/EM and Bayesian estimation were explicitly examined to ensure a clear and appropriate analysis of the results. Under the ML/EM estimation using Mplus Version 7.2, replications that gave proper solutions were counted as converged results. Agreement diverged upon whether a replication with a negative error variance (i.e., Heywood cases), should be considered as a converged result. Liu & Hancock (2014) excluded replications as non-converged results that produce inadmissible solutions, such as negative error variance (i.e., a Heywood case), whereas Tolvanen (2007) considered negative error variance as a normal variation of sampling. Many others did not consider Heywood cases at all (e.g., Peugh & Fan, 2012). In the present study, the overall replications with *negative error variance estimates* were about 18% across all replications and all three models. A further examination of the replications that bore negative error variance estimates showed that the estimates were statistically insignificant from zero, and the problem of Heywood cases was not a concern as they were most likely induced from sampling fluctuations (Chen, 1995; Dillon, Kumar, &

Mulani, 1987; Gerbing & Anderson, 1987). Therefore in our study, we considered Haywood cases as converged results and included those replications in our final analysis.

For the correctly specified two-class models, all of the 8000 replications had converged results. For the one-class models, the overall convergence rate is 75%. Apparently, forcing too few classes onto the data has led to a higher number of replications that failed to converge. In addition, the non-convergence occurred more often when the sample size was small (N=300) and the class separation was low. The non-convergence rate decreased when the sample size increased and when the class separation changed from low separation to high separation. For the three-class models, the overall convergence rate was 96.5%. Again, the non-convergence rate occurred most often with small sample size (N=300) and low class separation. In replications where the three-class failed to converge, instead of continuing to fit in the data with four-, five-, or six-class models (e.g., Liu & Hancock, 2014), we chose to fit a three-class model at most and to retain and interpret the two-class model provided that the index was found to show preference toward the two-class model (e.g., Tofighi & Enders, 2008).

Under the Bayesian estimation, the Raftery and Lewis diagnostic (1992) was first used to calculate the number of iterations and the number of burn-in iterations necessary for convergence within each data scenario. Results from the Raftery and Lewis convergence diagnostic indicated that the convergence with 2 latent classes could be obtained after running a minimum number of 40,000 iterations, and 10,000 of which were burn-in iterations. To be conservative, for each replication, a single chain was run with 40,000 iterations, a burn-in of 10,000 iterations, and a thinning interval of 10,

resulting in 4000 Gibbs samples that comprised the posterior distributions for the parameters. The convergence of the Markov chain was again evaluated by the Geweke (1992) statistic and the Heidelberger-Welch stationarity test (Heidelberger & Welch, 1981; Heidelberger & Welch, 1983). The Geweke convergence diagnostic test compares the mean estimates from two non-overlapping parts, usually from the early 0.1 and the latter 0.5 proportions of the Markov chain. The mean difference test is a two-sided test based on a z-score statistic. When an absolute z value is larger than 2, it indicates non-convergence. The Heidelberger-Welch stationarity test examines whether the Markov chain is from a stationary distribution. The test consists of two parts. In part one, the test statistic calculates the stationarity by discarding the first 10%, then 20% up to 50% until the null hypothesis of stationarity is not rejected; the convergence fails if after 50% of the data has been discarded, the null hypothesis is still rejected. Part two is a half-width test following the pass of the test in Part one. The test statistic is the ratio of the half-width of the 95% credible interval to the mean. The test fails if the test statistic is greater than 0.1. For each replication, the convergence test results were obtained from both the Geweke and the Heidelberg-Welch diagnostics and examined for convergence problems. The replication is considered converged if the posterior distribution of each parameter passed both convergence tests.

Overall, the non-convergence rate for the one-class under-extracted model is 48%. The rate is higher for data sets with low class separation and with 10 time points, and the non-convergence rate is particularly higher when the sample size is large (N=1000). For the correctly specified two-class models, the non-convergence rate is

around 25%. A high convergence rate up to 91 % was found in the condition of small sample size (N=300), balanced class proportion, and 6 time points. The convergence rate decreased when sample size increased from 300 to 1000 and the number of time points from 6 to 10. Holding other factors constant, the change from balanced class proportions to unbalanced ones incurred higher rate non-convergence. For the three-class over-extracted models, the non-convergence problem was serious. On average around 60% of all replications were not converged. Non-convergence problems occurred more often in replications with a small sample size in combination with low class separation and a large number of time points. A similar stance was taken here as in the ML/EM estimation: for replications that failed to converge with three-class models, we fit three-class modes at most, and to retain and interpret the two-class models given that the index showed preference for them.

4.3.2 Comparing Overall Performance of the Enumeration Indices

Figure 7 shows the average percentages of one-class, two-class, and three-class models identified by the AIC, AICC, CAIC, BIC, SABIC, NEC, ICL-BIC, CLC, and BLRT for all converged replications estimated by the ML/EM method. As shown in the figure, all model enumeration indices were able to identify the two-class solution (i.e., the correct model) in most of the replications. Specifically, the ICL-BIC had the highest percentage of correct identification (97%), followed by the CAIC (91%), DBIC (85%), BIC (84%), BLRT (74%), AICC (73%), AIC (57%) and CLC (65%). The NEC exhibited the least utility as it tended to overestimate the number of classes, giving a

higher overall percentage in identifying the three-class solution (46%) than the corresponding two-class solution (34%).

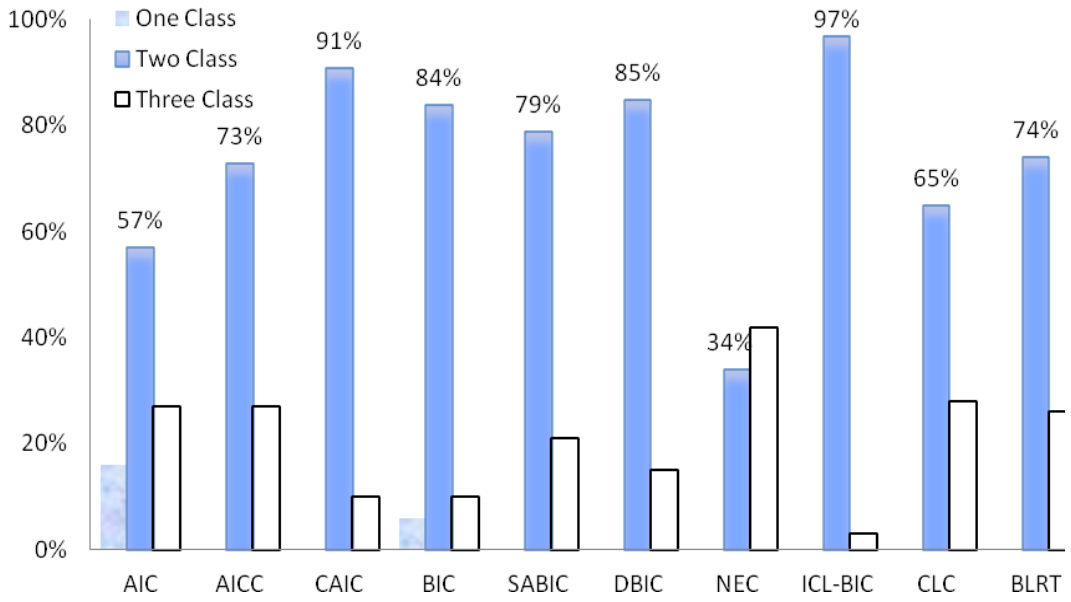

Figure 7 Percentage of 1-, 2-, and 3-class models identified by enumeration indices under the ML/EM estimation

Figure 8 shows the average percentages of one-class, two-class, and three-class models identified by the AIC, WAIC2, BIC, DIC3, LPML, and PsBF across all converged replications estimated by the Bayesian method. All model enumeration indices were able to identify the two-class solution (i.e., the correct model) for the majority of the replications. On average, the DIC3 was able to identify the correct model most frequently, up to 93% of all replications, followed by the WAIC2 (91%), BIC (85%), AIC (75%), LPML (73%), and PsBF (69%). The AIC, PsBF and LPML were less accurate in identifying the correct model. Comparatively speaking, the AIC, LPML,

and PsBF exhibited a higher tendency to over-extract classes than the other indices such as the WAIC2, DIC3, and BIC.

Overall, when using the ML/EM estimation method, ICL-BIC and CAIC were the most effective measures in identifying the correct number of trajectories in PGMM. When using the Bayesian method, WAIC2 and DIC3 were the most accurate indices followed by BIC.

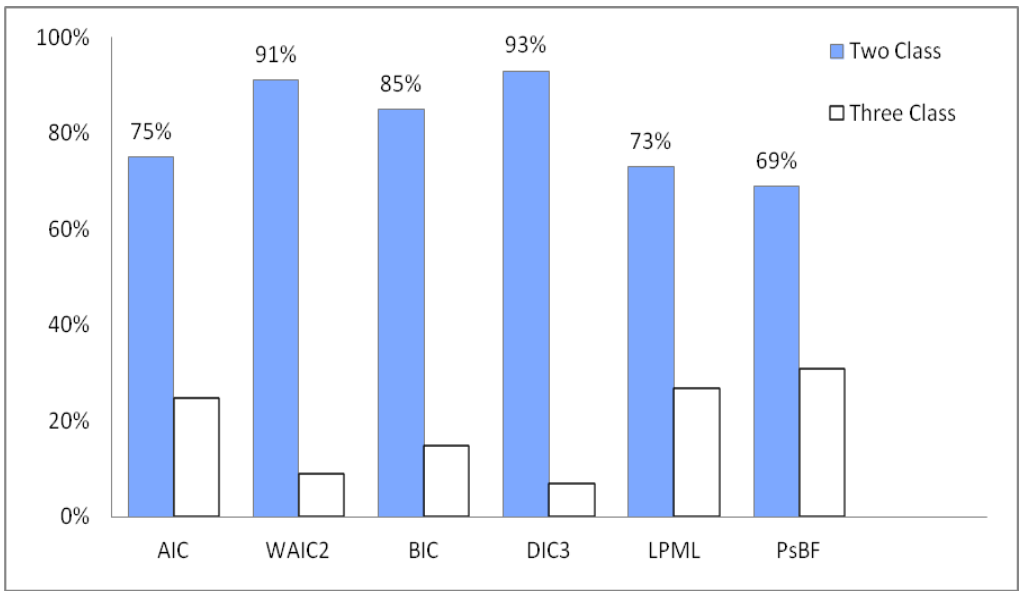

Figure 8 Percentage of 1-, 2-, and 3-class models identified by enumeration indices under the Bayesian estimation.

4.3.3 Impact of the Design Factors on Class Enumeration Accuracy

Full factorial ANOVAs were conducted to determine the impact of the four design factors (or five factors under the Bayesian estimation) on the performance of the enumeration indices in class identification accuracy. The effect size indicator, the semi-partial eta-squared (i.e., $\eta^2 = SS_{Effect} / SS_{Total}$), was computed to measure the impact of each

design factor on the class enumeration performance of a particular index. Table 3 and Table 4 presented the ANOVA results under the ML/EM and Bayesian framework respectively. A statistically significant effect (p<.05) of the factors on the indices was marked with an asterisk, and a semi-partial eta-squared larger than 0.1 was bolded. The interaction effects were found to be non-significant and thus were removed from the tables for brevity.

As presented in Table 3, under the ML/EM framework, class separation and the number of time points were found to have significant impact on the enumeration performance of certain indices (i.e., the AIC, ICL-BIC, AICC, SABIC, CLC, and BLRT), while class proportion and sample size had not shown any significant impact. As shown in Table 4, when using the Bayesian estimation method, the prior specification was a significant factor that had critical impact on the accuracy of some indices, such as the AIC, LPML, and PsBF. Class separation was also found to have substantial bearing on the indices of the BIC and PsBF. The number of time points, sample size, class proportion had shown no substantial impact on class enumeration accuracy.

4.3.3.1 Class Separation

Table 5 showed the percentage of replications that correctly identified the two-class population model under the ML/EM framework, broken down by the degree of class separation and time points. Table 6 presented similar information under the Bayesian method, broken down by the type of prior, degree of class separation, and time points. In both tables, the percentage over 90% was bolded.

Under the ML/EM estimation, because CAIC and ICL-BIC are the best performing indices, we focus on these two indices and examine how class separation affects their performance. When two classes were well-separated, the CAIC, ICL-BIC, and BIC were very accurate measures in identifying the correct model. However, when the class separation was poor, the ICL-BIC overwhelmingly outperformed the other indices, identifying the correct model in 94% of the replications. The CAIC was not severely impacted by class separation either, choosing the correct model 99% of the time when the classes were highly separated, and was still able to identify the correct model in 83% of the time even with poor class separation.

Under the Bayesian estimation, we focused on the DIC3, WAIC2, and BIC because they are the top performers. As expected, the average accuracy of correct model identification increased when the classes were more separated. Specifically, with well-separated class, the DIC3 (96%), BIC (96%), and the WAIC2 (92%) were very accurate in identifying the correct model, and their respective percentage decreased to 89% for DIC3, 82% for BIC, and 87% for WAIC2 when the class separation was poor.

When comparing across the two frameworks (ML/EM vs. Bayesian), we found that the accuracy of the AIC and BIC were particularly higher under the Bayesian estimation than their counterparts in the ML/EM method when the class separation was low. This is because within the Bayesian framework, the model predictive accuracy of the AIC and BIC are summarized by posterior distributions using the Laplace method, which increased their accuracy than when summarized using the maximum likelihood estimator.

Table 3

ANOVA Results of the Designed Factors' Effects on Model Fit Indices in Selecting the True Model across Model Types and Conditions under ML/EM Estimation

Factors	AIC	AICC	**CAIC**	BIC	SABIC	DBIC	NEC	**ICL-BIC**	CLC	BLRT
Class Separation	**0.40***	**0.20**	**0.15**	**0.25**	**0.26**	**0.23**	**0.17**	**0.73***	0.00	0.01
Time Points	**0.21**	**0.31***	**0.17**	**0.18**	**0.38***	**0.30**	0.01	0.00	**0.33***	**0.50***
Sample Size	**0.17**	**0.20**	0.08	0.07	0.07	**0.08**	**0.11**	0.00	**0.13**	0.05
Class Proportion	0.02	0.00	0.04	0.03	0.004	0.02	**0.16**	**0.12**	0.02	**0.24**

Table 4

ANOVA Results of the Design Factors' Effects on Model Fit Indices in Selecting the True Model across Model Types and Conditions under Bayesian Estimation

Factors	AIC	**WAIC2**	**BIC**	**DIC3**	LPML	PsBF
Class Separation	**0.16**	**0.10**	**0.20***	**0.12**	**0.17**	**0.25***
Time Points	**0.14**	0.08	**0.18**	0.06	**0.27***	**0.34***
Sample Size	**0.15**	**0.18**	**0.20**	**0.10**	0.08	0.00
Class Proportion	0.00	0.05	0.00	0.02	**0.10**	0.04
Prior	**0.18***	**0.13**	0.08	0.10	**0.30***	**0.35***

Table 5

Percentage of Correct Model Identification Collapsed by the Factors of Class Separation, Number of Time Points, and Sample Size under ML/EM Estimation

Design Conditions			Performance of Class Enumeration Indices									
Separation	Time Points	N	AIC	AICC	**CAIC**	BIC	SABIC	DBIC	NEC	**ICL-BIC**	CLC	BLRT
High	6	300	86%	**98%**	**100%**	**96%**	**97%**	**100%**	62%	**100%**	74%	**100%**
		1000	73%	84%	**100%**	**98%**	**97%**	**100%**	26%	**100%**	47%	**100%**
	10	300	83%	**91%**	**100%**	**97%**	**91%**	**98%**	44%	**100%**	58%	44%
		1000	54%	65%	**98%**	**95%**	78%	88%	54%	**99%**	40%	26%
Low	6	300	73%	**92%**	**100%**	**91%**	**93%**	**98%**	23%	**93%**	74%	**100%**
		1000	45%	75%	**100%**	**93%**	**88%**	**94%**	10%	**95%**	53%	**100%**
	10	300	29%	57%	**90%**	76%	58%	70%	48%	**93%**	51%	29%
		1000	8%	25%	42%	26%	26%	29%	4%	**94%**	10%	3%

Table 6

Percentage of Correct Model Identification Collapsed by Prior Specification, Class Separation, and Number of Time Points, under Bayesian Estimation

Design Conditions			Performance of Class Enumeration Indices					
Prior	Separation	Time Points	AIC	**WAIC2**	**BIC**	**DIC3**	LPML	PsBF
Data Driven	High	6	82%	**94%**	**97%**	**97%**	**91%**	86%
		10	77%	88%	**96%**	**95%**	78%	70%
	Low	6	75%	**93%**	**92%**	**94%**	84%	80%
		10	58%	80%	72%	84%	58%	50%
Informative	High	6	87%	**100%**	**98%**	**100%**	**91%**	88%
		10	80%	**96%**	**96%**	**97%**	82%	79%
	Low	6	81%	**95%**	**94%**	**96%**	86%	84%
		10	62%	84%	75%	85%	61%	57%
Weakly Informative	High	6	80%	**92%**	**95%**	**96%**	85%	79%
		10	70%	83%	**96%**	**92%**	70%	62%
	Low	6	72%	**91%**	**90%**	**93%**	79%	68%
		10	52%	79%	68%	80%	45%	40%

4.3.3.2 Time Points

The number of time points also had a critical impact on the percentage of correct model identification (η^2 ranged from 0 to 0.5), and the impact was more dramatic on the performance of the AICC, SABIC, and BLRT than on other indices. Specifically, when the number of time points increased from 6 to 10, the correct model identification rate for the AICC dropped from about 90% to 60%, BLRT from 100% to 26%, and SABIC from 94% to 63%. The accuracy for other indices also decreased. However, the accuracy of the ICL-BIC remained overwhelmingly high regardless of the change in the number of time points or class separation. The CAIC were still able to identify the correct model in 83% of the time when the number of time points changed to be 10, as compared to 100% when the number of time points was 6.

The impact of the number of time points under the Bayesian estimation was not as dramatic as that under the ML/EM method. The BIC showed a decrease from 94% to 84% when the number of time points increased from 6 to 10. The WAIC2 and DIC3 were able to maintain good performance with a slight drop in correct model dentification rate with the variation of the number of time points.

4.3.3.3 Sample Size

Under both estimation methods, the variation of sample size did not exhibit significant impact on the accuracy for the indices under either ML/EM or Bayesian framework (η^2 ranged from 0 to 0.2). There is a slight trend that as sample size increases, many indices tend to point to three-class models more often than when the sample size was small. However, regardless of the sample size, the ICL-BIC under the ML/EM

method had the best performance as it consistently provided accurate assessment of the number of trajectories with 93% of the replications when N=300, and 95% when N=1000. Under the Bayesian estimation framework, the WAIC2, BIC, and DIC3 gave reasonably good percentage in correct model identification with a slight decrease from around 95 % to 90% when sample size increases from 300 to 1000.

4.3.3.4 Class Proportion

The factorial ANOVAs results indicated no appreciable differences between balanced and unbalanced proportions under the ML/EM method (η^2 ranged from 0 to 0.24) and under the Bayesian estimation (η^2 ranged from 0 to 0.1). Again, with the ML/EM method, the ICL-BIC stood out to be the most accurate measure to identify the correct class solution across the two levels of this factor. The WAIC2, BIC, and DIC3 under the Bayesian method were reasonably good measures in identifying the correct model.

4.3.3.5 Prior Specification

For the Bayesian analysis, the different prior specifications had a substantial impact on class enumeration (η^2 ranged from 0.08 to 0.35). Overall, the more informative the priors were, the more accurate the indices were in enumerating the correct number of classes. Specifically, when using informative prior, the WAIC2 (94%), BIC (91%), and DIC3 (95%) were very accurate in identifying the correct model. However, the accuracy was compromised with the use of less informative priors. The WAIC2, BIC, and DIC3 showed decrease in accuracy using data driven priors, and the decrease became more dramatic when it came to the use of the weakly informative

priors. With weakly informative prior, the measures of DIC3 (90%), WAIC2 (86%), and BIC (87%) performed with only acceptable accuracy.

4.4 Discussion

This study evaluated and compared the class enumeration accuracy of a variety of model selection indices in Piecewise Growth Mixture models (PGMM) with unknown turning points estimated by the Bayesian and the Maximum Likelihood methods.

4.4.1 ML-Based Class Enumeration Indices

4.4.1.1 Information-based Criteria (IC) Statistics

Under the ML/EM estimation, for the information-based criteria (IC) statistics, the CAIC showed the highest correct model recovery percentages. It performed consistently the best with exception to the condition of low class separation in combination with 10 time points and N=1000. The BIC, SABIC, and DBIC performed similar to each other with exception to the conditions that combines 10 time points and N=1000, in which the BIC (95%) had substantially higher correct model identification percentage than the latter two indices. The AICC performed consistently better than the AIC, and both showed a strong tendency to over extract classes given a large sample size (N=1000 in our study) and even more so when the class separation in the data condition is poor plus with a large sample size and a large number of time points. Overall, the results in our study recommended to use the CAIC and BIC class enumeration with PGMMs.

Considering our results in conjunction with previous studies in GMMs, where Tofighi and Enders (2008), Nylund et al (2007), and Peugh and Fan (2012) found that on

average the SABIC functioned well across different data scenarios in their studies, particularly so when the sample size was large and the class proportion was equal (N=3000 in Peugh and Fan's study). Liu and Hancock (2014) recommended using the BIC and DBIC in the linear GMM context and DBIC across linear GMMs and unrestricted GMMs in various conditions combined.

There are at least three factors that can explain the different findings in our study. First, in our study, the within-class model specification was non-linear; the non-linear component within-class model specification has never been examined in the previous studies. A nonlinear within-class model involves more parameters, resulting in different performances in the enumeration indices from those with linear GMMs, as more severe penalty weight occurred with an increased number of parameters. Second, the largest sample size considered in our study was limited to N=1000; when combined with 10 time points, such sample size was not large enough to exhibit the advantages of sample size adjusted versions of IC statistics (i.e., SABIC) as found in the previous GMM studies. Third, even the low class separation considered in our study could be considered very high compared with the design, for example in Peugh and Fan (2013) study. The MD (multivariate distance) in our study is almost three times the largest distance in their study, which explains partly why in our study the indices performed well even with low class separation and a small sample size. The last but not the least important factor is the factor of the number of time points, with the increase of the number of time points from 6 to 10, the distance between the initial time point and the turning point also increased. As a result, more possibility of different time point estimates could occur given the

increased distance, which might explain why the factor of the number of time points has a substantial impact on the class enumeration performance.

4.4.1.2 Entropy-based Indices under the ML/EM Estimation

At this point, few studies have thoroughly examined the performance of entropy-based statistics in class enumeration with GMMs. The only study that thoroughly evaluated the entropy-based statistics was by Peugh and Pan (2012). However, the formula for calculating the entropy in their study was different from our study as they used the rescaled entropy, a default output in Mplus Version 7.2 for calculation. Our results showed that, the ICL-BIC consistently performed the best in extracting the correct number of growth trajectories, with the accuracy rate ranging from 93% up to 100% in many design cells. The NEC and the CLC, however, did not perform as well as the ICL-BIC and showed strong tendency to extract more complex model with the increase in sample size particularly when the class separation changed from high to low.

4.4.1.3 Likelihood Ratio Test under the ML/EM Estimation

The findings regarding the performance of BLRT were consistent with the results in previous studies (Nylund et al., 2007; Peugh & Fan, 2012; Liu & Hancock, 2014). The BLRT could be used with confidence to identify heterogeneity in growth trajectories, but at the same time we have to be cautious over its tendency to extract false latent classes. In our results, the BLRT were very likely to choose the 3-class over-extracted model in data conditions when the class separation is not sufficiently high plus a large number of time points.

4.4.2 Indices under the Bayesian Estimation

The WAIC2 and DIC3 were newly proposed criteria under the Bayesian framework for model selection. To the best of our knowledge, their performance in class enumeration with GMM has never been evaluated. Our results indicated that the WAIC2, DIC3, and BIC were very useful criteria in enumerating the correct number of latent growth trajectories in PGMM. The performance of the BIC was consistent with what was found in previous studies on mixture models (Li, et al., 2009; Steele & Raftery, 2009). With the knowledge of accurate prior information on the model parameters, the three indices could be used with strong confidence in identifying the heterogeneity in growth trajectories.

On the other hand, the AIC, LPML and PsBF should be used with caution as their tendency to extract spurious latent classes is detrimental when the number of time points is large but without a sufficiently large sample size, and even more damaging when there is very limited information over the prior distributions.

4.4.3 Convergence Issues

No consensus could be found on how to define and handle non-convergence issues in studies of Growth Mixture Models under the maximum likelihood estimation method, particularly with respect to how to address negative variances. For the over extracted models which were prone to having negative variances, Tofighi and Enders (2008) and Nylund et al. (2007) suggested to use highly restricted model in terms of class-varying parameters to avoid negative variances. However, such strategy did not work well in the study by Liu and Hancock (2014), and their study showed a large

discrepancy in convergence rate as compared to the GMM study by Nylund et al. (2007) due to the exclusion of replications that produced Heywood cases. Another approach was considering the problem harmless when the negative variance estimates were statistically insignificant from zero which were caused by random sampling fluctuation (Dillion, Kummar & Mulani, 1987; Gerbing & Anderson, 1987; Tolvanen, 2007).

In our study, we employed the second approach because we would like to keep our model less constrained and allow for class-varying parameters. In addition, an in-depth examination of the replications that bore negative variances revealed the estimates were not statistically significant from zero, a supporting massage that we could include those replications as proper solutions for our final result analysis. However, it should be noted that different data situations require a different approach, particularly when the negative variances were found to be statistically different from 0.

4.4.4 Impact of the Design Factors

Class separation had a significant impact on class enumeration accuracy under both ML/EM and Bayesian methods. The variation of the number of time points had a substantial effect on the accuracy in correct model identification, but such effect was restricted only to the ML/EM method. The accuracy of the class enumeration indices under the Bayesian method was highly sensitive to the prior distributions of the growth factor means. A higher level of accuracy in class enumeration could be obtained given a higher level of informativeness on model parameters, which, however, is not realistic in real data analysis.

4.4.5 Recommendations

For empirical researchers who are interested in using a PGMM with unknown turning points for comparing growth rates for two or more periods, or for estimating the turning point and the associated heterogeneous growth processes, our study results show that the ML/EM or Bayesian estimator perform variably with regard to the accuracy in class enumeration depending on data scenarios. Overall, for the ML/EM method, the ICL-BIC is the most effective measure in correct model identification, followed by the CAIC; for the Bayesian method, the WAIC2 and DIC3 are the most accurate indices, and the BIC is acceptable to a certain extent. For ideal data scenarios where underlying heterogeneous populations are well separated, there is no substantial difference in using either the ML/EM or the Bayesian estimator with regard to their respective best performed enumeration indices. However, for data situations that involve poorly separated latent classes, under the ML/EM estimator, only the ICL-BIC is able to distinguish the latent trajectories regardless of sample size and measurement occasions; for the Bayesian method, the accuracy of the WAIC2, DIC3, and BIC depends on the degree of prior information (or knowledge) that researchers could possibly have to incorporate into the Bayesian model estimation. The more informative the priors are, the more accurate the class enumeration indices are in determining the correct number of growth trajectories.

CHAPTER V

IMPLICATIONS AND CONCLUSIONS

Piecewise growth mixture model (PGMM) is a very flexible technique in analyzing longitudinal data in educational and behavioral research. The approach is particularly useful for the identification of heterogeneous growth trajectories with distinctive turning points that connects two or more different growth processes. When specifying a PGMM, the traditional approach relies on theoretical or design considerations to determine turning points a priori. However, such approach may not be always reasonable. For example, researchers often set the turning point at the time when intervention is given, however, the true turning point may occur after the intervention due to a delay in response to the intervention. A less restrictive alternative is to use the piecewise growth mixture model with unknown turning points (Kohli et al., 2013), which allows researchers to estimate unknown turning points based on data. This approach often yields a more optimal functional form describing the patterns of the observed data over time.

The present study investigated the performance of PGMMs with unknown turning points in three aspects: 1) the accuracy of commonly used enumeration indices in class enumeration, 2) accuracy of parameter estimates, and 3) individual classification accuracy. Two dominant estimation methods were compared, namely, the Bayesian estimation method via the Markov Chain Monte Carlo (MCMC) algorithm and the maximum likelihood estimation via the expectation maximization (EM) algorithm (ML/EM) in the Structural Equation Modeling (SEM) framework.

The present study had several significant findings. First, for the ML/EM method, among the three major categories of class enumeration indices, the ICL-BIC, an entropy-based criterion, has been found to be the most useful index in correct model identification across all the data conditions under examination. Among the information-based criteria (IC) statistics, the CAIC could generally identify the correct model in a higher percentage than the BIC, which was in turn higher than the other IC statistics, but the performance of these two indices was hampered when the data consist of poorly-separated classes in combination with a large number of time points. As for the BLRT, a likelihood ratio test derivative, it is safe to use the index to identify heterogeneity in growth trajectories, but due caution should be paid to its tendency to extract false latent classes. For the Bayesian method, the WAIC2, DIC3, and BIC were found to be very useful criteria in extracting the correct number of latent classes. If the researchers happen to have accurate prior knowledge on the model parameters, those three indices could be used with strong confidence in class enumeration with PGMMs. However, conservativeness should be taken when using the AIC, LPML, and PsBF, as their tendency to over-extract false latent classes could be very serious when the number of time points is large and valid information over the prior distributions is unavailable.

Second, regarding the accuracy in individual classification (i.e., the hit rate), the Bayesian method with informative priors gave generally the highest percentage of correct individual classification than the same method with data driven priors, which was slightly better than the ML/EM method. The Bayesian method with weakly informative priors was comparatively the least accurate in individual classifications. However, such

accuracy varied substantially depending on the class separation and the number of time points across the two estimation methods.

Third, only the Bayesian method with informative priors can recover the true parameters in both classes to a reasonable degree. The Bayesian method with data driven priors and the ML/EM method were found to have difficulty in accurately retrieving the true parameter of the second slope of the second class. Comparatively speaking, the estimates for the parameters of class two is the least reliable when using the Bayesian approach with weakly informative priors. Difficulty was also found for both the Bayesian and ML/EM methods in estimating random effects. Both approaches tend to overestimate the variances of the random effects, with exception to the intercept, which was largely underestimated.

The findings of the present study enhanced the understanding of the performance of a PGMM with unknown turning points estimated in both the ML/EM and Bayesian frameworks. Based on the results of the study, the following suggestions are worthy of consideration given that the researchers have similar data structure to our simulations. First, though the BIC has been generally recommended for class enumeration under both estimation methods, our study found that for the ML/EM method, the ICL-BIC is the most effective measure in identifying the correct number of latent classes; for the Bayesian method, the DIC3 could be used with confidence in correct model identification, followed by the WAIC2. The BIC should be used with caution as it shows substantial variability depending on the number of time points and class separation. Second, for the Bayesian estimation, the more accurate information the researchers have

for the prior distributions, the more reliable the results are for the class enumeration, estimation of the parameters, and individual classification.

The findings of the study should be considered in light of the limitations. First, the present study assumed a two-piece linear by linear growth mixture model with one unknown turning point. However, a PGMM specification could be more complex with each piece taking up different functional forms connected by more than one turning point. Second, the present study used fixed measurement occasions for all individuals, and constrained turning point to be fixed within classes. Such restrictions are necessary under the SEM framework, but can be relaxed under the Bayesian estimation method. More complex data scenarios such as individually-varying measurement occasions, and randomly varying turning points within classes can be accommodated under the Bayesian framework. In the future, researchers could further investigate the performance of PGMMs with random turning point and varying and/or missing measurement occasions under the Bayesian framework.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6), 716-723.

Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research, 40*(2), 235-243.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*(1), 3.

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence, IEEE Transactions On, 22*(7), 719-725.

Biernacki, C., & Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics*, 451-457.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Bozdogan, H. (1987). Model selection and akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*(24), 4279-4292.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis, 1*(4), 651-673.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195-212.

Chen, J. (1995). Optimal rate of convergence in finite mixture models. Annals of Statistics, 23, 221-233.

Chen, Q., Kwok, O., Luo, W., & Willson, V. L. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A monte carlo study. *Structural Equation Modeling, 17*(4), 570-589.

Chou, C., Yang, D., Pentz, M. A., & Hser, Y. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis, 46*(2), 213-225.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101*(1), 126.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus bayesian estimation. *Psychological Methods, 18*(2), 186.

Dominicus, A., Ripatti, S., Pedersen, N., & Palmgren, J. (2006). Modelling variability in longitudinal data using random change point models. *Research Reports in Mathematical Statistics.*

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological),* 45-97.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*(1), 75-95.

Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association, 90*(430), 577-588.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Galatzer-Levy, I. R., Bonanno, G. A., & Mancini, A. D. (2010). From marianthal to latent growth mixture modeling: A return to the exploration of individual differences in response to unemployment. *Journal of Neuroscience, Psychology, and Economics, 3*(2), 116.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. New York, NY: Chapman & Hall/CRC.

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika, 52*(1), 99-111.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association, 74*(365), 153-160.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* CRC press.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds) 169-193. Oxford Univ. Press.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing markov chain monte carlo. *Markov chain Monte Carlo in practice*, *1*, 19.

Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach* (Vol. 20). CRC press.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society.Series B (Methodological)*, 190-195.

Hardy, S. A., & Thiels, C. (2009). Using latent growth curve modeling in clinical treatment research: An example comparing guided self-change and cognitive behavioral therapy treatments for bulimia nervosa. *International Journal of Clinical and Health Psychology*, *9*(1), 51-71.

Harring, J. R., Cudeck, R., & du Toit, S. H. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research, 41*(4), 579-596.

Hathaway, R. J. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics & Probability Letters*, *4*(2), 53-56.

Heidelberger, P., & Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, *24*(4), 233-245.

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*(6), 1109-1144.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202-226.

Heybroek, L. (2011, October). Life satisfaction and retirement: A latent growth mixture modelling approach. In *HILDA Survey 10th Anniversary Research Conference*.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Irony, T. Z., & Singpurwalla, N. D. (1997). Non-informative priors do not exist A dialogue with josé M. bernardo. *Journal of Statistical Planning and Inference, 65*(1), 159-177.

Johnson, V. E. (2004). A bayesian chi2 test for goodness-of-fit. *Annals of Statistics,* 2361-2384.

Johnsson, K. O., Leifman, A., & Berglund, M. (2008). College students' drinking patterns: Trajectories of AUDIT scores during the first four years at university. *European Addiction Research, 14*(1), 11-18.

Kass, R. E., & Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association, 90*(431), 928-934.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A,* 49-66.

Khoo, S., West, S. G., Wu, W., & Kwok, O. (2006). Longitudinal methods. *Handbook of Multimethod Measurement in Psychology,* 301-317.

Kim, S., & Kim, J. (2012). Investigating stage-sequential growth mixture models with multiphase longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 293-319.

Kohli, N., Harring, J. R., & Hancock, G. R. (2013). Piecewise linear–linear latent growth mixture models with unknown knots. *Educational and Psychological Measurement, 73*(6), 935-955.

Koop, G. (2003). *Bayesian econometrics*. John Wiley & Sons.

Kwok, O., Luo, W., & West, S. G. (2010). Using modification indexes to detect turning points in longitudinal data: A monte carlo study. *Structural Equation Modeling, 17*(2), 216-240.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine, 24,* 2401–2428.

Lewis, S. M., & Raftery, A. E. (1997). Estimating bayes factors via posterior simulation with the Laplace—Metropolis estimator. *Journal of the American Statistical Association, 92*(438), 648-655.

Li, F., Duncan, T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling, 8*(2), 175-204.

Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications. Paper presented at the *NSF-CBMS Regional Conference Series in Probability and Statistics*, i-163.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, , 0013164413519798.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767-778.

Lu, X., & Huang, Y. (2014). Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. *Statistics in medicine, 33*(16), 2830-2849.

McAuley, E., Mailey, E. L., Mullen, S. P., Szabo, A. N., Wójcicki, T. R., White, S. M., . . . Kramer, A. F. (2011). Growth trajectories of exercise self-efficacy in older adults: Influence of measures and initial status. *Health Psychology, 30*(1), 75.

McGrory, C. A., & Titterington, D. (2007). Variational approximations in bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis, 51*(11), 5352-5367.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. *Applied Statistics*, 318-324.

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Mengersen, K. L., & Robert, C. P. (1993). *Testing for mixtures: A bayesian entropic approach.* INSEE/Dpt de la recherche.

Mora, P. A., Bennett, I. M., Elo, I. T., Mathew, L., Coyne, J. C., & Culhane, J. F. (2009). Distinct trajectories of perinatal depressive symptomatology: Evidence from growth mixture modeling. *American Journal of Epidemiology, 169*(1), 24-32.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change*. 291-322.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide.* Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. *Advances in Latent Variable Mixture Models, 1*, 1-24.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891.

Natarajan, R., & McCulloch, C. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics, 7*, 267-277.

Neelon, B., Swamy, G. K., Burgette, L. F., & Miranda, M. L. (2011). A bayesian growth mixture model to examine maternal hypertension and birth outcomes. *Statistics in Medicine, 30*(22), 2721-2735.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569.

Park, B., & Lord, D. (2009). Application of finite mixture models for vehicle crash data analysis. *Accident Analysis & Prevention, 41*(4), 683-691.

Park, B., Zhang, Y., & Lord, D. (2010). Bayesian mixture modeling approach to account for heterogeneity in speed data. *Transportation Research Part B: Methodological, 44*(5), 662-673.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 204-226.

Peugh, J., & Fan, X. (2015). Enumeration Index Performance in Generalized Growth Mixture Models: A Monte Carlo Test of Muthén's (2003) Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 115-131.

Plummer M, Stukalov A, & Denwood, M (2015). Package 'rjags'. Update 2015; 16:1.

Preacher, K. J. (2008). *Latent growth curve modeling* (No.157). Sage.

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raftery, A.E. and Lewis, S.M. (1992). "How Many Iterations in the Gibbs Sampler?" In *Bayesian Statistics*, 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.). Oxford, U.K.: Oxford University Press, p. 763–773.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-164.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103-124.

Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review, 26*(2), 195-239.

Richardson, S., & Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59*(4), 731-792.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*(5), 465-471.

Rutter, M. (1996). Autism research: Prospects and priorities. *Journal of Autism and Developmental Disorders, 26*(2), 257-275.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford university press.

Soromenho, G. (1994). Comparing approaches for testing the number of components in a finite mixture model. *Computational Statistics, 9*(1), 65-78.

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

Steele, R. J., & Raftery, A. E. (2009). Performance of bayesian model selection criteria for gaussian mixture models. *Dept.Stat., Univ.Washington, Washington, DC, Tech.Rep, 559*

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of Statistics*, , 40-74.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. *Advances in latent variable mixture models*, (Information Age Publishing, Inc), 317-341.

Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study* (Unpublished Doctoral dissertation). University of Jyväskylä, Finland.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., . . . Rietschel, M. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40*(08), 1367-1377.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wang, L., & McArdle, J. J. (2008). A simulation study comparison of bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling, 15*(1), 52-74.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research, 11*, 3571-3594.

Weiss, R. E. (2005). *Modeling longitudinal data: With 72 figures.* Springer.

Wood, P. K., & Jackson, K. M. (2013). Escaping the snare of chronological growth and launching a free curve alternative: General deviance as latent growth model. *Development and Psychopathology, 25*(03), 739-754.

Wu, A. D., Zumbo, B. D., & Siegel, L. S. (2011). General piecewise growth mixture model: Word recognition development for different learners in different phases. *Journal of Modern Applied Statistical Methods, 10*(1), 21.

Zhao, L., & Banerjee, M. (2012). Bayesian piecewise mixture model for racial disparity in prostate cancer progression. *Computational Statistics & Data Analysis, 56*(2), 362-369.

APPENDIX A

R PROGRAM FOR GENERATING PIECEWISE GROWTH MIXTURE DATA

```
library (mnormt)
library(reshape2)

setwd('C:\\mixturedata\\High Separation\\SixP')
source('writeDatafileR.txt')

sim.lme.slopeCP<- function(G=2, nn=c(50,50),
t0=c(seq(1,6,by=1)),mu.alpha=c(25,22.5), sigma.alpha=c(0.1,0.1),
COV=COV1,mu.beta1g=c(-3.78, -4), sigma.beta1g=c(0.05, 0.05),mu.beta2g=c(-0.24, -
0.18), sigma.beta2g=c(0.05, 0.05),
mu.taug=c(4.3,2.4), sigma.taug=c(0,0), sigma=1, p.zg=c(1,1)){
dir.create(path = paste0( per1,".", per2,".", nsubj))
for (irep in 1:reps) {
        if(length(mu.beta1g) != G)
                stop("The length of 'mu.beta1g' has to be equal to the number of groups")
        if(length(sigma.beta1g) != G)
                stop("The length of 'sigma.beta1g' has to be equal to the number of
        groups")
        if(length(sigma.beta2g) != G)
                stop("The length of 'sigma.beta2g' has to be equal to the number of
        groups")
        if(length(mu.taug) != G)
                stop("The length of 'mu.taug' has to be equal to the number of groups")
        if(length(sigma.taug) != G)
                stop("The length of 'sigma.taug' has to be equal to the number of
        groups"y <- NULL; t=NULL;  id=NULL; g=NULL; ID=NULL
        cumsubj<- c(0, cumsum(nn))
 n1 <-rep(0,G)
nt=rep(length(t0),2)
for(j in 1:G){
   n <- nn[j]
        tau<- rnorm(n, mu.taug[j], sigma.taug[j])
ab<- rmnorm(n = n, mean = c(mu.alpha[j],mu.beta1g[j],mu.beta2g[j]), COV)
alpha<- ab[,1]
beta1 <- ab[,2]
 beta2 <- ab[,3]
n1[j] <-sum(rbinom(n,1,p.zg[j]))
                if(n1[j] > 0){
        for (i in 1:n1[j]){
        t1 <- t0 * (t0 <= tau[i])+ tau[i] * (t0 > tau[i])
```

```
          t2 <- (t0-tau[i]) * (t0 > tau[i])
          X <- cbind(rep(1, nt[j]), t1, t2)
          y.part<- X%*%c(alpha[i], beta1[i], beta2[i]) +
          rnorm(nt[j], 0, sigma)
          y <- c(y, y.part)
                        }
        }

       t <- c(t, rep(t0, n))
id<- c(id, rep(1:n, each=nt[j]))
g <- c(g, rep(j, n*nt[j]))
ID <- c(ID, rep(1:n, each=nt[j])+cumsubj[j])
S=max(ID)
N=length(y)
}
    output.dat <- data.frame(y=y, t=t, ID=ID)
    output.dat1 <- list(y=y, t=t, ID=ID, S=S, N=N)
output.dat.long<- melt(output.dat, id.vars = c("ID", "t"))
output.dat.wide<- dcast(output.dat.long, ID ~ t)
writeDatafileR(output.dat1, paste0(per1,".", per2,".", nsubj, "/databugs", irep, ".txt"))
write.table(output.dat.wide, paste0(per1,".", per2,".", nsubj, "/data", irep, ".txt"),
          row.names = F, col.names = F)
    "/datalong", irep, ".txt"), row.names = F, col.names = F)
    }
  }

COV1<- matrix(c(0.1,0.025,0,0.025,0.05,0.0175,0,0.0175,0.05),ncol=3) # high
          separation

nsubj<- 300    ### sample size: 300/1000
RepMe<- 6         ### number of repeated measures: 6/10
per1 <- .5   ### mixing percentages: .5 & .5 or .75 & .25
per2<- .5
reps<- 200      ### replications
dd<- sim.lme.slopeCP(G=2, nn=c(nsubj*per1,nsubj*per2), t0=c(seq(1,RepMe,by=1)),
mu.alpha=c(25,22.5), sigma.alpha=c(0.1,0.1),COV=COV1,mu.beta1g=c(-3.78,
 -4), sigma.beta1g=c(0.05, 0.05),mu.beta2g=c(-0.24, -0.18), sigma.beta2g=c(0.05,
          0.05),mu.taug=c(4.3,2.4), sigma.taug=c(0,0),sigma=1, p.zg=c(1,1))
```

APPENDIX B

USE THE MPLUSAUTOMATION PACKAGE TO GENERATE MPLUS INPUT

FILES FOR FITTING A PIECEWISE GROWTH MIXTURE MODEL

```
[[init]]
iterators =  nsubj per1 per2 reps ;
nsubj = 300 1000 ;
per1 = 0.5;
per2 = 0.5;
reps = 1:200;
filename = "data[[reps]]_C2.inp";
outputDirectory = C:/mixturedata/Mplus/6.1/output/C2/[[per1]].[[per2]].[[nsubj]]";
[[/init]]

Title: C2-6.[[nsubj]].[[per1]].[[per2]]-data[[reps]];
  Data:
file is "C:/mixturedata/Mplus/6.1/[[per1]].[[per2]].[[nsubj]]/dataw[[reps]].txt";

VARIABLE:
  NAMES are ID t1-t6;
  USEVARIABLES are t1-t6;
  CLASSES = c(2);

  ANALYSIS:
  TYPE IS MIXTURE;
  STARTS = 75 25;
  K-1STARTS = 50 10;
  LRTSTARTS = 0 0 100 20;
  STITERATIONS = 50;
  ITERATIONS = 5000;
  SDITERATIONS = 250;
  MITERATIONS = 1000;

  MODEL:
  %OVERALL%
w1 BY t1-t6@1;
w2 BY t1@0 t2@1 t3@2 t4@3 t5@4 t6@5;
w3 BY t1*0(p1);
w3 BY t2-t6 (p2-p6);
  w1*(v1);
  w2*(v2);
  w3*(v3);
```

```
w1 WITH w2*;
w1 WITH w3*;
w2 WITH w3*;
  [t1-t6@0];
t1-t6;

  %c#1%
w3 BY t1* (p1);
w3 BY t2-t6 (p2-p6);
  [w1*17.9](mw11);
  [w2*-2.45](mw21);
  [w3*1.785](mw31);

  %c#2%
w3 BY t1 (q1);
w3 BY t2-t6 (q2-q6);
  [w1*17.86](mw12);
  [w2*-2.535](mw22);
  [w3*1.965](mw32);

  MODEL CONSTRAINT:
   mw11 >mw12;
NEW(gam1*3.6 gam2*1.75 b11 b21 b41 b12 b22 b42);
  p1 = (sqrt((0-gam1)^2));
  p2 = (sqrt((1-gam1)^2));
  p3 = (sqrt((2-gam1)^2));
  p4 = (sqrt((3-gam1)^2));
  p5 = (sqrt((4-gam1)^2));
  p6 = (sqrt((5-gam1)^2));
  q1 = (sqrt((0-gam2)^2));
  q2 = (sqrt((1-gam2)^2));
  q3 = (sqrt((2-gam2)^2));
  q4 = (sqrt((3-gam2)^2));
  q5 = (sqrt((4-gam2)^2));
  q6 = (sqrt((5-gam2)^2));
  b11=mw11 + mw31*gam1;
  b21=mw21-mw31;
  b41=mw21 + mw31;
  b12=mw12 + mw32*gam2;
  b22=mw22-mw32;
  b42=mw22 + mw32;

  OUTPUT:
  SAMPSTAT TECH11 TECH14 ;
```

```
SAVEDATA:
FILE =
        "C:/mixturedata/Mplus/6.1/output/C2/[[per1]].[[per2]].[[nsubj]]/cmember[[reps]]
        .dat";
 FORMAT IS F8.2;
SAVE = CPROBABILITIES;
```

APPENDIX C

USE THE MPLUSAUTOMATION PACKAGE TO CALLMPLUS FOR FITTING A PIECEWISE GROWTH MIXTURE MODEL

```
workdir<- switch(Sys.info()['sysname'],
                 'Windows' = "C:/mixturedata/Mplus")

setwd(workdir)

createModels("HS_6p_C2_55.txt")

savedir<- "C:/mixturedata/Mplus/6.1/output/C2/0.5.0.5.300"

runModels(directory = savedir, filefilter = "C2.inp")
```

APPENDIX D

R SYNTAX OF TWO-CLASS BAYESIAN MODEL

```
library(rjags)
library(foreign)
library(coda)

setwd('C:/mixturedata/6.1')
setwd('0.5.0.5.300')
modelstring="
model {
pi<-3.141593
for (i in 1:N) {
y[i] ~ dnorm(mu[i], w)
  mu[i] <- b[ID[i],G[ID[i]], 1]+(b[ID[i],G[ID[i]],2]*t[i])*step(r[ID[i],G[ID[i]]]-t[i])
         +b[ID[i],G[ID[i]],2]*r[ID[i],G[ID[i]]]*step(t[i]-r[ID[i],G[ID[i]]])
         +b[ID[i],G[ID[i]],3]*(t[i]-r[ID[i],G[ID[i]]])*step(t[i]-r[ID[i],G[ID[i]]])
log.like[i]<- -0.5*log(2*pi)+0.5*log(w)-0.5*(y[i]-mu[i])*(y[i]-mu[i])*w
like[i] <- exp(log.like[i])
 }
for (s in 1:S) {
for (k in 1:2) {
b[s ,k , 1:3] ~ dmnorm(b.mu[k,1:3],  tau[1:3,1:3])
r[s,k] ~  dnorm(Omu.r[k], tau.r[k])
 }
 M[s] ~ dbern(prior)
 G[s] <- M[s]+1
 }
#  prior for mixture probability vector
tau [1:3,1:3] ~ dwish(Omega,10)      # the precsion matrix for random effects
B1[1]~dnorm(-4,5)
B1[2]~dnorm(-3.78,5.29)
B2[1]~dnorm(-0.18,111)
B2[2]~dnorm(-0.24,83)
A[1]~dnorm(22.5,0.89)
A[2]~dnorm(25,0.8)
#  }
b.mu[1:2,1] <-sort(A)
b.mu[1:2,2]<-sort(B1)
b.mu[1:2,3]<-B2
Omu.r[1]~dnorm(2.4,8.3)
Omu.r[2]~dnorm(4.3,4.65)
prior ~ dbeta(0.5, 0.5)
```

```
tau.r[1]~dgamma(0.001, 0.001)
tau.r[2]~dgamma(0.001, 0.001)
w~dgamma(0.001, 0.001)
e<-1/w
}
"
writeLines (modelstring, con="C:/mixturedata/C2/info/C2_info.bug")
filenames<-list.files()
parameters =  c("b.mu","Omu.r", "e", "tau")
bugs.output<- matrix(nrow = length(filenames), ncol = 27)
G.output<- matrix(nrow = length(filenames), ncol = 300)
gweke<- matrix(nrow = length(filenames), ncol = 9)
heidel<-matrix(nrow = length(filenames), ncol = 9)
fitstats<- matrix(nrow = length(filenames), ncol = 6)

for (i in 1:length(filenames)){
dat<- read.csv(paste0('C:/mixturedata/6.1/0.5.0.5.300', '/data', i, '.csv'), header=T)
  t=dat$t
  ID=dat$ID
  y=dat$y
  N=NROW(y)
  S=max(ID)
  Omega=structure(.Data = c(
    0.001, 0, 0,
    0, 0.001, 0,
    0, 0, 0.001), .Dim = c(3, 3))
dat<-list(t=t,ID=ID, y=y, N=N, S=S,Omega=Omega)
burnInSteps = 10000
nChains = 1
thinSteps = 10
nPerChain=40000
  Model = jags.model ("C:/mixturedata/C2/info/C2_info.bug", data=dat,
          n.chains=nChains)
cat("Burning in the MCMC chain...\n")
update(Model, n.iter=burnInSteps)
cat("Sampling from the final MCMC chain ... \n")
codaSamples = coda.samples(Model, variable.names=c(parameters, "like", "G"),
n.iter=nPerChain,thin=thinSteps )

subsample<-codaSamples[,c(301:309)]   ### sub sample of omu.r, b.mu, e

like<- matrix(nrow = 4000, ncol = 300)
like_all<- codaSamples[,c(310:2109)]
```

```
for (L in 1:4000){
  like_1<-do.call(rbind,like_all[L,] )
likes<-matrix(like_1, nrow=6,ncol=300)
  R <- apply(likes, 2, prod)
like[L,]<-R
  }
   Deviance <--2*mean(apply(like,1, function(s) sum(log(s))))
dm<-21
  AIC <- Deviance + dm*2
  BIC <- Deviance + dm*log(max(ID))
  DIC3= 2*Deviance + 2*sum(log(colMeans(like)))
  CPO =  1/colMeans(1/like)
  LPML= sum(log(CPO))
lpd=log(colMeans(like))
p_waic=apply(log(like),2,var)
  elpd1_i=lpd-p_waic
  waic1= -2* sum(elpd1_i)
logPsBF=sum(log(CPO))
b.tau<- rbind(matrix((summary(codaSamples)$stat [c(2110:2118), 1]),
          ncol=3,byrow=T))
b.Sigma<-solve(b.tau)
b.sigma<-as.vector(b.Sigma)
subcodaS<- summary(codaSamples)$stat[c(301:309),c(1:2)]
par<-c(subcodaS,b.sigma)
bugs.output[i,]<- par #c(summary(subcodaS)$stat[,1])
G.output[i,] <- c(summary(codaSamples)$stat[c(1:300),1] )
fitstats[i,]<-c(Deviance, AIC, BIC, DIC3, waic1, LPML)
geweke.vec<- geweke.diag(subsample)[[1]]$z
heidel.vec<- heidel.diag(subsample)[[1]][, 6]
gweke[i,]<- geweke.vec
heidel[i,]<-heidel.vec
}
colnames(bugs.output)<("r1","r2","a1","a2","b11","b12","b21","b22","e","sd.r1","sd.r2"
          ,"sd.a1","sd.a2","sd.b11","sd.b12","sd.b21","sd.b22", "sd.e","taua",
          "tauab1","tauab2","taub1a","taub1","taub1b2","taub2a","taub2b1","taub2")
colnames(gweke) <- names(geweke.vec)
colnames(heidel) <- names(heidel.vec)
colnames(fitstats) <- c("Deviance","AIC", "BIC","DIC3","WAIC2", "LPML" )

write.csv(bugs.output,file='C:/mixturedata/C2/info/ParMean_6.1_55300_1.csv')
write.csv(gweke,file='C:/mixturedata/C2/info/Geweke_6.1_55300_1.csv')
write.csv(heidel,file='C:/mixturedata/C2/info/Heidel_6.1_55300_1.csv')
write.csv(G.output,file='C:/mixturedata/C2/info/G_6.1_55300_1.csv')
write.csv(fitstats,file='C:/mixturedata/C2/info/fit_6.1_55300_1.csv')
```