Title
Bayesian Non/Semi-Parametric Methods for Latent Growth Mixture Models

Author
Yang, Yuzhu

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Bayesian Non/Semi-Parametric Methods for Latent Growth Mixture Models

A dissertation for the degree
Doctor of Philosophy

in

Psychological Sciences

by

Yuzhu Yang

2018

Committee members:
Professor Sarah Depaoli, Chair
Professor Jack Vevea
Professor Keke Lai

The dissertation of Yuzhu Yang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

(Professor Jack Vevea)

(Professor Keke Lai)

(Professor Sarah Depaoli, Chair)

University of California, Merced

2018

To my family

Table of Contents

List of Figures

List of Tables

Acknowledgements

Now as I am writing the acknowledgements for my dissertation, I can't help but look back at the five years of my PhD study. It has been an incredible journey. There were ups and downs, tears and joys. But most importantly, it made a profound impact not only on my intellectual and career development but also on the path of my personal life. I would like to take this opportunity and express my deepest gratitude to those who have helped me get through this journey.

First, I would like to thank my academic advisor, Dr. Sarah Depaoli. Sarah has been a caring and supportive advisor from the first moment I started my PhD study. She introduced me into the research area of Quantitative Psychology. She guided me at each step towards my final goal of getting my PhD. She provided me with skill sets that could help me succeed in academia as well as in industry. She inspired me with new ways of thinking and problem solving in research challenges. She supported me when obstacles were presented in my life. I would not have started this invaluable learning experience if she had not admitted me to her lab, nor would I have achieved any academic success without her guidance.

In addition, I would like to thank the members of my dissertation committee, Dr. Jack Vevea and Dr. Keke Lai. They have shown their kind patience and provided enormous help to not only my dissertation work but also my previous milestone projects. I would also like to express my appreciation to my lab mates, my roommate and good friend, and fellow PhD students from other labs. We have supported each other by discussing research ideas, learning new things together, and lending a sympathetic ear. I am grateful to have their company.

Finally, I would like to thank my parents. Although they are half a world away, they always manage to let me know that they love me and support me unconditionally. Thanks to my husband who met me when I was struggling to survive grad school and then has always been by my side since. Thanks to my in-laws for always feeding me with delicious food that reminds me of home.

VITA

EDUCATION

Ph.D. Candidate in Psychological Sciences

University of California, Merced, Merced, CA, USA 2013 – 2018 (Expected)

Master of Science

Baruch College, City University of New York, New York, NY 2010 - 2012

Master of Science

Florida Institute of Technology, Melbourne, FL 2006 - 2008

Bachelor in Educational Psychology

Xuzhou Normal University, Xuzhou, Jiangsu, China 2001 - 2005

WORKING EXPERIENCE

Human Resources Intern

Brooklyn Academy of Music, New York, NY 2009 - 2010

Behavior Specialist

Tucci Learning Solutions, Watsonville, CA 2008 - 2010

Abstract of the Dissertation

Bayesian Non/Semi-Parametric Methods for Latent Growth Mixture Models

by

Yuzhu Yang

Doctor of Philosophy in Quantitative Psychology

University of California, Merced, 2018

Professor Sarah Depaoli, Chair

This dissertation consists of two studies that introduce and investigate two Bayesian non/semi-parametric estimation methods for latent growth mixture modeling (LGMM). LGMM is a useful statistical tool for modeling latent classes or unobserved subgroups in longitudinal data analysis. One of the major challenges of fitting an LGMM is deciding on the number of latent classes that exist in the population from which data were collected. In this dissertation, I introduce two non/semi-parametric estimation methods, that is Reversible jump Markov chain Monte Carlo (RJMCMC) and Dirichlet process modeling (DP) for LGMM. Specifically, I examined the estimation performance of these two non/semi-parametric methods along with traditional estimation methods, such as maximum likelihood (ML) and the Bayesian estimation framework. I also investigated some commonly discussed topics within the LGMM context, such as class enumeration and the impact of class separation. In particular, Study 1 examines the ability of RJMCMC, DP, and ML to recover the model parameters, especially the number of classes and class sizes via a simulation study. Simulation results showed that RJMCMC and DP performed comparable to ML and even better under some conditions for some parameters. An empirical example is included in Study 1 as an illustration of how to apply RJMCMC and DP; the example uses an education-related data set and covers how to interpret the results. In Study 2, the investigation is focused on the impact of class separation on class enumeration and model parameter recovery. Specifically, different degrees of class separation and several separation conditions were investigated. The performance of RJMCMC, DP and two Bayesian estimation methods with different prior specifications were examined for the LGMM via a simulation study. Results of Study 2 showed that RJMCMC and DP performed comparable to the Bayesian estimators under different degrees of class separation. Findings of the two studies suggested that RJMCMC and DP can be used as alternatives to traditional ML and Bayesian estimation methods in accurately recovering the number of latent classes for LGMM under most conditions. However, there are added benefits to the use of RJMCMC and DP over the other approaches. Other implications, suggestions for applied researchers, limitations, and future directions are also discussed.

Chapter 1

Overview of Dissertation

Latent growth mixture modeling (LGMM) has been a useful tool for identifying multiple unobserved subgroups and describing longitudinal change within each subgroup in social and behavior sciences. In this dissertation, I will introduce Bayesian non/semi-parametric methods into the latent growth mixture modeling framework. Specifically, I will discuss the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm and the Dirichlet Process (DP) technique as non/semi-parametric methods on deciding on the number of classes for latent growth mixture models. I will compare the performance of these two Bayesian non/semi-parametric methods with the frequentist method, as well as the Bayesian estimation method. Ultimately, I am interested in the capability of the Bayesian non/semi-parametric methods to detect the number of latent classes, specifically in the context of different class separation conditions (i.e., assessing the performance when classes are more alike versus more disparate).

This dissertation consists of two separate studies. The first study is entitled "Deciding on the Number of Classes for LGMM using Bayesian Non/Semi-Parametric Methods". The primary goal of the first study is to examine the performance of RJMCMC and DP on extracting the correct number of latent classes for LGMM. Another goal of the first study is to compare these two methods with a commonly used implementation (i.e., the frequentist method) of the LGMM. The third goal is to provide an empirical example to illustrate how to apply RJMCMC and DP through a case study using a real world data set.

The second study is entitled "Class Enumeration under Various Levels of Class Separation: Bayesian Non/Semi-Parametric Methods vs. a traditional Bayesian Approach". In the second study, the main goal is to investigate the performance of the RJMCMC algorithm and the DP technique under various degrees of class separation conditions. A full Bayesian estimation method with different prior specifications is also examined as comparisons to the Bayesian non/semi-parametric methods.

This dissertation is structured as follows. First I will discuss the general formulation of LGMM that is examined in the two studies. Then I will introduce the RJMCMC algorithm and the DP technique and their applications on LGMM. Next, two studies will be presented aiming to: 1) examine the performance of the two methods in the context of mixture models (Study 1 of the dissertation), and 2) address specific issues linked to class separation (Study 2 of the dissertation). Finally, I will conclude by discussing the implications of using the two Bayesian non/semi-parametric methods for mixture modeling, as well as provide recommendations for use in applied research settings.

Chapter 2

General Introduction

2.1 Basic Formulation of Latent Growth Mixture Modeling

In this section, I will briefly introduce LGMM and its basic formulation that will be examined using different modeling techniques presented in this dissertation.

LGMM is a statistical tool for capturing multiple latent (unobserved) subpopulations and examining the change within and between subpopulations over time (Muthén, 2001). In other words, LGMM identifies the unobserved subpopulations, describes the longitudinal change within the subpopulations, and tests the difference among the subpopulations.

LGMM incorporates the features of both latent growth curve models (LGCMs) and finite mixture models (FMMs) within one modeling technique. LGCM models the change over time and tests hypotheses about between-individual differences and within-individual change (Muthén, 2001). FMM, on the contrary, focuses on identifying and accounting for the unobserved heterogeneity in the data and assigning individuals into latent groups. I will present a brief introduction to LGCMs and FMMs in the following subsections.

2.1.1 Latent Growth Curve Models

As a useful tool for the longitudinal data analysis, LGCMs keep track of separate trajectories of each individual as well as capture the average growth trajectory for the whole group (Bollen & Curran, 2005). In other words, LGCMs summarize the group growth intercept and slope using parameters such as mean and variances while allowing each individual to have a distinct intercept and slope to simultaneously describe the unique path.

Following Bollen and Curran (2005)'s notation, the basic formulation of a LGCM can be specified as:

$$y_{it} = \alpha_i + \lambda_t \beta_i + \epsilon_{it}, \tag{1}$$

where y_{it} is the variable y for the case i at time t, α_i and β_i are the random intercept and the random slope for case i, and ϵ_{it} represents the random error. The λ_t parameter is a constant, which is commonly coded as $\lambda_1 = 0$, and $\lambda_2 = 1$, and $\lambda_t = t - 1$ for all t in a linear growth model. The random intercept and the random slope can further be modeled with the following equations:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i}, \tag{2}$$

$$\beta_i = \mu_\beta + \zeta_{\beta_i}, \tag{3}$$

where μ_α and μ_β are the mean intercept and mean slope across all cases, and ζ_{α_i} and ζ_{β_i} are disturbances for α_i and β_i, respectively. When we combine the two levels of models into a single model, we get:

$$y_{it} = \left(\mu_\alpha + \lambda_t \mu_\beta\right) + \left(\zeta_{\alpha_i} + \lambda_t \zeta_{\beta_i} + \epsilon_{it}\right). \tag{4}$$

In this single-level model, $\mu_\alpha + \lambda_t \mu_\beta$ is referred to as the fixed coefficients, while $\zeta_{\alpha_i} + \lambda_t \zeta_{\beta_i} + \epsilon_{it}$ is random. It is assumed that the disturbance variables have a mean of zero and variances of $\psi_{\alpha\alpha}$ and $\psi_{\beta\beta,}$ and the covariance between the intercept and slope is denoted as $\psi_{\alpha\beta}$. It is also assumed that the variance of the disturbance for case i at time t is $\theta_{\epsilon_{it}}$.

2.1.2 Finite Mixture Models

FMMs are a data-driven approach (as opposed to a substantive theory-based approach) to the modeling of random phenomena that consist of mixtures of distributions. This type of modeling can be very useful in classifying cases into discrete groups in the social and behavioral sciences.

We can write a basic FMM in the following form using notation partially from McLanchlan and Peel (2000) and Depaoli (2013). First, we assume that data are generated from a finite mixture distribution, $f(y_i; \Psi)$, in this case a normal distribution. This mixture distribution can be represented by the following mixture density function for mixture class c such that

$$f(y_i; \Psi) = \sum_{c=1}^{C} \pi_c f_c(y_i; \Delta_c), \tag{5}$$

where y_i is a vector of repeated measure outcomes for case i with $i = 1,2, \dots, n$, π_c represents the unknown mixture class proportion for the cth latent class with $c = 1,2, \dots, C$, $\sum_{c=1}^{C} \pi_c = 1$, and f_c (for $c = 1,2, \dots, C$) are the densities of the C latent classes that are assumed to be multivariate normal (MVN): $y_i | c \sim MVN(\mu_c, \Sigma_c)$, and μ_c and Σ_c are the mean vector and the covariance matrix of the multivariate normal distribution from which the random samples y_i are drawn for the cth class. Further, $\Psi = (\pi, \Delta)$ represents a vector of unknown parameters, which includes the mixture class proportions $\pi = \pi_1, \pi_2, \dots, \pi_C$ and the unknown parameter vectors $\Delta = \Delta_1, \Delta_2, \dots, \Delta_C$. The parameter Δ_c for each cth latent class contains a vector of model specific parameters, such as means and variances of regression coefficients in a regression model. We then introduce an assignment parameter z_i, which is a vector of associated component-labels for each y_i with $z = z_1, z_2, \dots z_n$. The parameter z_i is defined to be 1 when y_i is in class c or 0 when y_i is not in class c for cth class.

2.1.3 Latent Growth Mixture Models

In the social and behavioral sciences, it is common for research questions to involve classifying cases based on their repeated measures over time when their group membership is unknown. LGMMs provide a tool that combines the LGCM and FMM and makes use of the features from both modeling techniques. This type of model classifies cases into unobserved groups and estimates the latent growth curve of each group simultaneously (Bollen & Curran, 2005)

I use the formulation in Bollen and Curran (2005) (but with different notation from their book) to illustrate the specification of the LGMM below. The 2-level model can be written as

$$y_{it} = \sum_{c=1}^{C} \pi_{ic}[\alpha_{ic} + \lambda_{tc}\beta_{ic} + \epsilon_{itc}], \tag{6}$$

$$\alpha_{ic} = \mu_{\alpha_c} + \zeta_{\alpha_{ic}}, \tag{7}$$

$$\beta_{ic} = \mu_{\beta_c} + \zeta_{\beta_{ic}}, \tag{8}$$

where y_{it} is the measure of variable y for case i at time t, $\pi_{ic}$is the probability that the ith case belongs to the cth group with all $\pi_{ic} \geq 0$ and $\sum_{c=1}^{C} \pi_{ic} = 1$, and the subscript c denotes the latent class each parameter belongs to. λ_{tc} represents the coding of time t for all cases in class c; in the two studies of this dissertation, $\lambda_{tc} = 0,1,2,3$.

Therefore in the expression of the LGMM, y_{it} can be seen as a function of a vector of unknown parameters, which are the growth trajectories with parameters dictating each group, Δ, and the probability that the case belongs to that group, π. In the case of an unconditional linear growth model, Δ includes the means of the random intercept and the random slope ($\mu_{\alpha c}$ and $\mu_{\beta c}$), the variances of α_{ic} and β_{ic} ($\psi_{\alpha\alpha}$and $\psi_{\beta\beta}$), the covariance between α_{ic} and β_{ic} ($\psi_{\alpha\beta}$), and variances of disturbance for case i at time t is $\theta_{\epsilon_{it}}$. Although in many situations, $\psi_{\alpha\alpha}$, $\psi_{\beta\beta}$, $\psi_{\alpha\beta}$, and $\theta_{\epsilon_{it}}$ can vary across latent classes, they are constrained to be equal in the studies in this dissertation for the purposes of model simplification.

This basic LGMM of four time points with an intercept and a linear slope will be examined with the RJMCMC algorithm and the DP technique, as well as other modeling methods in this dissertation.

2.2 Reversible Jump Markov Chain Monte Carlo

In this dissertation, two types of Bayesian non/semi-parametric methods are examined for LGMMs. The first method is the reversible jump Markov chain Monte

Carlo (RJMCMC) process. RJMCMC is a type of random sweep Metropolis-Hastings (MH) algorithm; it extends the MH algorithm to more general state spaces (Richardson & Green, 1997). RJMCMC constructs Markov chains using the reversible jumping rules and enables jumps between the parameter subspaces, whose dimensions can vary across iterates of the Markov chain. For example, jumps are allowed to take place between two adjacent iterates or sweeps, in which the dimension of the mixture component parameter (i.e., latent classes[1]) is different from one iterate to the other in the case of mixture modeling.

One of the advantages of RJMCMC is that it allows jumps between parameter subspaces of differing dimensionality. This feature of RJMCMC makes it a very useful tool for solving statistical problems with inferences that are not fixed, such as the unknown number of models being selected, of mixture components, or of changing times and rates.

The Bayesian non/semi-parametric methods discussed in this dissertation are the modeling techniques that focus on estimating the number of mixture components. This process is different from other types of non/semi-parametric methods, which examine the functional form(s) of the relationship between variables; that method is not a topic addressed in this dissertation.

2.2.1 The Development of RJMCMC

Green (1995) applied RJMCMC as a solution to Bayesian model selection in several types of modeling contexts. In a Bayesian multiple change-point analysis, RJMCMC was used to compute the number of points in the step functions, which was allowed to vary instead of being fixed. In the next example, the researcher extended the usefulness of RJMCMC to an image segmentation process that was essentially a form of a two-dimensional step function. An RJMCMC algorithm was modified from the previous one-dimensional change-point problem and used on the multidimensional step function. In this same paper, Green then further used RJMCMC in a model partition problem for binomial data in Bayesian analysis of factorial experiments. Specifically, the number of the partition of the subgroups was unfixed and estimated based on the “birth and death” algorithm.

In the context of mixture modeling, RJMCMC treats the mixture representations as an unknown and hence varying component and models the number of mixture component and other model parameters in one process. RJMCMC has been intensively applied in estimating the number of mixture components for different types of mixture models, but this process has rarely been examined within the social or behavioral sciences.

Richardson and Green (1997) used RJMCMC to estimate the number of mixture components of a univariate normal mixture model. This mixture model was formulated in a hierarchical form, where the number of mixture components and other model

[1] Latent classes are also called “mixture components” in the literature of mixture modeling (e.g., McLachlan and Peel, 2000) and these two terms are used interchangeably in this dissertation.

parameters were regarded as unknown and were drawn from the prior distributions, respectively. The model can be written in the following form[2]:

$$y_i \sim \sum_{c=1}^{C} \pi_c f(y_i|\Delta_c), \text{indepently for } i = 1,2,\dots,n \text{ and } c = 1,2,\dots,C, \quad (9)$$

where $f(y_i|\Delta)$ represents a given density function parameterized by a generic vector Δ. Here in the univeriate normal mixture model, the unknown parameters include: Δ, which is a vector of a pair of means and variances, μ_c and ${\theta_c}^2$, and the mixture component weight (i.e., latent class proportion), π_c, for a specific class c. The class membership (also regarded as the group label or the latent allocation variable) is denoted as z_i for each $y_{i,}$. The class membership represents the latent class where the observation is drawn from and is unknown. The z_i values are independently drawn from a distribution, where

$$p(z_i = c) = \pi_c, \text{ for } c = 1,2,\dots,C. \quad (10)$$

The observations y_i can then be seen as drawn from their respective individual subpopulations c, given z_i:

$$p(y_i|\, z) \sim f(y_i|\Delta_{z_i}), \quad (11)$$

which, in the univariate normal modeling context, can be expanded as

$$f(y_i|\Delta_c) = f(y_i|\mu_c, \sigma_c^2) = \frac{1}{\sqrt{(2\pi)}\sigma_c} \exp\left\{-\frac{(y_i - \mu_c)^2}{2\sigma_c^2}\right\}. \quad (12)$$

In the Bayesian framework, the unknown parameters $c, z, \pi,$ and Δ are drawn from appropriate prior distributions. Given the above formulas, the joint distribution of this model can be expressed as

$$p(c, z, \pi, \Delta, y) = p(c)p(\pi|c)p(z|\pi, c)p(\Delta|c)p(y|\Delta, z). \quad (13)$$

The prior distributions for $\Delta = \mu_c, \sigma_c^2$ are

$$\mu_c \sim N(mean, variance) \quad (14)$$

and

[2] The original notations from Richardson and Green (1997) are not used here because some of them are conflict with the formulation in the current dissertation. Therefore I reconstructed the formulas from Richardson & Green (1997) in a way such that all the notation follows the formulation in this dissertation.

$$\sigma_c^2 \sim \Gamma(b_1, b_2), \tag{15}$$

where μ_c is drawn from a normal distribution with a hyperparameter mean, denotaed as $mean$, and a hyperparameter variance, denoted as $variance$; σ_c^2 follows a inverse gamma distribution with shape and scale parameters b_1 and b_2, respectively. The proportion weight parameter π is drawn from a Dirichlet distribution

$$\pi \sim Dirichlet(\delta_1, \delta_2, \ldots, \delta_C) \tag{16}$$

And the prior distribution for the number of latent class C is a uniform distribution between 1 and an integer, C_{max} denoted the maximum number of latent classes being specified in the analysis.

Richardson and Green (1997) described the process of RJMCMC for the univariate normal model as follows. Using x to denote the current state of the vector of the unknown parameters, including $\Delta, \pi, and\ c,$ with a posterior probability $p(dx)$. When a type of move m is proposed, it would take the state to x', with a probability $q_m(x, dx')$, which is called the Markov transition kernel (Green, 1995). The move m from state x to x' is accepted with probability

$$a_m(x, x') = \min\left\{1, \frac{p(dx')q_m(x', dx)}{p(dx)q_m(x, dx')}\right\}, \tag{17}$$

where $p(dx')$ is the posterior probability of x', $q_m(x', dx)$ is the probability of moving from state x' to state x. Based on this generic formula, an acceptance probability for dimension-changing moves in the mixture modeling context can be written as

$$a_m(x, x') = \min\left\{1, \frac{p(x'|y)r_m(x')}{p(x|y)r_m(x)q(u)}\left|\frac{\partial x'}{\partial(x, u)}\right|\right\}, \tag{18}$$

where $p(x'|y) = p(dx')$ and $p(x|y)$ represent the posterior probabilities of state x and state x'; r_m is a probability of choosing move type m when in state x, and $q(u)$ is the density function of an auxiliary variable, u, included to ensure the dimension-matching (i.e., to match the degrees of freedom of joint variation of the state and proposal as the dimension changes with c) in the dimension-changing moves; $\left|\frac{\partial x'}{\partial(x,u)}\right|$ is the determinant of the Jacobian matrix (i.e., regarded as "the Jacobian" in the literature) for the change of variable from (x, u) to x' (Richardson & Green,1997).

Richardson and Green (1997) illustrated the performance of RJMCMC for the univariate normal mixture model with several real data sets. They presented the predictive densities both conditional and unconditional on the number of the mixture components c and chose the proper number of components based on the results. This study also discussed the sensitivity of the posterior distribution for *c* to the prior distributions for the means and the variances. They found that the posterior distribution of *c* was insensitive to the hyperparameters of the fixed effects used to specify the prior distribution of the variance.

Ho and Hu (2008) extended the application of RJMCMC from the univariate normal model to the random effects normal mixture model. In this study, they provided an RJMCMC algorithm for a linear random effects model with a random intercept and a random slope. This Gaussian (normal) mixture random-effects model was specified in the form of a Bayesian hierarchical model. Equations 6-8 in the univariate normal mixture model became, for $i = 1, \dots, n$ and $j = 1, \dots, J$,

$$y_{ij} = \alpha_{1j} + \alpha_{2i} u_{ij} + \beta_1 v_{1ij} + \beta_2 v_{2ij} + \cdots + \beta_q v_{qij} + \epsilon_{ij}, \tag{19}$$

$$\epsilon_{ij} \sim N(0, \sigma^2), \tag{20}$$

$$\alpha_i \sim \textstyle\sum_{c=1}^{C} N_2(\mu_c, \Sigma_k). \tag{21}$$

In the above formulas, N_2 represents a bivariate (multivariate) normal distribution; (μ_c, Σ_k) are the hyperparameters for the mean and the covariance matrix, respectively; and σ^2 is the variance of the error term in the level 1 model. The $\{u_{ij}, v_{1ij}, \dots, v_{qij}\}$ parameters are known covariates associated with observed data $\{y_{ij}\}$. $\alpha_i = (\alpha_{1i}, \alpha_{2i})^T$ represents a matrix of random effects. For each random-effects vector α_i, a latent group label variable z_i is provided, which takes values $\{1, \dots, C\}$. This group label variable (or allocation variable) is drawn from a distribution that is specified in Equation 11.

Similar to the univariate normal mixture models, the estimation of this Gaussian mixture random effects model can be considered as a general semi-parametric density analysis using $\alpha_i | z_i$ as a Gaussian random vector to construct efficient Gibbs samplers (Ho & Hu, 2008). The joint distribution of all parameters in Equation 13, now become

$$\begin{aligned} &p(y, \alpha, \beta, \sigma^2, \pi, c, \Phi, z) = \\ &p(\sigma^2, \beta) p(c) p(\Phi | c) p(\pi | m) p(z | \pi, m) p(\alpha | \Phi, z) \times p(y | \alpha, \beta, \sigma^2), \end{aligned} \tag{22}$$

where,

$$\begin{aligned} y = \{y_{ij}\}, \alpha = \{\alpha_1, \dots \alpha_n\}, \beta = \{\beta_1, \dots, \beta_q\}, \pi = \{\pi_1, \dots \pi_C\}, z \\ = \{z_1, \dots z_n\}, \text{and } \Phi = \{(\mu_1, \Sigma_1), \dots, (\mu_C, \Sigma_C)\}. \end{aligned} \tag{23}$$

Ho and Hu (2008) adopted the methods from Richardson and Green (1997) for calculating the acceptance probability for the moves, which is specified in Equations 17 and 18. However, the dimension of the model parameters now has increased as the model changed from univariate to multivariate. Therefore, the "combine" and "split" moves in the RJMCMC process also need to be adjusted for the computation of the moment conditions. Specifically, Ho and Ho (2008) calculated the moment conditions in a "combined" move for the Gaussian random effect mixture model in the following way:

$$\pi_j = \pi_{j1} + \pi_{j2}, \tag{24}$$

$$\pi_j \mu_j = \pi_{j1}\mu_{j1} + \pi_{j2}\mu_{j2}, \tag{25}$$

$$\pi_j\left(\Sigma_j + \mu_j\mu_j^T\right) = \pi_{j1}\left(\Sigma_{j1} + \mu_{j1}\mu_{j1}^T\right) + \pi_{j2}\left(\Sigma_{j2} + \mu_{j2}\mu_{j2}^T\right). \tag{26}$$

The above expressions are the RJMCMC transformation for generating the zero*th*, first, and second moments for the density of a "combined" component based on its previous state. Specifically, Equation 24 creates the class proportions of the new component, Equation 25 creates the expectation of the density of the new component (i.e., mean), and Equation 26 creates the covariance matrix of the density.

The "split" move takes a more complex form, where an auxiliary variable u is introduced in order to assist with the dimension matching as the dimensions of the model parameters increased in a split move. In Ho and Hu (2008), u was set to be drawn from a beta distribution such that $u \sim beta(2,2)$. Then two proportion components parameters for the new component can be created by:

$$\pi_{j1} = u\pi_j, \pi_{j2} = (1-u)\pi_j. \tag{27}$$

Then the expectations of the new component are created by Cholesky decomposition

$$\Sigma_j = L_{\boldsymbol{j}} L_j^T, \tag{28}$$

where

$$L_j = \begin{pmatrix} l_{11}^j & 0 \\ l_{21}^j & l_{22}^j \end{pmatrix}. \tag{29}$$

A random 2×1 vector $\nu = (\nu_1, \nu_2)^T$ is created for splitting the mean vector for the new component. The ν_i parameter in Ho and Hu (2008) was set in a way such that $\nu_i = \frac{\pi_i}{(1+\pi_1^2+\pi_2^2)^{\frac{1}{2}}}$, and that the Euclidean norm $||\nu|| < 1$. Therefore, the means are

$$\mu_{j1} = \mu_j - \sqrt{\frac{1-u}{u}} L_j \mathrm{v}, \tag{30}$$

$$\mu_{j2} = \mu_j + \sqrt{\frac{1-u}{u}} L_j \mathrm{v}. \tag{31}$$

The covariance matrix for the new component then can be computed with assistance from some extra auxiliary variables as follows:

$$\pi_j\left(\Sigma_j + \mu_j \mu_j^T\right) = \pi_{j1}\left[\Sigma_{j1} + \left(\mu_j - \sqrt{\frac{\pi_{j2}}{\pi_{j1}}} L_j \nu\right)\left(\mu_j - \sqrt{\frac{\pi_{j2}}{\pi_{j1}}} L_j \nu\right)^T\right]$$
$$+ \pi_{j2}\left[\Sigma_{j2} + \left(\mu_j - \sqrt{\frac{\pi_{j1}}{\pi_{j2}}} L_j \nu\right)\left(\mu_j - \sqrt{\frac{\pi_{j1}}{\pi_{j2}}} L_j \nu\right)^T\right], \tag{32}$$

$$\pi_j\left[\Sigma_j - L_j \nu \nu^T L_j^T\right] = \pi_{j1}\Sigma_{j1} + \pi_{j2}\Sigma_{j2}. \tag{33}$$

Let $\boldsymbol{Q}_j = \Sigma_j - \boldsymbol{L}_{j\nu\nu^T}\boldsymbol{L}_{\boldsymbol{j}}^{\boldsymbol{T}}$ and by Cholesky decomposition, $\boldsymbol{Q}_j = \boldsymbol{C}_j \boldsymbol{C}_j^T$, where $\boldsymbol{C}_j$ is the lower triangular matrix with positive diagonal entries. A positive definite matrix $\boldsymbol{H}$ is drawn from a multivariate β distribution such that $\boldsymbol{H} \sim MVB\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ with $n_1 \geq d$ and $n_2 \geq d$, where d is the dimension of the parameter. Then the covariance matrices of the density for the new component are

$$\Sigma_{j1} = \frac{\pi_j}{\pi_{j1}} C_j H C_j^T, \tag{34}$$

$$\Sigma_{j2} = \frac{\pi_j}{\pi_{j2}} C_j (I - H) C_j^T. \tag{35}$$

Ho and Hu (2008) evaluated the performance of the RJMCMC algorithm for the Gaussian mixture random effects model on simulated data sets. In the simulation study, they examined the effect of the hyperparameter values of two parameters: 1) the precision of the normal prior distribution for the fixed effect coefficient μ_j, and 2) the positive definite matrix of the inverse Wishart prior distribution for the covariance matrix. Their findings were consistent with previous studies (e.g., Richardson and Green, 1997 and Ho, 1995) in that the posterior number of components of the random effects distribution would be moderately affected by the hyperparameter values, while the fixed effects estimation results were almost unchanged.

The specification of the prior distributions for the Bayesian inferences was based on Richardson and Green (1997) and Lee and Song (2003). The Dirichlet prior distribution was assigned on the mixture component parameter, and both informative and uninformative priors were assigned for other model parameters. In the RJMCMC implementation process, they followed the steps and moves in the previous studies (e.g., Richardson & Green, 1997; Boys & Henderson, 2001; Papastamoulis & Iliopoulos, 2009; and Roberts et al,2000.) and proposed the combined and split moves for the dimension-changing in the mixture structure equation modeling (SEM).

Unlike the treatments of the steps in the implementation in some of the previous studies, Liu and Song (2017) discussed the unnecessary use of the birth-and-death step that was originally proposed in Green (1995) and was commonly used in the RJMCMC literature. They hence excluded the birth-and-death steps from their implementation.

In their simulation studies, Liu and Song (2017) examined recovery of the number of mixture components (e.g., comparing a one, two, and four class solution) using different prior specifications on model parameters. They also calculated the deviance information criterion for the models with different numbers of mixture components, and they compared the DIC approach with the RJMCMC methods. They found that the RJMCMC algorithm was highly computationally efficiency and that it yielded relatively small bias levels and root mean square values. Their simulation results concluded that the model selection and estimation results were not very sensitive to the prior specifications under consideration.

RJMCMC has also been widely used in Bayesian model averaging (e.g., Green, 1995; Hastie & Green, 2012; Huelsenbeck, Larget, & Alfaro, 2004; etc.), neural networks (Andrieu, de Freitas, & Doucet, 2001; Holmes & Mallick, 1998) and signal processing (Andrieu & Doucet, 1999; Larocque & Reilly, 2002). The advantages of using RJMCMC for Bayesian mixture modeling were discussed in several studies (e.g., Richardson and Green, 1997; Ho and Hu, 2008, Liu and Song, 2017, etc.). First, RJMCMC yields convenient, accurate and flexible outcomes compared with other analytic approximations or MCMC techniques. Second, mixture modeling is conventionally considered as the estimation of separate models. Model comparison criteria or other non/semi-Bayesian tests are used to infer the number of mixture components. On the contrary, RJMCMC provides full Bayesian treatments of mixture estimation. It models not only the model parameters but also the number of mixture components by treating the number of mixture components as random variables that are drawn from a distribution. Third, RJMCMC solves the technical issues that are associated with the sampling methods for the posterior distribution. RJMCMC enables the jumps between states that are of different dimensionalities. It provides the means of computing moments for the density functions when a new component is created.

Despite its popularity in the literature of statistics and computer sciences, RJMCMC has seldom been applied in social and behavioral sciences. Therefore, in this dissertation, I will introduce this Bayesian non/semi-parametric method to LGMMs, which are commonly implemented within the social and behavioral sciences. My main focus will be on using this approach to aid in the estimation of the number of latent classes. A specific emphasis will be placed on how this approach can benefit substantive research being conducted within Psychology and related fields.

In the next few subsections, I will present and discuss the procedure of the RJMCMC method for the LGMM in the two studies in this dissertation.

2.2.2 RJMCMC for LGMM

2.2.2.1 Prior Specifications

The formulation of the LGMM is presented in Equations 6-8, which can be written in the Bayesian hierarchical model format:

$$y_{it} = \lambda_t B_F + \mathrm{ID} B_R + \epsilon_{it}, \tag{36}$$

$$\epsilon_{it} \sim N(0, \sigma^2), \tag{37}$$

$$B_R \sim \sum_{c=1}^{C} \pi_{ic} MVN(\mu_c, \Sigma)\,, \text{for } i = 1, \dots, n; t = 0, \dots, 3, \tag{38}$$

where on the level-1 model, λ_t is a vector of time that is coded as $\lambda_t = 0{,}1{,}2{,}3$, and ID is a vector of the identification number of the observation coded as $\mathrm{ID} = 1, \dots, n$. B_F is a vector of fixed effects coefficients, and $B_R = (\alpha_{ic}, \beta_{ic})$ represents a matrix of random coefficients that can be modeled with a level-2 model, where B follows a multivariate normal distribution with the mean of μ_c and covariance matrix of Σ for each latent class *c*. On the level-2 model, $\mu_c = (\mu_{\alpha c}, \mu_{\beta c})$ represents the mean vector for the multivariate normal distribution in class *c*, whereas $\Sigma = \begin{pmatrix} \psi_{\alpha\alpha} & 0 \\ \psi_{\alpha\beta} & \psi_{\beta\beta} \end{pmatrix}$ represents the covariance matrix, which is invariant across the latent classes.

The prior distributions assigned on the model parameters in the LGMM are specified as follows.

- $\mu_c \sim MVN(\xi, D)$. The mean vector follows a multivariate normal distribution, with ξ as the hyperparameter mean and D as the hyperparameter covariance matrix.
- $\Sigma \sim IW(p, \tau)$. The covariance matrix follows the *inverse Wishart* distribution, with the degrees of freedom p and a positive definite matrix τ.
- $\sigma^2 \sim IG(b_1, b_2)$. The residual variance follows an *inverse gamma* prior with the shape and scale hyperparameters (b_1, b_2).
- $(\pi_1 \dots, \pi_C) \sim Dirichlet(\delta_1, \dots \delta_C)$. The mixture class proportions are drawn from the *Dirichlet* distribution with the hyperparameter δ, which is linked to the class size.
- $B_F \sim N(\mu_F, D_F)$. The fixed effect coefficients follow a normal distribution with mean hyperparameter μ_F and hyperparameter covariance matrix D_F.

2.2.2.2 Steps in the RJMCMC algorithm

Similar to the standard MCMC algorithm, the transitions of RJMCMC follow the detailed balance condition, which ensures that moves from State 1 to State 2 are made as often as moves from State 2 to State 1 (Gelman et al, 2014). Let x denote the state variable, and $p(dx)$ denote the target distribution, which represents the posterior distribution of the model parameters given the observed data. When the current state is x, with a move type m being made and the destination state x' being proposed, the joint distribution probability is $p(x, dx')$. We can construct the Markov transition kernel with $p(x, dx')$, which should satisfy the detailed balance condition: $\int_A \int_B p(dx)p(x, dx') = \int_B \int_A p(dx')p(x', dx)$, for all appropriate subspaces A and B (Green 1995).

Based on the Gibbs sampler and the Metropolis-Hastings method, the probability that the move (from state x to state x') is accepted can be calculated by Equation 17. With the measure $a_m(x, x')$, the move is accepted; otherwise, no move is attempted and the current values are retained.

The steps of an RJMCMC sweep can be constructed as follows:

- Initialize latent class *c* and corresponding model parameters, including $\pi_c, \mu_c, \Sigma, \sigma^2$ and the allocation z_i at iteration = 1.
- When iteration > 1, for the within-model move, within a fixed model *c*, update parameters $(\pi_c, \mu_c, \Sigma, \sigma^2)$.
- Still when iteration > 1, for the between-model move, conduct a "combine" or a "split" move that either combines two mixture components into one or splits one into two according to the acceptance mechanism described above. Simultaneously update other model parameters $(\pi_c, \mu_c, \Sigma, \sigma^2)$ and z_i.
- While the number of iterations < total number of iteration, increment iteration = iteration + 1 and repeat steps 2-4.

We adopted the approach from Ho and Hu (2008) for moments matching for the density in the new mixture component in the "combine" and "split" moves. Specifically, the zero*th* moment (proportion weight), the first moment (mean), and the second moment (covariance matrix) become

$$\pi_c = \pi_{c_1} + \pi_{c_2}, \tag{39}$$

$$\pi_c \mu_c = \pi_{c_1}\mu_{c_1} + \pi_{c_2}\mu_{c_2}, \tag{40}$$

$$\pi_c(\Sigma_c + \mu_c\mu_c^T) = \pi_{c_1}\left(\Sigma_{c_1} + \mu_{c_1}\mu_{c_1}^T\right) + \pi_{c_2}\left(\Sigma_{c_2} + \mu_{c_2}\mu_{c_2}^T\right), \tag{41}$$

in the "combine" move, where components c_1 and c_2 are combined into a new component *c*. In the "split" move,

$$\pi_{c_1} = u\pi_c, \pi_{c_2} = (1-u)\pi_c, \tag{42}$$

$$\mu_{c_1} = \mu_c - \sqrt{\frac{1-u}{u}} L_c \mathrm{v}, \tag{43}$$

$$\mu_{c_2} = \mu_c + \sqrt{\frac{1-u}{u}} L_c \mathrm{v}, \tag{44}$$

$$\Sigma_{c_1} = \frac{\pi_c}{\pi_{c_1}} C_c H C_c^T, \tag{45}$$

$$\Sigma_{c_2} = \frac{\pi_c}{\pi_{c_2}} C_c (I - H) C_c^T, \tag{46}$$

where component c is split into two new components c_1 and c_2. For the details of the transformation and the auxiliary variables, see Equations 29-35 and Ho & Hu (2008).

The acceptance probability specified in Equation 17 then can be computed as follows. Let A denote the acceptance probability, pr denotes the prior ratio, mp denotes the move probability, pp denotes the proposal distribution probability, and $|J|$denotes the determinant of the *Jacobian* matrix:

$$pr = \frac{p(m+1)}{p(m)} (m+1) \frac{\left(w_{j1}^{\delta-1+n_{j1}} w_{j2}^{\delta-1+n_{j2}} \right)}{w_j^{\delta-1+n_j} B(\delta, m\delta)}, \tag{47}$$

$$mp = \frac{1}{2\pi |D|^{\frac{1}{2}}} \exp[-\frac{1}{2} \{ \left(\mu_{j1} - \xi\right)^T D^{-1} \left(\mu_{j1} - \xi\right) + \left(\mu_{j2} - \xi\right)^T D^{-1} \left(\mu_{j2} - \xi\right) - \left(\mu_j - \xi\right)^T D^{-1} \left(\mu_j - \xi\right) \}], \tag{48}$$

$$pp = \frac{1}{2^r \pi^{\frac{1}{2}} |\Theta|^{\frac{r}{2}} \Pi_{i=1}^2 \Gamma(\frac{1}{2}(r+1-i))} \times \frac{\left|\Sigma_{j1}^{1}\right|^{\frac{r+3}{2}} \left|\Sigma_{j2}^{-1}\right|^{(r+3)/2}}{\left|\Sigma_j^{-1}\right|^{(r+3)/2}}, \tag{49}$$

$$|J| = \frac{w_j |\Sigma_j|^2 \left(1 - ||v||^2\right)^{3/2}}{s}, \tag{50}$$

$$A = (\text{lieklihood ratio}) \times pr \times \ mp \times pp \times |J|. \tag{51}$$

Then the acceptance probability for the "split" move is $min(1, A)$, and the acceptance probability of the "combine" move is $min(1, A^{-1})$.

2.2.2.3 Label Switching Issues

Label switching issues occur when the posterior distribution is invariant under permutations in the labeling of the parameter (Chung & Schafer, 2004). This results in the parameters having identical marginal posterior distributions. Label switching causes problems in the interpretation of the MCMC output since parameter estimates do not necessarily represent those from a single latent class.

One conceptually simple solution to the label switching problem is to impose artificial identifiability constraints on the parameters. For example, putting a constraint on the mean parameters, μ_c, of a normal mixture model such as $\mu_1 < \dots < \mu_c$. Another approach is to handle the non-identifiability in the post-processing of MCMC output. Some studies proposed inferential methods based on the relabeling of components or adopted fully decision-theoretic methods in order to minimize the posterior expected loss or construct an appropriate loss function (e.g., Stephens, 2000; Celeux et al, 2000; Hurn et al, 2003; Sisson and Hurn, 2004, Sisson and Fan, 2010, etc.). Each of these methods can be computationally expensive. In this dissertation, I have used the former approach where the constraint is put on the parameters to handle label switching.

2.3 Dirichlet Process Mixture Modeling for LGMM

The Dirichlet process models are a family of non/semi-parametric Bayesian models, which are commonly used for density estimation, non–parametric and semi-parametric modeling, and model selection or averaging. DP is non-parametric in the sense that it allows the model to contain an infinite number of parameters and "let the data speak for themselves." In other words, DP does not require the specification of the number of latent classes *a priori* because it assumes that the number is infinite. It also allows the mixture model to adapt (i.e., increase or decrease) the number of "active" (i.e., non-empty) classes as more data are fed in the model over time (Teh, Dirichlet Process, 2011). The non-parametric feature of DP can also be used in finite mixture models and I elaborate on this point in the following subsections.

2.3.1 Dirichlet Distribution

The Dirichlet distribution can be seen as a distribution over multinomial distributions. Specifically, it is a distribution over the C-dimensional simplex, $(\pi_1, \dots \pi_C)$, where $\Sigma_c \pi_c = 1$ and $\pi_c \geq 0$. $(\pi_1, \dots \pi_C)$ follows a Dirichlet distribution:

$$(\pi_1, \dots \pi_C) \sim Dirichlet(a_1, \dots a_C), \tag{52}$$

with parameters $(a_1, \dots a_C)$. The likelihood function of the Dirichlet distribution can be specified as:

$$p(\pi_1, \dots \pi_C) = \frac{\Gamma(\Sigma_c a_c)}{\Pi_c \Gamma(a_c)} \prod_{c=1}^{C} \pi_c^{a_c - 1}. \quad (53)$$

2.3.2 Dirichlet Process

The Dirichlet process is defined as a distribution over distributions, or a measure on measures. Let G be a function that is assumed to have infinite dimensions, and $G \sim DP(.|a, G_0)$. G_0 represents a base distribution. Based on the properties of the Dirichlet distribution discussed earlier, we can derive, for all $(\pi_1, \dots \pi_C)$,

$$G(\pi_1), \dots, G(\pi_C) \sim Dirichlet(a(G_0(\pi_1), \dots, a(G_0(\pi_C)). \quad (54)$$

Samples from a Dirichlet process are discrete with probability one:

$$G(\theta) = \sum_{c=1}^{\infty} \pi_{ic} f(\theta_c), \quad (55)$$

where $f(\theta_c)$ represents a generic mixture model that is assumed with to have infinite number of classes, and θ is the parameter vector of the model. One advantage of using DP as a prior distribution is that DPs are conjugate to themselves. That is, the posterior of a DP is also a DP:

$$P(G|\theta) = DP\left(\frac{a}{a+1} G_0 + \frac{1}{a+1} f(\theta), a+1\right). \quad (56)$$

DP has been widely used in statistics and machine learning. Teh (2010) summarized some simple and prevalent applications of DPs, including Bayesian model validation, density estimation and clustering via mixture models. In the validation of model fit process, a base model that is assumed to generate the observed data is compared with other possible models. The Bayesian non-parametric approach uses the space of all possible distributions in comparison with a prior over these distributions. DP is a popular prior distribution in this case. In the comparison process, a given parametric model is chosen as the base model, while DP serves as a non-parametric relaxation around this parametric model. If the parametric model performs as well as or better than DP relaxed model, then the validity of the model is convinced (Teh, 2010; also see e.g., Carota, Filippone, & Polettini, 2018). Another application of DP is in density estimation, where the interest lies in modeling the density from which a given set of observations is drawn. DP as a Bayesian nonparametric method is often chosen as a prior over all densities. When using DP for density estimation, the draws from the Dirichlet distribution are usually smoothed out with a kernel. (see e.g., Escobar & West, 1995; Lo, 1984) Let

$G \sim DP(.\,|a, G_0)$, and let $f(y|\theta)$ be a family of densities (kernels) indexed by θ. Then the nonparametric density of the observed data y can be presented as:

$$p(y) = \int f(y|\theta)G(\theta)d(\theta). \tag{57}$$

This smoothing function is also equivalent to the DP mixture models, which I will discuss in the following few sections.

2.3.3 DP Mixture Models

DP mixture models are commonly used in the context of infinite mixture models, where the mixture has an infinite but countable number of clusters or classes. The advantage of DP mixture models is that the number of classes is not fixed and can be automatically inferred from data using the Bayesian posterior inference framework (Neal,2000). For finite mixture models, where the number of classes is fixed, there can be different approaches to the inference of clustering. One approach is equivalent to model selection or model averaging for appropriate number of clusters (see e.g., Gershman and Blaei, 2011; Kim, Tadesse, and Vannucci, 2006; Wang and Dunson, 2011).

Another approach for the finite mixture models using DP is to treat the number of clusters as a very large value, which nearly approaches infinity. But note that even under the assumption that the number of clusters in a finite mixture model is nearly infinite, the actual number of active clusters (the clusters with at least one observation) cannot be greater than the number of observations. This treatment of the number of clusters allows us to apply DP on the finite mixture models and to avoid the complicated operation of the model selection and averaging approach.

Various computation solutions for implementing DP mixture models have been proposed in the fields of statistics and computer sciences. Blei and Jordan (2006) compared a mean-field inference algorithm to the Gibbs sampling methods for DP mixtures of Gaussians and presented an application to a large-scale image analysis problem. The authors used the variational inference method to compute the posterior distribution under a DP mixture prior. With a simulation study and an image data example, the authors concluded that the variational inference methods ran faster than the two Gibbs samplers (i.e., collapsed Gibbs and blocked Gibbs sampler). Gelfand and Kottas (2012) proposed a computational approach to obtain the posterior distribution for more general functions for the underlying distributions for mixture models using DP as a Bayesian nonparametric method. The authors investigated the extreme value distributions associated with a single population, compared the medians in a *k*-sample problem, and applied their methods in a survival data analysis example.

Teh et al (2012) applied DP on a hierarchical model (i.e., multilevel model) with groups of data where observations within each group were drawn from a mixture model. They assumed that the number of mixture components was unknown *a priori* and was to be inferred from the data. DP was used as a non-parametric prior on the parameters of the hierarchical mixture model. This paper presented three schemes of the construction of the hierarchical DP, including the stick-breaking representation, the “Chinese restaurant franchise” representation, and a presentation of the process in terms of an infinite limit of finite mixture models. The authors demonstrated the application of the hierarchical DP mixture model on document modeling. They also compared the performance of the hierarchical DP combined with hidden Markov models (HMM) with other non-DP HMM approaches. Based on the results of their experiments, this paper concluded that this hierarchical nonparametric Bayesian approach to clustering provided a generally useful extension of model-based clustering.

Miller and Harrison (2016) used the analogues of the mixture of finite mixtures (MFM) to estimate the unknown number of mixture components for DP mixture models. The paper reviewed some essential properties of MFM, including the clustering algorithms that were also exhibited in DP mixture models. The authors compared the MFMs and the DP mixture methods through a simulation study and the discriminate cancer subtypes data. They concluded that the methods for inference in DP mixtures (such as the exchangeable partition distribution, the Chinese restaurant process, the random discrete measure formulation, etc.) could be implemented on MFM in a simpler manner.

Despite the application of DP mixture models in statistics and machine learning (see Maceachern & Muller, 1998, Gelfand, Kottas, & MacEachern, 2005, Jiang, Kulis, & Jordan, 2012, etc. for more examples), it has not been introduced into the structural equation modeling framework. Therefore, I am proposing to adopt this Bayesian non-parametric method and apply it to the LGMM.

2.3.4 Model Specification for a DP Mixture Model

We assume a generic mixture model with a density that can be specified as:

$$f(y) = \sum_{c=1}^{c_{max}} \pi_c f(y|\Delta_c), \tag{58}$$

where $f(y|\Delta_c)$ is a generic function of the model with the parameter vector, Δ_c, and π_c denotes the mixture weights. A mixed-effects normal (or Gaussian) mixture model can be parameterized in the form of the DP mixture modeling in the following way:

$$y_i \sim F, \tag{59}$$

$$F \sim G, \tag{60}$$

$$G \sim \mathrm{DP}(aG_0), \tag{61}$$

$$G_0 \sim N(\mu_c, \Sigma), \tag{62}$$

where F is a density function, in this dissertation, a mixed-effects normal function with

$$F = \alpha_{ic} + \lambda_{tc}\beta_{ic} + \epsilon_{itc}, \tag{63}$$

$$\alpha_{ic} = \mu_{\alpha c} + \zeta_{\alpha_{ic}}, \tag{64}$$

$$\beta_{ic} = \mu_{\beta c} + \zeta_{\beta_{ic}}. \tag{65}$$

This model formulation can be illustrated with a diagram in Figure 1.

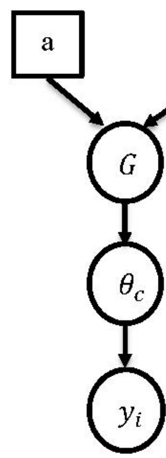

Figure 1. Diagram of a Generic DP Mixture Model

In this diagram, the observation data y_i given the parameter vector $\theta_c$follows the mixed-effects normal function F, that is, $y_i|\theta_c \sim F(\theta_c)$. G is a distribution of θ_c and $G \sim Dirichlet(a, G_0)$. In other words, the mixture models are sampled from DP with parameter a and G_0, with $\theta_c \sim G$ independently for *c = 1,…,* ∞, and G_0 denotes the base distribution that θ_c is drawn from. a is the dispersion parameter, or precision parameter that determines the number of clusters or classes we are going to obtain. Larger values for a tend to lead to a greater number of clusters. G_0 can be further modeled as $G_0 = \sum_{c=1}^{\infty} \pi_c \delta_{\theta_c}$. δ specifies that when $\delta = \delta^*$, this term will take the value 1, otherwise it will take zero. $\delta^*$is the parameter estimate for the class membership of a specific class. In an infinite mixture model, G_0 is the unknown mixing measure. The finite mixture models can be considered as a special case of infinite mixture models. In this case, G_0 can be treated as discrete with masses at a finite number of *c* components, *C*, and then we can obtain a finite mixture model as was discussed in the previous section (Gelman, et al., 2014).

2.3.5 Algorithms of DP Mixture Models

2.3.5.1 Blackwell-MacQueen Urn Scheme

The Blackwell-MacQueen urn scheme is commonly used to represent a Dirichlet process. It is based on the Polya urn model that can be considered as the opposite of sampling with replacement. In the Polya um model, we assume that we have an urn with colored marbles and we draw the marbles randomly from the urn. Each time we draw a marble, we observe the color. We put it back in the urn and add an additional marble of the same color. We use the similar scheme proposed by Blackwell-MacQueen (Blackwell & MacQueen, 1973) to construct the Dirichlet process.

Consider drawing an i.i.d. sequence of $\theta_1, \dots, \theta_n$ from G, where $G \sim DP(a, G_0)$. The conditional probabilities of drawing a θ_n given its previous draws is $\theta_{n+1}|\theta_{1:n} \sim \frac{aG_0 + \Sigma\delta_{\theta_i}}{a+n}$. In this scheme, we assume $G_0$is a distribution over colors and each θ_n represents the marble that is added in the urn. The posterior base distribution given $\theta_1, \dots, \theta_n$ is also the predictive distribution of θ_{n+1}. This sequence of predictive distributions for $\theta_1, \dots, \theta_n$forms the MacQueen urn scheme.

Given the conditional probabilities of $\theta_1, \dots, \theta_n$, we may construct a distribution over sequences $\theta_1, \dots, \theta_n$ by iteratively drawing each θ_i given $\theta_1, \dots, \theta_m$. The joint distribution of $\theta_1, \dots, \theta_m$ is invariant to any finite permutations, and thus it is exchangeable (see Teh, 2010 for the details of the proof). Then according to Finetti's theorem, for any infinitely exchangeable sequence $\theta_1, \dots, \theta_n$, there is a random distribution G such that the sequence is composed of i.i.d. draws from it:

$$P(\theta_1, \dots, \theta_n) = \int \prod_{i=1}^{n} G(\theta_i) dP(G). \tag{66}$$

In this specification, the prior over the random distribution $P(G)$ is the Dirichlet Process $DP(a, G_0)$.

2.3.5.2 Chinese Restaurant Process

The Blackwell-MacQueen scheme can be proved mathematically equivalent to the Chinese restaurant process (CRP). CRP describes the distribution over partitions. Imagine we have an empty restaurant with an infinite numbers of tables. Then here comes the first costumer. The probability that this costumer will go to the first table is 1 and to other tables is 0. Then comes the 2nd costumer, with the probability she will go to the first table and share with the 1 customer begin set at $\frac{1}{1+a}$. The probability she will take a new table is $\frac{a}{1+a}$. Then the $(n+1)th$ customer always has 2 choices: she can either share an occupied table or take a new table. The probability of taking a new table is $\frac{a}{n+a}$, and the probability of sharing an occupied table is $\frac{c}{n+a}$; *n* is the number of customers before her, and *c* is the number of customers already sitting at the table. Essentially, the Chinese restaurant process is the distribution over the partition.

When we use DP on a mixture model, we assume that P, in this case the number of mixture components, is infinite. Then we no longer have joint conjugacy in which the posterior of P since $y^n = y_1, \dots y_n$. A solution to this problem is to marginalize out P to obtain an induced prior distribution on the model parameter θ. By doing this, we obtain the Polya urn predictive rule (Gelman, et al., 2014),

$$p(\theta_i|\theta_1, \dots \theta_{i-1}) \sim \left(\frac{a}{a+i-1}\right) P_0(\theta_i) + \sum_{i=1}^{n} \left(\frac{1}{a+i-1}\right) \delta_{\theta_i}. \tag{67}$$

This conditional prior distribution consists of a mixture of the base measure P_0 and probability at the previous subject's parameter values. The Chinese restaurant process describes this scheme. Consider a restaurant with an infinite number of tables. The first customer sits at a table with dish θ_1. The second customer sits at the first table with probability $\frac{a}{a+1}$ or a new table with probability $\frac{1}{a+1}$. This process continues with the *i*th customer sitting at an occupied table with the probability proportional to the number of previous customers at that table and sitting at a new table with probability proportional to α. In this process, each occupied table in the restaurant represents a different cluster of subjects, with new clusters added at a rate proportional to $\alpha \log(n)$ in the asymptotic limit. Therefore, the number of clusters depends on the number of subjects *n* with new clusters introduced as needed and as additional subjects are added to the sample (Miller & Harrison, 2016).

2.3.5.3 Gibbs Sampler

The Chinese restaurant process can be implemented using Gibbs sampler:

1) First randomly assign the cluster ID z_i to a ξ, while keep other cluster IDs fixed.
2) Calculate the probability using the CRP algorithm to determine which cluster should this observation ξ should be associated with, and then assign a new cluster ID z_j to ξ, where $j \neq i$.
3) Update the parameters.
4) Repeat this process until it converges.

2.3.5.4 Prior Specification for DP

The formulation of the LGMM, which is presented in Equations 6-8, can be written in the Bayesian hierarchical model format:

$$y_{it} \sim N(\lambda_t B_F + \mathrm{ID} B_R, \epsilon_{it}), \tag{68}$$

$$\epsilon_{it} \sim N(0, \sigma^2), \tag{69}$$

where

$$B_R \sim G, \tag{70}$$

$$G \sim DP(a, G_0), \tag{71}$$

and

$$G_0 \sim \sum_{c=1}^{C} \pi_{ic} MVN(\mu_c, \Sigma)\,, \text{for } i = 1, \dots, n; t = 0, \dots, 3. \tag{72}$$

The prior distributions assigned to the model parameters in the LGMM can be specified as follows:

- $\mu_c \sim MVN(\xi, D)$. The mean vector follows a multivariate normal distribution, with ξ as the hyperparameter mean and D as the hyperparameter covariance matrix.
- $\Sigma \sim IW(p, \tau)$. The covariance matrix follows the *inverse Wishart* distribution, with the degrees of freedom p and a positive definite matrix τ.
- $\sigma^2 \sim IG(b_1, b_2)$. The residual variance follows an *inverse gamma* prior with the shape and position hyperparameters (b_1, b_2).
- $B_F \sim N(\mu_F, D_F)$. The fixed effect coefficients follow a normal distribution with mean hyperparameter μ_F and hyperparameter covariance matrix D_F.
- $B_R \sim G$ and $G \sim DP(a, G_0)$. G represents a probability function, specifically a normal distribution, over a parameter space and can be written as $G \sim DP\big(a, N(\mu, \Sigma)\big)$, where a is the dispersion parameter of the DP mixture, which can take on a single value, such as 1, or follow a gamma distribution such that $a \sim Gamma(a_1, a_2)$. In the probability density function of normal mixture models (i.e., Gaussian mixture models) G can be marginalized out. Taking the integral of function G with respect to μ, the random effects coefficient B_R then follows an integrated function such that $B_R \sim \int N(\mu, \Sigma_b)(d\mu)$. In this distribution, $\Sigma_b \sim IW(\nu_b, T_b)$.

Chapter 3

Study 1: Deciding on the Number of Classes for LGMM using Bayesian Non/Semi -Parametric Methods

3.1 Introduction

LGMM is a useful method in the structure equation modeling (SEM) framework for modeling latent classes in longitudinal data. Despite its usefulness in the social and behavioral sciences, one major challenge is deciding on the number of latent classes that exist; there are several techniques used to help drive this decision (e.g., comparative indices such as information criteria). Parallel to the existing methods used to decide on the number of classes, Bayesian non/semi-parametric methods, such as RJMCMC and DP, have been developed and popularized in statistics and computer science. These two modeling techniques may prove useful within the LGMM framework by contributing to the practice of "testing" the number of latent classes.

The primary goal of Study 1 is to examine the performance of RJMCMC and the DP on extracting the number of latent classes for LGMM. Another goal is to compare these two Bayesian non-parametric methods with the traditional (i.e., frequentist) implementation of the LGMM. Study 1 is structured as follows. First I will briefly discuss the traditional approaches to assessing the number of latent classes for mixture models and the issues (i.e., problems) linked to them. Then I will present a simulation study that examines the performance of RJMCMC and the DP and compare these approaches with the frequentist estimation framework. The simulation study is followed by an empirical example in which RJMCMC and the DP are applied to an LGMM using a real life data set. Study 1 will conclude with a discussion of the performance of the Bayesian non-parametric methods for LGMMs, as well as a discussion of the implications for applied researchers implementing these methods.

3.1.1 Assessing the Number of Classes for Finite Mixture Models

When researchers have no *a priori* information about the classes (e.g., the substantive differences between the classes, the number of classes, or the substantive interpretation of classes), then the number of classes has to be inferred from the data; the parameters in the component (i.e., class) densities are estimated from the data. In this case, the selection of the number of classes is akin to cluster analysis (Roeder & Wasserman, 1997). Several problems may arise from this cluster analysis approach. First, the separation (or distinction) between classes must be large enough in order to detect that they are indeed distinct classes. This large degree of separation is not always the case in applied settings, and classes can be difficult to identify when the sample size is relatively small (McLachlan & Peel, 2000). Second, the distribution of the data may influence class separation. For example, non-normally distributed repeated measures may lead to the extraction of multiple classes when only a single class exists in the population (Bauer & Curran, 2003).

Another approach to selecting the number of classes is to use a number of statistical tests and model fit measures. Some popular information-based fit indices are the Bayesian information criterion (BIC), the sample size adjusted BIC (SBIC; Sclove, 1987), Akaike's information criterion (AIC; Akaike, 1987), and the consistent AIC (CAIC; Bozdogan, 1987). In addition to the information-based indices, the mixture models can also be examined and compared using the nested model tests, such as the likelihood-ratio (LR) test and the Lo-Mendell-Rubin test (LMR; Lo, Mendell, & Rubin, 2001), as well as some goodness of fit tests (e.g., the multivariate skewness test (MST) and the multivariate kurtosis test (MKT) proposed by Muthén, 2003).

Simulation studies that examined some of these fit indices within the LGMM context arrived at conflicting conclusions. For example, Tofighi and Enders (2008) suggested that SBIC and LMR consistently performed well, whereas other tests and measures (e.g., BIC and CAIC) tended to provide inconsistent information on the number of extracted latent classes. Tofighi and Enders (2008) also examined the bootstrap likelihood ratio test (BLRT; McLachlan & Peel, 2000) and found that it outperformed the other indices in detecting the correct number of latent classes. Nylund et al. (2007) obtained some contrasting results regarding LGMM class enumeration. For instance, the BIC was found to be superior to all other information criteria for LGMMs for correctly identifying the true number of classes. They also found that the BLRT performed as well as BIC, and both of these indices performed better than the others in the mixture model context.

Despite their popularity, the statistical tests and model fit measures also have several challenges when being used for class enumeration purposes. The major issue is the disagreement among the statistical tests and the fit measures, which can make the determination of the number of classes highly subjective and sometimes difficult (or even impossible). The simulation studies by Tofighi and Enders (2008) and Nylund et al. (2007) can be considered as the two primary investigations on model comparison measures. The model fit indices and statistical tests examined in these two studies are regularly used for deciding on the number of latent classes in the SEM literature. However, as detailed earlier, Tofighi and Enders (2008) and Nylund et al. (2007) concluded with contradicting findings on which comparison measure(s) performed the best for class enumeration for LGMMs. The contradictions in simulation findings may cause confusion and difficulties for substantive researchers trying to decide on the number of latent classes when implementing the LGMM.

Another issue with the model comparison measures is that the result of the class solution can be sensitive to the model estimation method being implemented. For instance, the starting values for maximum likelihood estimation and the informativeness of the prior specification for Bayesian estimation can each impact the final model results and class structure obtained. Moreover, the performance of statistical tests and fit measures can depend on factors such as class separation, sample size, and class proportions (Nylund, et al, 2007). This artifact makes the implementation of these measures highly dependent on the specific settings or features of a particular application. Thus, generalizability of the performance of these indices may not be appropriate.

As I have discussed, the traditional clustering methods and the model selection approach, which are dependent on the statistical tests and fit measures, have their disadvantages and challenges. Therefore, in this study, I propose the alternative approaches to selecting the number of latent classes via Bayesian non/semi-parametric methods. RJMCMC and DP, as non/semi-parametric methods, are able to circumvent the issues rooted in the traditional clustering and model selection approaches when determining the number of classes. Most importantly, RJMCMC and DP do not require a presumption of the number of classes; instead, this is *estimated*. Therefore RJMCMC and DP may effectively avoid the potential contradicting conclusions that are derived from the comparisons of different models based on some statistical tests or fit measures. This feature of RJMCMC and DP can make them more efficient and straightforward compared with the traditional approaches. In the next section, I detail the simulation design used to examine these claims more thoroughly.

3.2 Design of Study 1

In Study 1, I evaluated the ability of RJMCMC and DP to accurately detect the number of latent classes for LGMM via a simulation study. I also compared these two Bayesian non-parametric methods with the traditional model estimation approach, which is the maximum-likelihood estimation through the expectation maximization algorithm (ML/EM) with regards to the accuracy of parameter recovery.

3.2.1 Population Values

In this simulation study, data were generated and analyzed using an LGMM with four time points, with one latent intercept and one latent linear slope. To minimize model complexity, the covariance structure was held equal across latent classes.[3] The generative model was constructed based on Equations 6-8. In total, 2000 replications of datasets were generated based on this model; a small sensitivity analysis was conducted to ensure the Monte Carlo study converged with 2000 replications.

I empirically derived the parameter population values for the data generation model from Kaplan (2002). Kaplan (2002) examined data from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) ([NCES], 2001) using a growth curve model with multiple time points. I used estimates for the latent factor means and covariance matrix from Kaplan (2002) as the population values for the intercept and the slope terms for the first latent class in this study. Population values for the remaining latent classes were determined statistically in order to create certain levels of class separation. Specifically, the multivariate Mahalanobis distance (MD) was used to measure the class separation between two adjacent latent classes.[4] The MD value was set to MD $=1.5$ for

---

[3] Note that this structure can also be allowed to vary across classes if desired.

[4] The multivariate MD is calculated by: $\mathrm{MD} = \sqrt{(\mu_1 - \mu_2)^T S^{-1} (\mu_1 - \mu_2)}$, where μ_1 and μ_2 are two vectors that represent the means of the latent growth factors (i.e., the intercept

all simulation conditions in Study 1 to mimic a situation with a moderate level of class separation. All of the population values that were used to generate the data are listed in Table 1, and data were generated using M*plus* 8 (Muthén & Muthén, 1998-2017).

Table 1. Population Values of Growth Parameters in Study 1 and Study 2.

Parameter	GCM	GMM-2Class				GMM-3Class				
		MD=1.5	MD=1	MD=2	MD=3	MD=1.5/1.5	MD=1/1	MD=1/3	MD=3/1	MD=3/3
Mean-I										
C1	31.370	31.370	31.370	31.370	31.370	31.370	31.370	31.370	31.370	31.370
C2	-	35.730	34.290	37.182	40.090	35.730	34.290	34.290	40.090	40.090
C3	-	-	-	-	-	40.090	37.210	43.010	43.010	48.810
Mean-S										
C1	1.802	1.802	1.802	1.802	1.802	1.802	1.802	1.802	1.802	1.802
C2	-	3.350	2.834	3.857	4.885	3.350	2.834	2.834	4.885	4.885
C3	-	-	-	-	-	4.880	3.867	5.917	5.917	7.968
Variance										
I	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000
S	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
Covariance	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Residual	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500

3.2.2 Simulation Conditions

Factors that were varied in this simulation study included: the sample size (3 levels), the number of latent classes (3 levels), and the class proportions (2 levels for 2-class conditions and 3 levels for 3-class conditions).

There were three levels of sample size included here as to mimic small, medium, and large data sets that are commonly found in the social and behavioral sciences literature, as well as the LGMM simulation literature (see e.g., Depaoli, Yang, & Felt, 2017; Depaoli & Boyajian, 2014). I selected the following sample size conditions to reflect the following levels: $n = 200$, $n = 400$, and $n = 600$. The three levels of the number of classes include: $C = 1$, $C = 2$, and $C = 3$. Condition $C = 1$ represents a latent growth curve model (i.e., without a mixture structure). Conditions $C = 2$ and $C = 3$ represent LGMMs with 2 latent classes and 3 latent classes, respectively. This factor allowed me to assess the performance of these estimation approaches with and without the presence of a mixture structure.

For the conditions implementing $C = 2$, I specified two levels of class proportions. The first level held classes at equal sizes in the population (Proportions = $50\%/50\%$), and the second assessed the impact of a minority class containing only 20% of the cases (Proportions = $80\%/20\%$). For the conditions implementing $C = 3$, I specified three levels of class proportions: Proportions = $33\%/33\%/33\%$, Proportions =

and the slope) in two adjacent latent classes, and S represents the covariance matrix of the latent factors.

45%/45%/10% (testing the impact of a true minority class), and Proportions = 70%/20%/10% (testing the impact of a true majority class).

3.2.3 Model Estimation Techniques

3.2.3.1 RJMCMC

Three different model estimation techniques were examined in Study 1, which were: RJMCMC, DP, and ML/EM. In the RJMCMC conditions, generated data were analyzed using the RJMCMC technique. The RJMCMC algorithm for analyzing the LGMM was developed based on the RJMCMC technique described in Ho and Hu (2008) that is detailed in the Introduction. This algorithm was created with R code and the functions for the analysis were modified based on the R package "miscF" (Feng, 2016).[5] The maximum number of models in the analysis process was set to 30. The program performed 5000 iterations for each MCMC run, with the first 2500 iterations designated as the burn-in phase; issues related to convergence are discussed in the Results section. The RJMCMC analysis model can be written in the Bayesian hierarchical model format (i.e., Equations 36-38). The prior specifications of the model parameters represent weakly informed prior distributions, which indicate some degree of uncertainty or not having "sufficient" information in the nature of an exploratory study (Gelman, et al, 2008). The prior distributions can be specified as follows:

- μ_c represents the vector of the means of the growth parameters (i.e., the latent intercept and slope terms). This vector follows a multivariate normal distribution, $\mu_c \sim MVN(\xi, D)$, where ξ represents the mean hyperparameter vector of the MVN prior distribution and D represents the covariance matrix hyperparameter of the MVN prior distribution.
    - For all the two-class model conditions, $\xi = (33{,}2)$. The first element of the vector ξ represents the mean hyperparemeter of the intercept; it is calculated by averaging the population values of the growth parameters across two latent classes. For example, the population values for the means of the intercept in a two-class model were 31.75 and 35.73 for C1 and C2, respectively (all population values are listed in Table 1). The mean hyperparameter of the intercept was therefore calculated by $\frac{(31.75+35.73)}{2} = 33.74$. Then the result, 33.74, was truncated to an integer, 33, and used as the hyperparameter of the intercept in the ξ vector. The second element of the ξ vector represents the mean hyparameter of the slope; it is calculated in the similar way as for the mean hyparameter of the intercept. That is, averaging the population values of the latent classes. In two-class conditions, the mean hyperaparameter of the slope was $\frac{(1.802+3.35)}{2} = 2.58$, then truncated to 2 and used as the second element of the ξ vector. In 3-class conditions, $\xi = (35{,}3)$, where 35 is the mean hyperparameter of the intercept and 3 is the mean hyperparameter of the slope. They were calculated in the similar way as in the two-class conditions,

[5] A proof of concept simulation was conducted to ensure that all code was working correctly. More information about this proof of concept can be found in the next section.

Specifically, the mean hyperparameter of the intercept was $\frac{(31.75+35.73+40.09)}{3} = 35.86$, then was truncated to 35 as the mean hyperparameter of the intercept in the ξ vector. The mean hyperparameter of the slope was $\frac{(1.802+3.35+4.88)}{3} = 3.34$, then was truncated to 3. These two mean hyperparameters formed the ξ vector (35, 3) for the mean hyperparameters for the growth factors in the 3-class model conditions. The hyperparameters of the growth factors (i.e., intercept and slope) were created in this way because RJMCMC did not assume and specify the number of latent classes a priori. Therefore, the hyperparameters were not implemented on the specific growth parameters in each latent class. In other words, there was only one set of mean hyperparameters for the growth parameters specified in the prior implementation (i.e., one for the intercept and one for the slope), regardless of the number of latent classes. To average the population values across all latent classes was a simple and reasonable way to derive the hyperparameters.

  - $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$ were implemented on the covariance matrix, where the variances were fixed at 10 and the covariance at 0 as to provide some degrees of infomrativeness.

- Σ is the covariance structure of the growth parameters and $\Sigma \sim IW(p, \tau)$, where $p = 2$, representing the number of parameters in the covariance matrix, and $\tau = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$; this specification represents an IW prior distribution with little information.
- σ^2 is the residual variance on level-1 model and $\sigma^2 \sim IG(b_1, b_2)$, where $b_1 = .01$ and $b_2 = .01$, which represents an uninformative prior distribution.
- B_F is a fixed effect coefficient and $B_F \sim N(\mu_F, S_F)$, where $\mu_F = 0$, and $S_F = 10$. This specification of hyperparameters provides a normal distribution with very little information.
- The mixture class proportions π were drawn from a Dirichlet distribution where $(\pi_1 \ldots, \pi_C) \sim Dirichlet(10, \ldots 10)$, where the values of π represent the number of cases in each latent class, and this particular specification of the prior distribution can be considered uninformative since very few cases are assigned to a given class under this specification; note that $\pi = 10$ was the default Dirichlet prior implemented in Mplus as to avoid the situations (such as when $\pi_c \sim Dirichlet(1,1)$) in which the formation of small or inadmissible class solutions might occur. (Muthén & Muthén, 1998-2017).

3.2.3.2 DP

The second Bayesian non-parametric model estimation technique being examined is DP. In the DP conditions, I implemented the DP process mixture modeling technique for the LGMM, and I used the R package “DPpackage” (Jara, Hanson, Quintana, Mueller, & Rosner, 2017) for data analysis. In each MCMC run, I requested 5000 total iterations in the chain, with the first 2500 iterations discarded as the burn-in phase. In the DP conditions, I also implemented the weakly informed prior distributions on the model parameters, which are described as follows:

- μ represents the vector of means for the latent growth parameters (i.e., the intercept and the slope). It follows a multivariate normal distribution $\mu_c \sim MVN(\xi, D)$, where $\xi = (33{,}2)$ for 2-class conditions and $\xi = (35{,}3)$ for the 3-class conditions, and $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$. The hyperparameters in the DP conditions were specified and calculated in the same manner as in the RJMCMC conditions. Specifically, the mean hyperparameters of the intercept and the slope were calculated by averaging the population values of the means of the intercept and the slope across latent classes, respectively. The hyperparameter of the covariance matrix represents an informative prior specification.
- Σ is the covariance structure of the growth parameters and $\Sigma \sim IW(p, \tau)$, where $p = 2$ and $\tau = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$. The degrees of freedom p and the τ parameter of the IW prior represent an uninformative prior specification.
- σ^2 is the residual variance on level-1 model and $\sigma^2 \sim IG(b_1, b_2)$, where $b_1 = .01$ and $b_2 = .01$, which represents an uninformative prior specification.
- B_F is a fixed effect coefficient and $B_F \sim N(\mu_F, S_F)$, where $\mu_F = 0$, and $S_F = 10$, which represents an uninformative prior specification.
- $B_R \sim G$ and $G \sim DP(a, G_0)$. G represents the linear normal mixture where the parameters were drawn from, and G_0 represents a base distribution of the DP mixture, which is a normal distribution in this study. G therefore follows a DP distribution written as $G \sim DP\big(a, N(\mu, \Sigma)\big)$, where a is the dispersion parameter of the DP mixture, which takes on a value of 1 in this study. In the probability density function of normal mixture models (i.e., Gaussian mixture models), after G has been integrated out (with respect to μ), the random effects coefficient B_R then follows an integrated function such that $B_R \sim \int N(\mu, \Sigma_b)(d\mu)$. In this distribution, $\Sigma_b \sim IW(\nu_b, T_b)$, where $\nu_b = 2$ and $T_b = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$. In this IW prior specification, ν_b represents the degrees of freedom, which is equal to the number of parameters, 2, and T_b represents an uninformative prior specification.

3.2.3.3 ML/EM

The third model estimation technique I evaluated in this study is the conventional frequentist approach, ML/EM, which was used to compare to the Bayesian non-parametric methods. In the ML/EM conditions, data analysis was implemented using M*plus* 8. The number of perturbations of user-specified starting values was set at 100, and the number of final stage optimizations was set at 25. This ratio was used in order to achieve proper convergence for LGMMs (see e.g, Hipp and Bauer, 2006).

All of the manipulated factors in the simulation were fully crossed with these three model estimation techniques, which resulted in 18 cells for each estimation technique.

3.3 Proof of Concept Simulation Study for RJMCMC and DP

In order to illustrate the performance of the software programs that were used to conduct the analysis for RJMCMC and DP, I included a proof of concept simulation study in this section. The goal of this proof of concept study was to ensure that the two programs, including the R code that was adopted from the R package "miscF" (Feng, 2016) for RJMCMC and the functions from R package "DPpackage" (Jara et al, 2017) for DP produce valid and reasonable results for the simulation studies in this dissertation. So far, there is no existing R packages or functions for applying RJMCMC on LGMMs so the code that was developed and used in this dissertation was a novel approach. The R script for conducting RJMCMC and DP for LGMM is available upon request from the author.

The proof of concept study was conducted through a brief simulation, where the same models that were examined in the dissertation were used to generate and analyze the data. In this brief simulation study, I generated one data set of $n = 10000$ cases (i.e., mimicking a population study) using a two-class growth mixture model with 50/50% class proportion and MD = 1.5 class separation. Data were generated using M*plus* version 8. The population values of the generative model were specified as follows. The means of the intercept and the slope were 31.37 and 1.802 in Class 1 and 35.73 and 3.35 for Class 2. The variances for the growth parameters were 16 and 2 and the covariance between the intercept and the slope was 0.3. The residual variance was 0.5. All the population values are listed in Table 1.

Then I tested the programs for RJMCMC and DP on the generated data using the two-class growth mixture model. For the RJMCMC condition, the analytic model is specified in Equations 36-38. The prior specifications are as follows. $\xi = (33,2)$ and $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$; $\Sigma \sim IW(2, \tau)$ and $\tau = [\begin{matrix} 10 & 0 \\ 0 & 10 \end{matrix}]$; $\sigma^2 \sim IG(0.01,0.01)$; $B_F \sim N(0,10)$; and $(\pi_1 \ldots, \pi_C) \sim Dirichlet(10, \ldots 10)$. A detailed description and explanation of the prior specifications are included in the Methods section in Study 1 (Section 3.2.3). In the RJMCMC model, one Markov chain with 5000 iterations was requested, of which the first 2500 iterations were discarded as burn-in. In the DP condition, the generated data were analyzed on the model specified in Equations 58-65. The priors of the DP model are specified as follows. $\xi = (33,2)$ and $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$; $\Sigma \sim IW(2, \tau)$ and $\tau = [\begin{matrix} 10 & 0 \\ 0 & 10 \end{matrix}]$; $\sigma^2 \sim \mathrm{IG}(0.01,0.01)$; $B_F \sim N(0,10)$; $B_R \sim G$ and $G \sim DP\big(a, N(\mu, \Sigma)\big)$ where $a = 1$, $B_R \sim \int N(\mu, \Sigma_\mathrm{b})(d\mu)$, and $\Sigma_\mathrm{b} \sim IW(2, T_b)$ where $T_b = [\begin{matrix} 10 & 0 \\ 0 & 10 \end{matrix}]$. Please see Section 3.2.3 for the detailed description and explanation of the prior specifications of the DP condition.

The results with respect to the parameter estimate and the percent bias of the parameter estimate for each model parameter are listed and compared with the population values in Table. In the DP conditions, the estimate of the parameter of the number of the latent classes is 1.84, which can be rounded to the integer 2. The class proportions are 51% in Class 1 and 49% in Class 2. Simply looking at the parameter estimates for the number of latent classes and the class proportions, we can see that the DP technique produced an estimate of the number of the classes that was approximately accurate. Although the parameter estimate was a decimal number, it was very close to the population value after being rounded to an integer. The class proportions (prop = 51%/49% for a 50%/50% condition) were also estimated accurately by DP. Based on the percent bias levels of the estimates, the means of the growth parameters in the DP condition were recovered well. Only the bias level of the mean in Class 2 exceeded 10%, which can be considered as "relatively high." However, the percent bias levels of the covariance matrix indicated that the variances and the covariance of the growth parameters were recovered poorly; they were all greater than 20%. The residual variance was recovered well.

Table 2. Parameter Estimates and Percent Bias for Proof of Concept Study

	Pop	DP	RJMCMC	% Bias of DP	% Bias of RJMCMC
# classes	2	1.840	2	-8%	0.000
Prop	0.5/0.5	0.51/0.51	0.51/0.51	0.02/-0.02	0.02/-0.02
Mean-I					
C1	31.370	31.430	31.410	0.191	0.128
C2	35.730	35.170	35.280	-1.567	-1.259
Mean-S					
C1	1.802	1.790	1.800	-0.666	-0.111
C2	3.350	3.010	3.110	-10.149	-7.164
Var-I	16.000	19.940	16.470	24.625	2.937
Var-S	2.000	2.440	2.190	22.000	9.500
Cov	0.300	1.360	0.270	353.333	-10.000
Residual	0.500	0.490	0.500	-2.000	0.000

In the RJMCMC condition, the number of latent classes and the class proportions were estimated very accurately. The percent bias levels of the parameter estimates indicated that all the parameters were recovered well, except the covariance with a bias level of 10%.

The results of the proof of concept simulation study suggested that the programs used to conduct the analysis for the DP and RJMCMC performed in an approximately accurate way. Although the covariance matrix was recovered poorly in the DP condition, this technique was still able to accurately estimate the number of latent classes, class proportions, mean parameters of the growth factors and the residual variance. The estimation issues for the covariance matrix did not only occur in the DP condition, they also appeared in other model estimation conditions. I will elaborate on these issues in the Results sections presented below.

When the DP and RJMCMC techniques were implemented in the main simulation studies via the R code, one issue needs to be noted. This proof of concept simulation study was conducted on a data set that contains a sample size of n = 10000, which is far greater than the actual sample size in each condition. Therefore, the accuracy of estimation might decrease to some extent as the sample size decreases. But the goal of this brief simulation study was to prove that these two techniques were feasible to implement with the currently developed programs using a large sample size; the results showed that the goal was achieved.

3.4 Results of Study 1

In Study 1, I examined the performance of ML, RJMCMC, and DP on recovering the number of latent classes, class proportions, and the means and covariance structure of the growth parameters. The final estimate of each parameter was calculated by averaging the estimates across all iterations. Specifically, the estimate for the number of classes was the average of all estimates from total number of post burn-in iterations; the estimates for the growth parameter means and variances were averaged across all iterations where the certain class solution was extracted. For example, the intercept means of a 2-class model were calculated by averaging the estimates of the intercept means in iterations that extracted 2 classes. The estimate of the number of classes was an integer within each iteration.

I evaluated the accuracy of the recovery of the parameter estimates by computing the relative percentage bias, where values greater than 10% were deemed problematic and are represented with bold values in the tables.[6] In addition to the parameter estimates and the percent bias levels, I also present the mean square error (MSE) for the means and covariance structure of the growth parameters as a composite measure of accuracy and efficiency. All of the simulation results in Study 1 are listed in Tables 3-8. Tables for MSE values (Tables A1 and A2) are presented in Appendix A.

3.4.1 Convergence

In Study 1, I used the Geweke (1992) statistic to assess whether the sampling procedure had converged to the target distribution within a Markov chain for the RJMCMC and DP conditions. The Geweke diagnostic evaluates the convergence for the Markov chains based on a test for equality of the means of the first 10% and the last 50% of a Markov chain (as to ensure that the two sections of the chain being compared are well separated, and presumably independent from one another). The test statistic of a Geweke diagnostic is a standard *z*-score, which is calculated by taking the difference between the two sample means and then dividing by its estimated standard error. If the samples are drawn from the stationary distribution of the chain, then the two means are equal and Geweke's statistic (i.e., the *z*-score) is asymptotically normally distributed. If the chain reaches convergence, then the *z*-score should have a relatively small value, for instance less than 2. The Geweke statistic for the means and covariance structure of the growth parameters was calculated using the "coda" package in R (Plummer, Best, Cowles, & Vines, 2006). The *z*-scores were all less 2.000 for all parameters in the RJMCMC and DP conditions in Study 1. I also included the trace plots for the estimates of the intercept and the slope parameter in the 2-class 80/20 condition as a proof of Markov chain convergence. Trace plots are presented in Figures 2 and 3.

[6] Relative percentage bias is computed using the following equation: $\frac{\text{Estimate}-\text{Population value}}{\text{Population value}} \times 100\%$

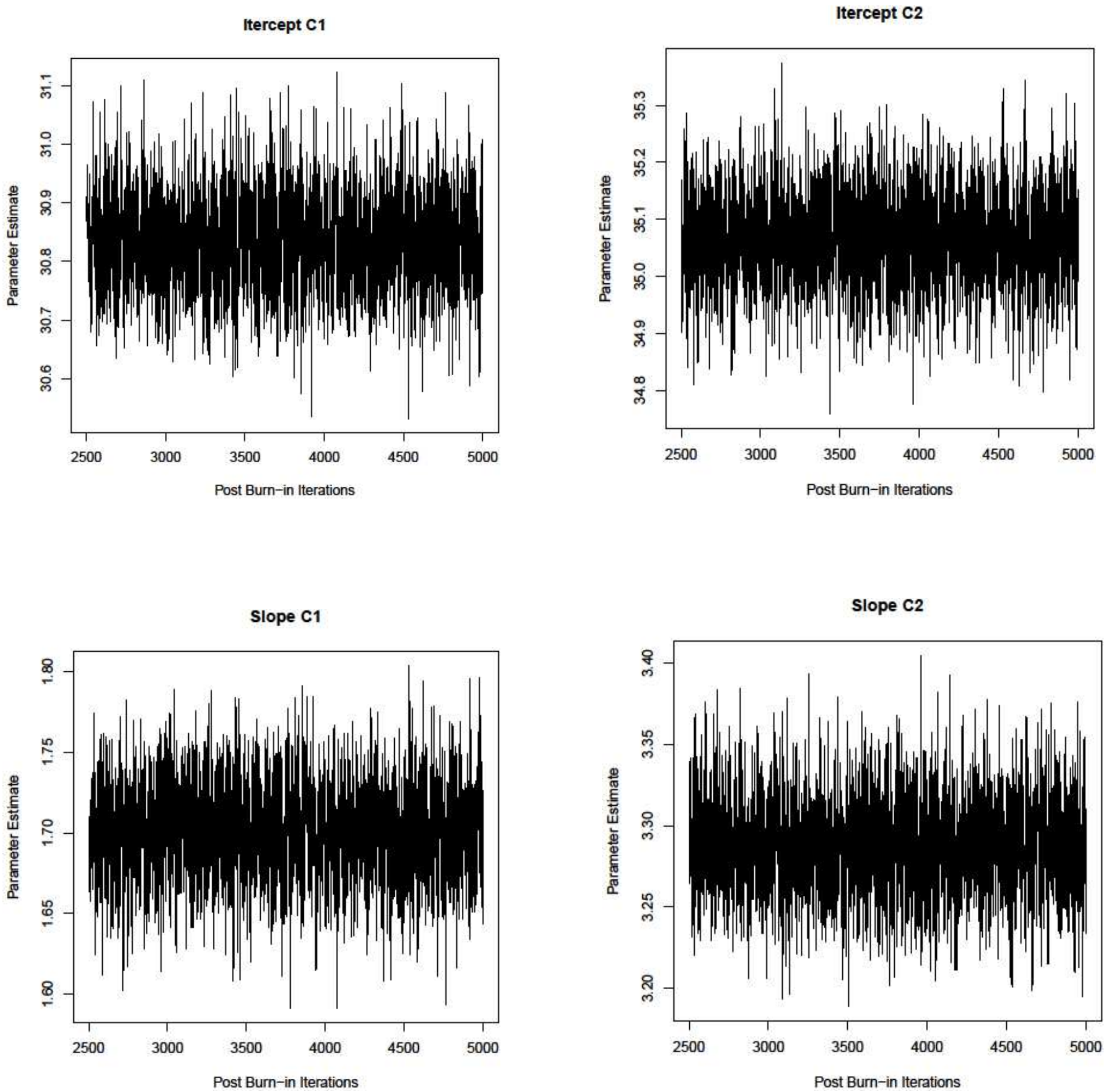

Figure 2. Trace Plots for the Growth Parameters in 2-Class 80/20 Condition for RJMCMC in Study 1. Top row: intercepts of Class 1 and Class 2 under RJMCMC; bottom row: slopes of Class 1 and Class 2 under RJMCMC.

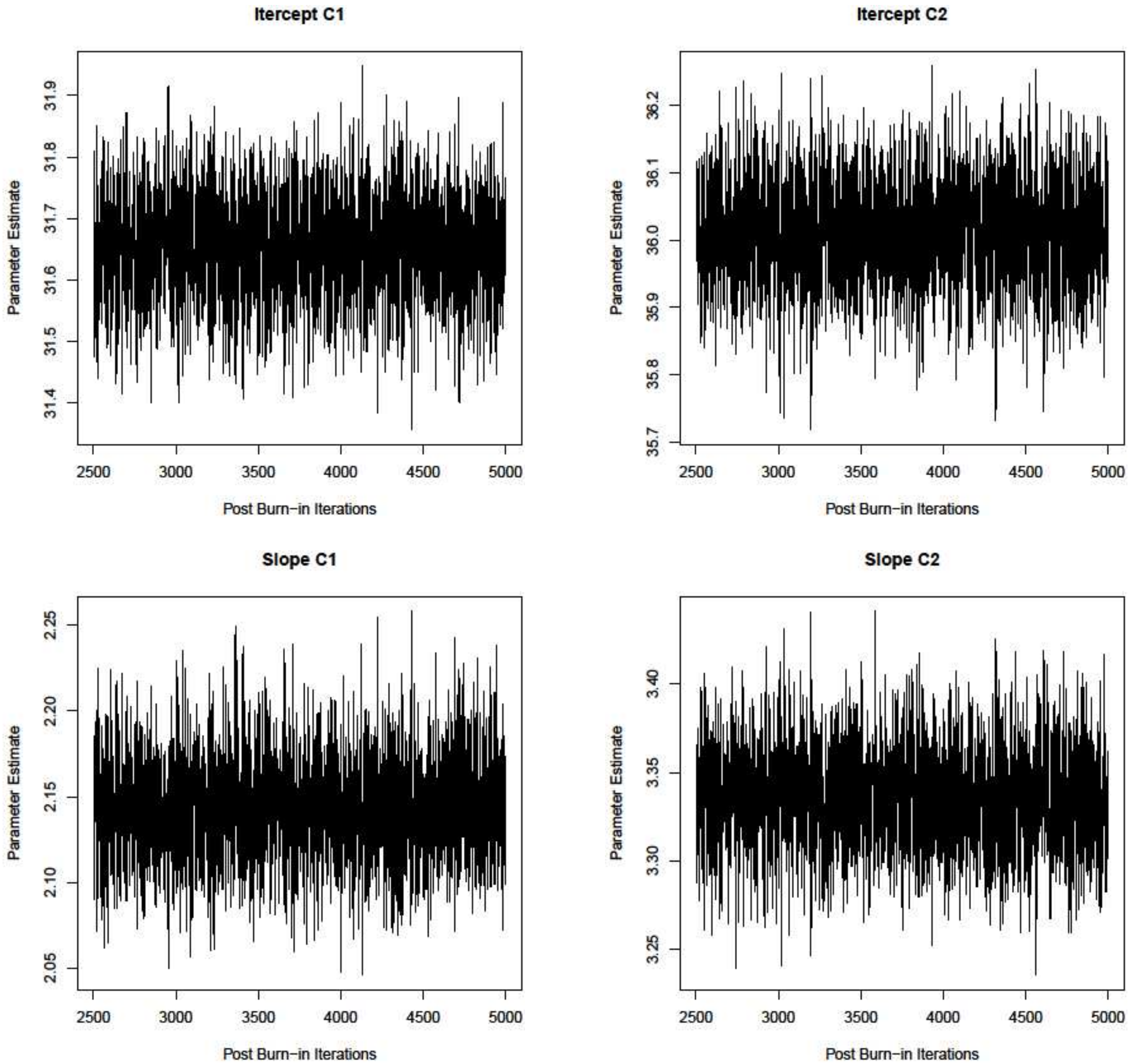

Figure 3. Trace Plots for the Growth Parameters in 2-Class 80/20 Condition for DP in Study 1. Top row: intercepts of Class 1 and Class 2 under DP; bottom row: slopes of Class 1 and Class 2 under DP.

In addition to the chain convergence for all of the parameters in RJMCMC and DP, model convergence was also obtained for 2000 out of 2000 replications being requested in each simulation cell. All conditions for the three model estimation techniques had a 100% model convergence rate. All results that are presented here were calculated using the parameter estimates from the 2000 converged replications.

3.4.2 Class Enumeration and Parameter Estimate

3.4.2.1 GCM

Results of the parameter estimates and the percent bias values for the GCM with ML, RJMCMC, and DP in Study 1 are presented in Tables 3 and the top section of Table A1. Since the number of latent classes was predetermined by the ML estimation method, the latent class parameter was only estimated by RJMCMC and DP. In the DP conditions, the number of latent classes was estimated accurately with decimal numbers that are very close to 1 (i.e., indicating only one class emerged). DP yielded quit low bias levels for this parameter. In addition to the number of classes, RJMCMC also provided the percentage of the number of iterations that picked this specific number of class in each replication,[7] presented in the parentheses in Tables 3-38 of the parameter estimates. This feature of RJMCMC can be greatly useful as it provides us with the certainty of the final class solution, as well as the number of classes extracted. For instance, in the RJMCMC n = 200 condition, the estimate for the number of classes was 1.002, indicating that the percentage of iterations within the RJMCMC algorithm selecting a 1-class solution was 99.8%. The high percent of picking a specific number of class solutions could represent a high confidence of our determination of the final class enumeration. Another way of phrasing this could be that if this percentage of selecting a particular class solution was markedly lower, then our confidence in the final class solution would also be diminished.

The growth parameter means were well recovered by ML, RJMCMC, and DP. The percent bias levels were all below 10% for all three sample sizes. The covariance structure was recovered very well in the ML conditions with low bias levels. However, both RJMCMC and DP produced moderate-to-high bias levels for the covariance structure. The MSE values are quite small, some close to 0, for the means and the covariance structure parameters in GCM with ML, RJMCMC, and DP.

[7] This percentage is calculated in 3 steps: 1) counting the number of iterations that selected a certain of number of class, 2) dividing the number by the total number of iterations in each replicate, and 3) averaging the percentages across all 2000 replications.

Table 3. Parameter Estimates and Percent Bias for GCM

	Pop	n=200 ML	n=200 RJMCMC	n=200 DP	n=400 ML	n=400 RJMCMC	n=400 DP	n=600 ML	n=600 RJMCMC	n=600 DP
		Parameter Estimates								
# classes	1.000	1.000	1.002(0.998)	1.029	1.000	1.002(0.998)	1.005	1.000	1.002(0.998)	1.004
Mean-I	31.370	31.373	31.929	31.473	31.364	31.527	31.112	31.369	31.559	31.339
Mean-S	1.802	1.801	1.964	1.741	1.801	1.789	1.744	1.801	1.799	1.920
Var-I	16.000	15.910	15.603	14.206	15.936	15.874	14.448	15.933	16.248	17.249
Var-S	2.000	1.991	2.654	1.730	2.000	2.363	1.964	1.999	2.187	1.990
Cov	0.300	0.305	0.128	0.082	0.302	0.375	0.804	0.306	0.340	0.146
Residual	0.500	0.504	0.516	0.755	0.502	0.496	0.686	0.502	0.499	0.697
		Percent Bias								
Mean-I	31.370	0.009	1.783	0.329	-0.019	0.500	-0.821	-0.004	0.604	-0.100
Mean-S	1.802	-0.083	8.977	-3.412	-0.044	-0.694	-3.192	-0.067	-0.153	6.550
Var-I	16.000	-0.564	-2.482	**-11.215**	-0.399	-0.789	-9.700	-0.421	1.548	7.808
Var-S	2.000	-0.475	**32.720**	**-13.480**	0.020	**18.135**	-1.797	-0.045	9.360	-0.501
Cov	0.300	1.667	**-57.369**	**-72.771**	0.733	**24.922**	**168.007**	2.133	**13.274**	**-51.482**
Residual	0.500	0.740	3.165	**51.085**	0.420	-0.820	**37.201**	0.360	-0.275	**39.316**

3.4.2.2 2-Class Prop=50/50 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 50/50 conditions are presented in Table 4 and the middle section of Table A1. RJMCMC and DP yielded accurate estimates of the number of classes. In the RJMCMC conditions, the percent of picking a 2-class solution out of all 2000 iterations improved as the sample size increased (e.g., from 65.1% for n = 200 to 94.2% for n = 600). This trend indicates that we can be more certain that the final class solution estimated by RJMCMC is 2 when the sample size is larger compared to when it is relatively smaller. In other words, we would expect proper class enumeration to be more difficult under smaller sample sizes, and the RJMCMC algorithm provides an indication that this is indeed true. The class proportions parameter was recovered well for all three estimation methods across different sample sizes. Only the RJMCMC n = 200 condition produced bias levels that were slightly higher than 10% for the 2-class proportion parameters.

The means of the intercept parameters were estimated with low bias levels for ML, RJMCMC, and DP, while the means of the slopes had relatively high bias levels for ML and DP in some conditions. All three estimation methods yielded moderate-to-high bias levels for the covariance structure. The MSE values for all the parameters were relatively low (below 2) across estimation methods and sample sizes. The few exceptions were the variance of the intercept parameter for DP, which had relatively larger MSE values (i.e., above 5).

Table 4. Parameter Estimates and Percent Bias for 2-Class 50/50

			n=200			n=400			n=600		
			ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
		Pop	Parameter Estimates								
# classes		2.000	2.000	2.331(0.651)	1.938	2.000	2.132(0.852)	1.975	2.000	2.218(0.942)	1.880
Prop	C1		0.497	0.448	0.527	0.510	0.514	0.512	0.507	0.503	0.512
	C2		0.503	0.552	0.474	0.490	0.486	0.488	0.493	0.497	0.488
Mean-I											
C1		31.370	31.103	32.113	31.953	31.252	31.008	31.018	31.313	31.747	31.140
C2		35.730	36.033	36.894	34.486	35.979	35.596	35.581	35.893	34.493	36.767
Mean-S											
C1		1.802	2.272	1.705	1.994	2.235	1.562	1.642	2.104	2.293	1.689
C2		3.350	2.848	3.050	3.396	2.910	3.060	3.087	3.010	3.457	3.277
Var-I		16.000	15.946	14.821	19.276	16.329	16.780	20.721	16.299	16.160	20.249
Var-S		2.000	1.968	2.061	2.918	2.033	1.865	2.789	2.030	2.118	2.613
Cov		0.300	1.182	0.829	1.510	1.073	0.137	1.459	0.917	0.644	1.240
Residual		0.500	0.503	0.712	0.519	0.502	0.706	0.506	0.502	0.700	0.513
			Percent Bias								
Prop	C1		-0.616	**-10.333**	5.300	1.944	2.833	2.450	1.454	0.556	2.433
	C2		0.616	**10.333**	-5.300	-1.944	-2.833	-2.450	-1.454	-0.556	-2.433
Mean-I											
C1		31.370	-0.850	2.370	1.860	-0.376	-1.155	-1.122	-0.180	1.201	-0.733
C2		35.730	0.848	3.257	-3.482	0.697	-0.376	-0.416	0.457	-3.461	2.901
Mean-S											
C1		1.802	**26.099**	-5.397	**10.670**	**24.007**	**-13.315**	-8.900	**16.781**	**27.242**	-6.248
C2		3.350	**-14.982**	-8.960	1.365	**-13.137**	-8.656	-7.855	**-10.152**	3.197	-2.174
Var-I		16.000	-0.338	-7.368	**20.475**	2.054	4.876	**29.505**	1.869	1.001	**26.558**
Var-S		2.000	-1.625	3.053	**45.890**	1.660	-6.750	**39.450**	1.475	5.907	**30.669**
Cov		0.300	**293.967**	**176.312**	**403.395**	**257.800**	**-54.208**	**386.225**	**205.533**	**114.581**	**313.405**
Residual		0.500	0.560	**42.403**	3.797	0.320	**41.266**	1.264	0.320	**39.946**	2.637

3.4.2.3 2-Class Prop=80/20 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 80/20 conditions are presented in Table 5 and the bottom section of Table A1. The estimate of the number of classes in the RJMCMC and DP conditions was 2, which was accurately recovered. In the RJMCMC conditions, the percentage of a 2-class solution was higher when the sample size was 600 compared to n = 200 or 400. ML, RJMCMC, and DP recovered the class proportions poorly when n = 200, especially for the minority class (i.e., prop = 20%). While ML and DP tended to overestimate the minority class, RJMCMC tended to underestimate it. The recovery of the class proportions improved as the sample increased; overall, RJMCMC and DP were slightly better than ML.

The means of the intercepts were recovered well for all three estimation methods. ML yielded slightly higher bias (higher than 10%) for the slope parameter means when n = 200 and n = 400. RJMCMC and DP yielded lower bias levels (lower than 10%), except for RJMCMC when n = 600, where the bias level was around 35%. The covariance structure parameters were estimated poorly, with high bias levels (all above 90%) for all three estimation methods across all simple sizes. Only ML yielded lower bias levels for the covariance structure as the sample size became larger. All three estimation methods produced low MSE values (below 2) for all parameters, while DP still had higher MSE values for the variance of the intercept parameter (around 5-23).

Table 5. Parameter Estimates and Percent Bias for 2-Class 80/20

			n=200			n=400			n=600		
			ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
		Pop	Parameter Estimates								
# classes		2.000	2.000	2.335(0.766)	1.720	2.000	2.125(0.742)	1.819	2.000	2.126(0.863)	2.186
Prop	C1		0.644	0.838	0.702	0.732	0.817	0.785	0.771	0.813	0.825
	C2		0.356	0.162	0.299	0.268	0.183	0.216	0.229	0.187	0.175
Mean-I											
C1		31.370	30.587	30.832	31.653	30.876	31.141	31.147	31.063	31.780	32.469
C2		35.730	35.646	35.060	36.012	35.836	35.352	35.335	35.853	35.835	35.064
Mean-S											
C1		1.802	2.088	1.702	2.145	1.995	1.610	1.651	1.930	2.434	1.963
C2		3.350	2.778	3.287	3.336	3.007	3.044	3.222	3.092	3.425	3.652
Var-I		16.000	15.124	13.751	20.805	15.618	16.798	18.546	15.702	16.252	19.561
Var-S		2.000	1.870	2.111	2.990	1.943	1.915	2.775	1.959	2.157	2.495
Cov		0.300	0.818	0.011	1.405	0.676	0.025	1.800	0.593	0.627	1.527
Residual		0.500	0.503	0.737	0.513	0.502	0.730	0.491	0.502	0.718	0.503
			Percent Bias								
Prop	C1		**-19.500**	4.792	**-12.313**	-8.555	2.083	-1.938	-3.673	1.583	3.125
	C2		**78.000**	**-19.167**	**49.250**	**34.220**	-8.333	7.750	**14.690**	-6.333	**-12.500**
Mean-I											
C1		31.370	-2.496	-1.715	0.902	-1.574	-0.729	-0.710	-0.980	1.307	3.504
C2		35.730	-0.236	-1.875	0.789	0.298	-1.059	-1.105	0.344	0.295	-1.863
Mean-S											
C1		1.802	**15.871**	-5.549	**19.018**	**10.699**	**-10.631**	-8.405	7.120	**35.053**	8.958
C2		3.350	**-17.063**	-1.881	-0.407	**-10.242**	-9.146	-3.808	-7.690	2.232	9.022
Var-I		16.000	-5.474	**-14.058**	**30.028**	-2.391	4.988	**15.910**	-1.865	1.576	**22.256**
Var-S		2.000	-6.505	5.527	**49.504**	-2.840	-4.256	**38.727**	-2.060	7.859	**24.762**
Cov		0.300	**172.767**	**-96.380**	**368.208**	**125.300**	**-91.746**	**500.087**	**97.500**	**108.984**	**408.916**
Residual		0.500	0.680	**47.316**	2.646	0.380	**45.970**	-1.812	0.360	**43.622**	0.693

3.4.2.4 3-Class Prop=33/33/33 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 33/33/33 conditions are presented in Table 6 and the top section of Table A2. RJMCMC and DP estimated the number of classes accurately. RJMCMC yielded a higher percentage (above 16%) of a 3-class solution as the sample size increased from n = 200 to n = 400. ML poorly recovered the class proportions across sample sizes, bias levels all above 17%. The class proportion estimates improved for RJMCMC as sample sizes increased from n = 400 to n = 600. DP performed comparatively better, with lower bias levels than ML and RJMCMC, in recovering the class proportions; this was especially the case when sample sizes were larger.

The estimates for the growth parameter means had low bias levels in all three estimation methods. The growth parameter variances were recovered well by ML but had relatively higher bias levels (about 43%) for RJMCMC and DP when the sample sizes were n = 200 and n = 400. The covariance structure was estimated poorly in all estimation methods (i.e., bias levels were extremely high, all were above -34%). The MSE values for all parameters were relatively low (all below 3) except for the variance of the intercept in the DP conditions (which were around 7 to 67).

Table 6. Parameter Estimates and Percent Bias for 3-Class 33/33/33

			n=200			n=400			n=600		
			ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
		Pop	Parameter Estimates								
# classes		3.000	3.000	2.626(0.652)	2.607	3.000	2.652(0.671)	2.834	3.000	2.882(0.753)	2.847
	C1		0.300	0.278	0.323	0.307	0.215	0.317	0.305	0.304	0.332
Prop	C2		0.428	0.309	0.374	0.425	0.400	0.329	0.422	0.371	0.342
	C3		0.271	0.413	0.304	0.269	0.385	0.354	0.273	0.325	0.326
Mean-I C1		31.370	31.018	31.010	30.624	30.977	30.557	31.063	30.979	31.781	30.747
Mean-I C2		35.730	35.860	34.072	34.397	35.893	34.669	34.974	35.861	36.833	34.859
Mean-I C3		40.090	40.909	41.998	41.844	40.961	39.199	39.190	40.947	39.513	38.991
Mean-S C1		1.802	2.006	1.724	1.719	1.871	1.671	1.674	1.802	1.951	1.958
Mean-S C2		3.350	3.484	3.311	3.132	3.511	3.395	3.393	3.510	3.577	3.568
Mean-S C3		4.880	4.635	4.232	4.347	4.730	4.815	4.813	4.772	4.931	4.774
Var-I		16.000	15.587	14.769	24.203	15.688	18.764	23.031	15.738	14.937	16.987
Var-S		2.000	1.837	2.059	3.390	1.901	1.867	3.128	1.931	2.234	2.533
Cov		0.300	1.055	0.953	3.327	0.694	0.198	2.729	0.567	0.514	1.159
Residual		0.500	0.503	0.739	0.513	0.502	0.705	0.502	0.502	0.696	0.490
			Percent Bias								
	C1		-9.024	**-15.657**	-2.273	-7.048	**-34.848**	-3.811	-7.430	-7.912	0.556
Prop	C2		**29.785**	-6.313	**13.182**	**28.642**	**21.212**	-0.432	**27.824**	**12.458**	3.586
	C3		**-17.730**	**25.000**	-7.879	**-18.564**	**16.667**	7.273	**-17.364**	-1.515	-1.111
Mean-I C1		31.370	-1.121	-1.149	-2.377	-1.253	-2.592	-0.979	-1.247	1.311	-1.985
Mean-I C2		35.730	0.365	-4.640	-3.730	0.455	-2.968	-2.115	0.366	3.086	-2.438
Mean-I C3		40.090	2.043	4.759	4.375	2.172	-2.221	-2.245	2.138	-1.440	-2.740
Mean-S C1		1.802	**11.293**	-4.323	-4.610	3.812	-7.282	-7.121	0.011	8.283	8.658
Mean-S C2		3.350	3.985	-1.168	-6.505	4.791	1.343	1.274	4.773	6.782	6.508
Mean-S C3		4.880	-5.029	**-13.270**	**-10.918**	-3.076	-1.339	-1.379	-2.219	1.053	-2.170
Var-I		16.000	-2.581	-7.692	**51.270**	-1.951	**17.275**	**43.943**	-1.639	-6.645	6.170
Var-S		2.000	-8.135	2.932	**69.482**	-4.940	-6.665	**56.377**	-3.440	**11.699**	**26.672**
Cov		0.300	**251.600**	**217.611**	**1009.021**	**131.200**	**-34.109**	**809.720**	**88.933**	**71.444**	**286.482**
Residual		0.500	0.500	**47.723**	2.517	0.420	**40.999**	0.348	0.320	**39.114**	-2.001

3.4.2.5 3-Class Prop=45/45/10 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 45/45/10 conditions are presented in Table 7 and the middle section of Table A2. The number of classes was estimated accurately in RJMCMC and DP. ML consistently overestimated the minority class for the class proportion parameters with moderate-to-high bias levels (around 16%-80%) across sample sizes. While RJMCMC also slightly overestimated the minority class for n = 200 and n = 400 conditions (e.g., bias levels were around 11% to 17%), it estimated the class proportions accurately when the sample size was 600 (bias levels decreased to below 10%). DP consistently performed the best among all three estimation methods, with low bias levels (all levels were under 10% bias) for the class proportions.

The intercept growth parameter means were estimated well in all three estimation methods across sample sizes (bias levels all under 10%). The slopes growth parameter means had slightly higher bias levels (e.g., slightly higher than 10%) when n = 200, but parameter recovery was improved as the sample size increased. The growth parameter variances was estimated with relatively high bias levels for all estimation methods when n = 200 (e.g., bias levels were slightly higher than 10% under ML and RJMCMC and were over 66% under DP); the bias of the variances decreased under ML and RJMCMC but maintained relatively high levels under DP (e.g., above 20%), even as the sample size increased to n = 200. The covariance parameter bias levels were consistently high (all levels were above -72%) for all estimation methods across sample sizes. The MSE values were relatively smaller for most parameters (e.g., MSE < 3.2) except for the intercept variance, which had relatively high MSE values (higher than 10) in the DP condition.

Table 7. Parameter Estimates and Percent Bias for 3-Class 45/45/10

			n=200			n=400			n=600		
			ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
		Pop				Parameter Estimates					
# classes		3.000	3.000	3.2(0.633)	3.229	3.000	3.25(0.767)	3.127	3.000	3.047(0.949)	3.041
	C1		0.343	0.406	0.464	0.370	0.438	0.459	0.377	0.480	0.425
Prop	C2		0.477	0.477	0.445	0.475	0.451	0.437	0.479	0.422	0.469
	C3		0.180	0.117	0.091	0.155	0.111	0.105	0.143	0.098	0.106
Mean-I C1		31.370	30.779	30.245	29.484	30.946	30.208	30.227	31.044	31.869	32.068
Mean-I C2		35.730	34.816	36.890	38.778	35.026	36.541	36.615	35.126	34.828	36.340
Mean-I C3		40.090	39.747	38.337	39.937	39.971	38.880	38.301	40.043	40.422	40.290
Mean-S C1		1.802	2.033	1.666	1.756	1.983	1.585	1.622	1.955	1.973	1.978
Mean-S C2		3.350	2.990	3.081	3.772	3.082	3.024	3.090	3.123	3.445	3.439
Mean-S C3		4.880	4.230	4.321	4.324	4.373	4.605	4.464	4.459	4.836	4.594
Var-I		16.000	15.293	14.235	26.612	15.841	17.170	26.844	15.918	16.160	19.229
Var-S		2.000	1.726	2.048	3.610	1.850	1.943	3.710	1.899	2.164	2.740
Cov		0.300	1.160	0.843	4.139	1.026	0.081	3.789	0.909	0.691	1.461
Residual		0.500	0.504	0.739	0.500	0.503	0.721	0.515	0.503	0.698	0.504
						Percent Bias					
	C1		**-23.844**	-9.778	3.133	**-17.724**	-2.593	1.889	**-16.120**	6.667	-5.481
Prop	C2		6.002	5.926	-1.222	5.593	0.122	-2.922	6.542	-6.296	4.200
	C3		**80.290**	**17.333**	-8.600	**54.590**	**11.119**	4.650	**43.100**	-1.667	5.767
Mean-I C1		31.370	-1.885	-3.588	-6.013	-1.351	-3.705	-3.642	-1.039	1.589	2.225
Mean-I C2		35.730	-2.557	3.246	8.531	-1.969	2.271	2.478	-1.690	-2.524	1.706
Mean-I C3		40.090	-0.855	-4.372	-0.381	-0.297	-3.017	-4.463	-0.116	0.829	0.499
Mean-S C1		1.802	**12.819**	-7.549	-2.559	**10.044**	**-12.017**	-9.963	8.474	9.503	9.741
Mean-S C2		3.350	**-10.740**	-8.043	**12.601**	-8.006	-9.738	-7.756	-6.767	2.837	2.657
Mean-S C3		4.880	**-13.316**	**-11.458**	**-11.391**	**-10.391**	-5.627	-8.528	-8.627	-0.893	-5.865
Var-I		16.000	-4.421	**-11.034**	**66.325**	-0.993	7.314	**67.775**	-0.513	1.002	**20.182**
Var-S		2.000	**-13.690**	2.405	**80.512**	-7.500	-2.851	**85.481**	-5.075	8.183	**37.020**
Cov		0.300	**286.500**	**181.135**	**1279.761**	**242.133**	**-72.961**	**1163.161**	**202.967**	**130.226**	**386.905**
Residual		0.500	0.760	**47.861**	0.061	0.500	**44.170**	3.083	0.520	**39.699**	0.890

3.4.2.6 3-Class Prop=70/20/10 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 70/20/10 conditions are presented in Table 8 and the bottom section of Table A2. The number of classes was estimated accurately by RJMCMC and DP. The percentage of selecting a 3-class solution yielded by RJMCMC was relatively low in the n = 200 condition (i.e., 55.4%) and the n = 400 condition (i.e., 58%), compared to the percentage in the n = 600 condition (i.e., 84.4%). This pattern suggests that we have greater certainty in the number of classes as sample sizes are increased. The class proportions were recovered with relatively high bias levels by ML and RJMCMC when n = 200 (bias levels were around 15% to 103% under ML and around 11% to -32% under RJMCMC). Class proportion recovery under ML and RJMCMC was improved slightly as the sample size increased to n = 200 (e.g., bias levels decreased to as low as -23% under ML and to slightly higher than 10% under RJMCMC). DP performed well across sample sizes in estimating the class proportions with relatively low bias levels (bias levels never exceeded 19%). The class size in the majority class was consistently underestimated by ML (e.g., it was -31.763% for the n = 200 condition) and was slightly over estimated by RJMCMC (e.g., it was 11.143% for the n = 200 condition).

The growth parameter means were estimated with relatively low bias levels (most did not exceed 13.5% bias) except for a few slope mean parameters in ML and RJMCMC under n = 200 (these bias levels were all over -16%). ML recovered the covariance structure poorly, but bias levels decreased as the sample size increased. The performance of RJMCMC and DP was inconsistent across sample sizes in estimating the covariance structure. The MSE values are small (e.g., below 2) for most parameters in all three estimation methods conditions, but they were relatively high (e.g., some of the MSE values were higher than 64) for the intercept variance under RJMCMC and DP.

Table 8. Parameter Estimates and Percent Bias for 3-Class70/20/10

			n=200			n=400			n=600		
		Pop	ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
			Parameter Estimates								
# classes		3.000	3.000	3.259(0.554)	3.489	3.000	3.268(0.58)	3.272	3.000	2.845(0.844)	3.199
	C1		0.478	0.778	0.653	0.519	0.757	0.678	0.537	0.746	0.662
Prop	C2		0.406	0.136	0.237	0.379	0.162	0.233	0.366	0.172	0.233
	C3		0.116	0.086	0.110	0.102	0.081	0.089	0.097	0.082	0.106
Mean-I C1		31.370	30.386	29.909	30.140	30.564	30.757	30.761	30.660	31.726	29.348
Mean-I C2		35.730	34.545	36.660	38.471	34.798	35.983	36.763	34.913	35.241	33.724
Mean-I C3		40.090	40.554	38.748	38.923	40.690	38.980	38.382	40.756	38.650	38.667
Mean-S C1		1.802	1.706	1.629	1.760	1.693	1.733	1.733	1.760	1.860	1.861
Mean-S C2		3.350	2.808	2.664	3.535	2.984	3.049	3.121	3.079	3.386	3.387
Mean-S C3		4.880	4.743	4.241	4.517	4.775	4.590	4.656	4.769	5.450	4.463
Var-I		16.000	14.236	14.246	27.144	14.837	16.986	25.101	15.202	24.003	15.447
Var-S		2.000	1.661	2.086	3.684	1.775	1.971	3.298	1.832	2.015	3.250
Cov		0.300	0.375	0.065	4.046	0.309	0.034	2.848	0.303	0.405	2.903
Residual		0.500	0.503	0.715	0.493	0.502	0.727	0.508	0.500	0.710	0.486
			Percent Bias								
	C1		**-31.763**	**11.143**	-6.667	**-25.820**	8.086	-3.186	**-23.274**	6.571	-5.500
Prop	C2		**103.180**	**-32.000**	**18.500**	**89.335**	**-18.900**	**16.650**	**83.130**	**-13.778**	**16.333**
	C3		**15.980**	**-14.000**	9.670	2.070	**-18.800**	**-11.000**	-3.340	**-18.444**	5.833
Mean-I C1		31.370	-3.137	-4.658	-3.922	-2.570	-1.955	-1.940	-2.264	1.133	-6.444
Mean-I C2		35.730	-3.316	2.604	7.671	-2.608	0.708	2.890	-2.287	-1.370	-5.616
Mean-I C3		40.090	1.158	-3.348	-2.912	1.497	-2.768	-4.261	1.662	-3.592	-3.549
Mean-S C1		1.802	-5.311	-9.618	-2.341	-6.060	-3.848	-3.812	-2.331	3.220	3.299
Mean-S C2		3.350	**-16.179**	**-20.490**	5.534	**-10.919**	-8.989	-6.822	-8.078	1.062	1.117
Mean-S C3		4.880	-2.816	**-13.097**	-7.430	-2.156	-5.945	-4.593	-2.285	**11.680**	-8.535
Var-I		16.000	**-11.024**	**-10.963**	**69.653**	-7.269	6.160	**56.881**	-4.988	**50.020**	-3.457
Var-S		2.000	**-16.955**	4.300	**84.221**	**-11.245**	-1.472	**64.901**	-8.400	0.744	**62.494**
Cov		0.300	**24.967**	**-78.337**	**1248.527**	3.100	**-88.748**	**849.360**	1.067	**34.891**	**867.792**
Residual		0.500	0.640	**42.998**	-1.380	0.480	**45.433**	1.555	-0.020	**41.958**	-2.774

3.5 An Empirical Example

This section provides a simple example as a tutorial to illustrate the application of the Bayesian non/semi-parametric methods in a substantive research area. I applied RJMCMC and DP on an LGMM using a dataset from the Early Childhood Longitudinal Study, Kindergarten Class program (ECLS-K; [NCES], 2001). This dataset includes a sample from the ECLS-K program that focuses on children's early school experiences beginning with kindergarten and following children through middle school. It consisted of approximately 1000 schools in the U.S., with a series of longitudinal measurements of students, teachers, and schools.

The subset of the ECLS-K data I used in this example included *n* = 400 cases that were randomly drawn from the original dataset. This subset consisted of four waves (i.e., time points) of measures: fall-kindergarten, spring-kindergarten, fall-first grade, and spring-third grade. The base year was Fall, 1998 and the spacing of the four waves was handled in the code to represent the unequal time spacing when data were collected (see for more details on time spacing: Kaplan, 2002). The outcome measure of this study was reading assessment (i.e., children's item response theory (IRT) scores on reading). The reading IRT scores in the ECLS-K data set were assumed normally distributed with means (and standard deviations in the parentheses) for the four waves: 22.67 (8.58), 32.47 (10.85), 37.97 (12.67), and 54.77 (14.17) (ECLS-K; [NCES], 2001).

The model used to analyze the data was an LGMM of 4 time points with an unknown number of latent classes to be tested. The estimation methods for data analysis were RJMCMC and DP. In the RJMCMC condition, the analytic model is specified in Equations 36-38. The priors in the RJMCMC condition were specified in a way such that very little information was incorporated so as to reflect a scenario where the researcher knows very little about the population. The prior specifications are as follows: $\xi = (0,0)$ and $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$; $\Sigma \sim IW(2, \tau)$ and $\tau = [\begin{smallmatrix} 10 & 0 \\ 0 & 10 \end{smallmatrix}]$; $\sigma^2 \sim IG(0.01{,}0.01)$; $B_F \sim N(0{,}10)$; and $(\pi_1 \ldots{,}\pi_C) \sim Dirichlet(10, \ldots, 10)$. A detailed description and explanation of the prior specification of RJMCMC are included in the Methods section in Study 1 (Section 3.2.3). One Markov chain with 5000 iterations was requested, of which the first 2500 iterations were discarded as burn-in. In the DP condition, the data were analyzed with the model specified in Equations 59-65. Akin to the RJMCMC condition, the non-informative priors were also implemented in the DP condition: $\xi = (0,0)$ and $D = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$; $\Sigma \sim IW(2, \tau)$ and $\tau = [\begin{smallmatrix} 10 & 0 \\ 0 & 10 \end{smallmatrix}]$; $\sigma^2 \sim IG(0.01{,}0.01)$; $B_F \sim N(0{,}10)$; $B_R \sim G$ and $G \sim DP\big(a, N(\mu, \Sigma)\big)$ where $a = 1$, $B_R \sim \int N(\mu, \Sigma_{\mathrm{b}})(d\mu)$, and $\Sigma_b \sim IW(2, T_b)$ where $T_b = [\begin{smallmatrix} 10 & 0 \\ 0 & 10 \end{smallmatrix}]$. A detailed description and explanation of the prior specifications of DP can be found in the Methods section in Study 1 (Section 3.2.3).

Results of the parameters estimates (mean and standard deviation) of the LGMM for RJMCMC and DP are presented in Table 9. RJMCMC and DP yielded conflicting outcomes. In the RJMCMC condition, the estimate of the number of latent classes was 1.991 according to the algorithm and around 89.1% out of all the iterations selected the 2 class solution. This indicates that there was strong support, according to RJMCMC, that the number of classes was 2. The class proportions estimated by RJMCMC were 13.7% in Class 1 and 86.3% in Class2. This suggests that majority of the children fell within one group in terms of the growth of the reading ability, while a few children were in another group. The intercept mean was 23.750 for Class 1 and 37.199 for Class 2, indicating the reading ability started at different points for the two latent classes. The slope mean for Class 1 was 4.308 and the slope for Class 2 was 4.753, which suggests that the growth rates of the reading ability for the two latent classes were close. The covariance structure was constrained equal across the two classes for ease of estimation for this example. The intercept parameter variance was 18.289, indicating a relatively large variation around the starting point. The slope parameter variance was 0.612 which suggests a relatively small variation of the growth rate. Overall, the separation between the classes appeared to be mostly at the fall-kindergarten reading level (i.e., through the intercept).

In the DP condition, the estimate of the number of classes was 1.443, which can be rounded to an integer of 1. This indicates that only 1 class was selected by DP estimation algorithm. Within this one class, the mean of the intercept was 29.961 and the mean of the slope was 3.940. The intercept parameter variance was close to the estimate in the RJMCMC condition, suggesting a large variation around the starting point. The estimate of the slope parameter variance, 13.832, indicates a relatively large variation in the growth rate among the individuals.

Overall, this example highlights how the results from RJMCMC and DP can be interpreted. It may be, as was the case here, that results do not align across estimation methods. This finding is perfectly fine, and the researcher would need to use substantive context to aid in selecting the final model solution to interpret. The element that RJMCMC and DP have that is lacking in the traditional approach to model testing (i.e., comparing competing class solutions via information criteria and other fit measures) is that these approaches actually *estimate* the number of latent classes. Then the approaches provide an index that allows the research to establish a level of (un)certainty in the final model. This level of (un)certainty can be defined through the raw estimate of the number of classes (i.e., RJMCMC = 1.991, DP = 1.443), as well as the number of iterations that favored a particular class solution. These features are potentially more informative and interpretable compared to the traditional approach to model selection.

Table 9. Parameter Estimates of the LGMM on the Reading IRT Scores

Parameter	RJMCMC		DP	
	Mean	SD	Mean	SD
# classes	1.991(89.100)	-	1.443	-
Prop	0.137	-	-	-
	0.863	-	-	-
Mean-I				
C1	23.750	5.538	29.961	1.072
C2	37.199	10.753	-	-
Mean-S				
C1	4.308	0.683	3.940	1.497
C2	4.753	0.606	-	-
Var-I	18.289	6.441	16.589	22.170
Var-S	0.612	0.306	13.832	3.489
Cov	6.446	1.828	0.915	0.467
Residual	59.596	14.772	43.923	8.889

3.6 Study 1 Discussion

In Study 1, the performance of ML, RJMCMC, and DP was examined in terms of the recovery of number of latent classes, class proportions, and the growth parameter means and the covariance structure. Then an empirical example was presented to illustrate the application of RJMCMC and DP using a substantive dataset. The findings of Study 1 are discussed as follows.

3.6.1 Estimation Methods

Overall, RJMCMC and DP performed as well as, or better than, compared to ML for the recovery of some model parameters. First, RJMCMC and DP were able to provide the correct estimates of the number of latent classes. RJMCMC and DP extracted the number of classes through their internal algorithm instead of comparing multiple models based on statistical and fit measures that are usually employed by ML. Therefore, RJMCMC and DP are more efficient than ML.

Second, the useful feature that is specifically linked to RJMCMC is able to provide the percentage of selecting a particular class solution through the algorithm. This percentage may be used as the likelihood or degree of (un)certainty about this particular class solution. It can provide the researcher with some form of guideline when making the decision of the number of classes. For instance, the researcher would be more certain or confident about her decision for a 2-class solution if 95% of the RJMCMC iterations selected the 2-class solution compared to if only 75% selected the 2-class solution.

Third, when estimating the class proportions, RJMCMC and DP outperformed ML, which tended to underestimate the class size of the minority class. DP performed particularly better than ML and RJMCMC on recovering class proportions, regardless of whether a majority or a minority class existed.

Fourth, RJMCMC and DP outperformed ML in recovering the growth parameter means, especially for the 2-class models. However, RJMCMC and DP performed worse than ML in recovering the covariance structure of the growth parameters in most conditions.

3.6.2 Model Parameter Recovery

Model parameters were recovered with varied degrees of accuracy under different conditions. RJMCMC and DP correctly detected the number of classes under all conditions. The percentage of selecting a certain class solution by RJMCMC was higher in GCMs than in the 2-class and 3-class LGMMs. Within the LGMMs, the percentage tended to increase as the sample size increased under the same conditions. The split of the class proportions did not affect the percentage of selecting a class solution.

The class proportions were recovered better under RJMCMC and DP than under ML, as I discussed earlier. While ML tended to underestimate the minority class size, RJMCMC tended to overestimate the minority class size and underestimate the majority class size under some conditions (e.g., in the 33%/33%/33% condition when n = 200). DP performed well in recovering the class proportions under almost all conditions. The recovery of the class proportions usually improved as the sample size increased. The class proportions were recovered better under the evenly split class proportions (i.e., 50%/50%) than under the unevenly split class proportions (i.e., 80%/20%) in 2-class models. In 3-class models, the split of the class proportions was not a factor that clearly affected the recovery of class proportions.

Among all of the model parameters, the intercept means were recovered the best under all estimation methods across all sample sizes. DP outperformed ML and RJMCMC in recovering the growth parameter means under almost all conditions. The slope means were recovered best in the GCMs and better in the 3-class models compared to the 2-class models. Within the 2-class models, the unevenly split class proportions (i.e., 80%/20%) had better recovered slope means compared to the evenly split class proportions (i.e., 50%/50%). Within the 3-class models, the recovery of the slope means under the 70%/20%/10% conditions were slightly worse than the other two class proportion conditions. The estimation of the slope means usually improved as the sample size increased.

The growth parameter covariance structure was recovered worse than other parameters. In general, ML performed better than RJMCMC and DP, while DP performed the worst out of the three estimation methods in recovering the covariance structure and the residual variances. The growth parameter variances were recovered better under ML and RJMCMC compared to DP in most conditions. The recovery of the growth parameter variances tended to slightly improve as the sample size increased in the 3-class models, while they were not affected by the sample size in the 2-class models. The covariance was recovered poorly under all three estimation methods in almost all conditions; it was recovered well only under ML in the GCMs. The residual variances were recovered well under ML and DP and poorly recovered under RJMCMC in most conditions. The exception was in the GCMs, where the residual variances were recovered better under RJMCMC than under DP. The sample size did not affect the recovery of the covariance structure or the residual variances. The split of the class proportions did not seem to be a factor affecting the recovery of covariance structure or the residual variances.

3.6.3 Empirical Example

In the example, DP and RJMCMC were implemented through an LGMM on the 4-wave reading IRT scores from the ECLS-K dataset. The results of the data analysis showed contradicting findings. Specifically, RJMCMC extracted 2 latent classes, while DP extracted only 1 class.

When the estimates of RJMCMC are different from those of DP in terms of the number of latent classes, the percentage of selecting a certain class solution can be quite informative and useful. In this case, 89.1% of the iterations selected the 2-class solution, which suggests a relatively high proportion of a 2-class model versus other class solutions. This information provides us not only the possible number of classes being extracted but also the likelihood of this class solution appearing in all of the iterations. Therefore, the researcher can conclude that, since a relatively high majority (i.e., 89.1%) of the RJMCMC iterations selected the 2-class solution, she can be quite certain to make a decision on the 2-class solution as opposed to other numbers of classes.

The number of latent classes extracted by DP in this example can be tricky to interpret. The estimate for the class solution parameter is 1.443, which can conventionally be rounded to an integer 1. However, this value (1.443) is also very close to 1.5 and is therefore close to being rounded to 2, which will result in a completely different model and parameter estimates (not to mention a different substantive interpretation of the findings). It is common for non-parametric methods, such as DP and RJMCMC, to provide the estimate for the number of classes as a decimal value. This fact can make the decision making on the class solution quite arbitrary. An estimate such as 1.443 is normally decided as a 1-class solution and meanwhile it can also be considered as (or at least very close to) a 2-class solution considering some rounding errors during the data analysis process. These two different decisions will result in two different models, and the interpretations and substantive implications derived from the two models can be quite different and even contradictory. Therefore, applied researchers need to be very careful when making the decision on the final number of classes when they implement DP and get a more ambiguous estimate. In any case, substantive knowledge about the dataset should always be taken into account, along with the statistical outcomes.

3.6.4 Implications

There are a few implications of the findings of Study 1. First, the performance of RJMCMC and DP indicated they are reliable and efficient alternatives for the traditional ML estimation method that relies on model selection approaches. In the simulation study, I did not test multiple class solutions for ML. Instead I used ML with true models (i.e., models with the true number of classes as in the population) and compared them with RJMCMC and DP. The simulation study suggested that RJMCMC and DP yielded results that were comparably valid as the model selection approach could have done. Yet, RJMCMC and DP did not require running extra models with more competing class solutions.

Second, RJMCMC provided information about the number of classes that is not available from ML. Specifically, the percentage of selecting a certain number of classes quantifies the degree of certainty in a class solution. This information can provide us with some form of "guideline" in class enumeration for LGMMs.

Third, RJMCMC and DP provided an estimate for the class solution parameter as a decimal value, which should be interpreted with caution. As I previously discussed, decimal values can be interpreted arbitrarily. Applied researchers should always base their decision of the final class solution in part on the substantive information of their study, instead of being dictated solely by the statistical models.

Fourth, simulation results suggested that the covariance structure, especially the covariance parameter linked to the growth parameter means, was not recovered very well by RJMCMC and DP. One of the reasons that the covariance parameter had relatively high bias levels (i.e., almost always greater than 10%) could be that the bias was inflated by the population value of the covariance, which is relatively small (i.e., 0.3). In addition, the difficulties in estimating the covariance structure have been an issue in the mixture modeling context (see e.g., Depaoli, 2013). Unfortunately, RJMCMC and DP, with the current model and prior specifications in Study 1, are not likely to be a solution to this problem.

Chapter 4

Class Enumeration under Various Levels of Class Separation: Bayesian Non/Semi-Parametric Methods versus a Traditional Bayesian Approach

4.1 Introduction

In Study 2, I will focus on one of the most important issues that is linked to mixture modeling: namely, the influence of class separation. I am specifically interested in how class separation impacts the proper recovery of the mixture component parameter and the class-specific model parameters.

Class separation is usually characterized as degree of similarity (or difference) in the growth trajectories for multiple latent classes. Poor class separation may cause estimation issues in mixture modeling. For example, cases (or people) may be inaccurately assigned to the latent classes, causing the class size and growth parameters to be inaccurately estimated (see e.g., Tofighi & Enders, 2008; Nylund et al, 2007; Depaoli, 2013). In this dissertation, I will detail how class separation is an issue that is tied directly to class enumeration. Previous work has indicated that poor separation may produce a collapsed class structure (i.e., improper class enumeration) that may be substantively problematic (see e.g., van de Schoot et al, 2018).

Class separation has been studied via frequentist methods, as well as the Bayesian framework for various types of mixture models (e.g., Depaoli, 2013; Nylund et al, 2007; Tofighi & Enders, 2008; Tueller & Lubke, 2010, etc). In this study, I would like to extend the examination to the Bayesian non/semi-parametric methods (i.e., RJ MCMC and DP) and to compare their performance with the traditional Bayesian estimation methods.

The purpose of this study is to investigate the performance of the Bayesian non/semi-parametric methods under various degrees of class separation conditions. Specifically, I will examine how RJ MCMC and DP recover the number of latent classes and corresponding growth parameters under different levels of class separation and sample sizes. In addition, I will also include several Bayesian estimation methods as a comparison.

Study 2 is structured as follows. First, I will discuss the issues of class separation and its connection with class enumeration in mixture models. Second, I will conduct a simulation study on the performance of the Bayesian non/semi-parametric and the traditional Bayesian estimation methods for LGMMs. Third, I will present the results of the simulation in terms of the recovery of the number of latent classes and parameter estimates, percentage bias, and MSE of the model parameters.

4.1.1 Class Separation for Mixture Models

In LGMM, class separation may refer to the amount of overlap of the growth trajectories in each of the latent classes. Depending on the degree of overlap in the trajectories, the class separation can vary from poor to high. For example, when the classes are separated well in an LGMM, the growth trajectories between two classes will be clearly apart from each other in the intercept, growth trend, or both of these factors. In contrast, when class separation is poor, the trajectories from different latent classes may largely overlap, making them harder to distinguish statistically and substantively from one another. There are often other scenarios in between these two extreme conditions, where the growth trajectories do not have clear boundaries nor do they completely overlap; these middle-ground situations are likely more representative of the applied literature.

One of the measures that can be used to capture class separation is the (multivariate) Mahalanobis distance (MD). We can assume a two-component latent growth model with univariate normal mixtures. The MD value for the latent classes can be calculated with the following formula:

$$\mathrm{MD} = \sqrt{(\mu_1 - \mu_2)^T S^{-1} (\mu_1 - \mu_2)}, \tag{73}$$

where μ_1 and μ_2 are two vectors that represent the means of the latent growth factors (i.e., the intercept and the slope) in two adjacent latent classes, and S represents the covariance matrix of the latent factors.

Previous studies have investigated the influence of different MD levels on the model parameter recovery for several types of mixture models. For example, findings for structural equation mixture models suggested that a smaller MD (i.e., poorer class separation) could produce a larger estimate bias when the sample size was also small (Tueller & Lubke, 2010). Depaoli (2013) assessed the recovery of the mixture components and other parameters using frequentist and Bayesian estimation methods. Her findings suggested that, although separation was important, class separation did not affect the parameter estimates as much as the estimation methods (e.g., ML or Bayes) or other factors (e.g., sample size and class proportions) did.

Class separation is an important issue that is tied to the enumeration of latent classes in mixture models. For example, a simulation study by Tofighi and Enders (2008) indicated that class separation had a dramatic impact on the enumeration of growth mixture models. Specifically, well-separated trajectories of latent classes made it easier to correctly identify the number of classes.

Previous studies, whether employing frequentist estimation methods or the Bayesian framework, presume the number of latent classes before the analysis. Thus, the determination of the number of classes can be crucial. In substantive areas, the enumeration of classes can be accomplished by taking into account the substantive theories or findings from previous studies. In Study 1, I have introduced the Bayesian non/semi-parametric methods, which do not require knowledge of the number of classes *a priori*. I therefore would like to examine how class separation may affect these Bayesian non-parametric methods. Within this investigation, I will compare these methods to the traditional Bayesian approach under different prior settings representing different levels of (un)certainty about the class structure.

4.2 Design of Study 2

In Study 2, I investigated how class separation might affect the accuracy of class enumeration when implementing the RJMCMC and DP estimation methods. I also compared these Bayesian non-parametric methods to a traditional Bayesian approach under different prior settings, representing different levels of informativeness about the latent class structure.

4.2.1 Population Values and Simulation Conditions

In this simulation study, I used the same model specified in Study 1 (based on Equations 6-8) as the population model and the analysis models. In total, 2000 replications of datasets were generated using M*plus* 8 (Muthén & Muthén, 1998-2017). The population values used to generate the data were the same as in Study 1 except for values tied to the mechanism controlling class separation. Different from Study 1, the MD values became one of the factors being examined in this simulation study, which will be elaborated in detail. All population values of Study 2 are listed in Table 1.

Factors being investigated in this study include: the number of latent classes (2 levels), sample size (3 levels), class proportions (2 levels for 2-class conditions and 3 levels for 3-class conditions), and class separation (3 levels for 2-class conditions and 4 levels for 3-class conditions). The three levels of sample size examined were $n = 200$, $n = 400$, and $n = 600$. The number of latent classes ($C = 2$ and $C = 3$) was crossed with class proportions, resulting in five levels of class proportions (prop = $50\%/50\%$, prop = $80\%/20\%$, prop = $30\%/33\%/33\%$, prop = $45\%/45\%/45\%$, and prop = $70\%/20\%/10\%$).

Finally, as in Study 1, I was interested in the impact of class separation on the accuracy of results obtained through these approaches. Class separation was measured using the multivariate MD with seven different levels of separation specified. Specifically, I altered the degree of separation (through varying MD values), as well as the *location* of higher versus lower separation. In the 2-class conditions, I examined three levels of separation. That is, MD values that were specified between two latent classes were $\mathrm{MD} = 1$, $\mathrm{MD} = 2$, and $\mathrm{MD} = 3$. The levels of separation were crossed with the class proportions. See Columns 2 and 3 in Table 10 for detailed specification.

Table 10. Simulation Conditions for the Number of Class, Class Proportions, and Class Separation Levels

# of classes	2-Class		3-Class		
Prop	50/50	80/20	33/33/33	45/45/10	70/20/10
MD	1	1	1/1	1/1	1/1
	2	2	3/3	3/3	3/3
	3	3	1/3	1/3	1/3
	-	-	-	3/1	3/1

In the 3-class conditions, I assessed separation level, as well as the location of separation. According to the MD values specified between Class 1 and Class 2 and between Class 2 and Class 3, there could be different types of separation location. The degree of separation could be "high" and "high" between classes (i.e., MD = 3 between Class 1 and Class 2, and MD = 3 between Class 2 and Class 3). The degree of separation could also be "low" and "low" (i.e., MD = 1 between Class 1 and Class 2, and MD = 1 between Class 2 and Class 3). There could also be "low" separation between two adjacent classes and "high" separation between the other two adjacent classes (i.e., MD = 1 between Class 1 and Class 2, and MD = 3 between Class 2 and Class 3, vice versa.)

For the 3-class conditions with equal class sizes (33%/33%/33%), I examined two levels of separation ("high" and "low") and three separation locations. Take the 33%/33%/33% condition for example. First, two "low" separation levels were specified between latent classes, that is, MD = 1 between Class 1 and Class 2 and MD = 1 between Class 2 and Class 3. Second, two "high" separation levels (MD=3) were specified between Class 1 and Class 2 and between Class 2 and Class 3, respectively. Last, one "low" separation (MD = 1) was specified between Class 1 and Class 2 and one "high" separation (MD = 3) was specified between Class 2 and Class 3. Details of specification are listed in Column 4 in Table 10.

For other 3-class conditions with unequal class sizes, two levels of class separation and four separation locations were assessed. For example, the first three types of separation location in the 45%/45%/10% condition were akin to the 33%/33%/33% condition. Then the fourth type of separation location for the 45%/45%/10% condition switched the two MD values between Class 1 and Class 2 and between Class 2 and Class 3 as to manipulate the location of the low versus high separation. Details of 45%/45%/10% condition and other conditions are listed in Table 10. The MD values between latent classes were varied and arranged in such way that I could assess the location of class separation as it was linked to different class sizes.

4.2.2 Model Estimation Techniques

In this study, I also examined the performance of four types of model estimation techniques, including RJMCMC, DP, ML, and Bayesian estimation with two levels of prior specification. Simulation conditions were fully crossed with these model estimation techniques, which resulted in 204 simulation cells in total.

4.2.2.1 Prior Specifications for RJMCMC and DP

The analysis models for the RJMCMC and DP techniques used in this study were the same as in Study 1. The model for the RJMCMC algorithm was specified in Equations 36-38. The prior distributions are the same as described in Study 1. Specifically, $\mu_c \sim MVN(\binom{32}{2}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix})$, $\Sigma \sim IW(2, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix})$, $\sigma^2 \sim IG(0.01, 0.01)$, $B_F \sim N(0,10)$, and $(\pi_1 \ldots, \pi_C) \sim Dirichlet(1, \ldots 1)$. The model for the DP conditions was specified in Equations 58-65. Prior specifications implemented in Study 1 were also used in this study: $\mu_c \sim MVN(\binom{32}{2}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix})$, $\Sigma \sim IW(2, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix})$, $\sigma^2 \sim \mathrm{IG}(0.01, 0.01)$, $B_F \sim N(0,10)$, $G \sim DP(1, N(\mu, \Sigma))$, and $\Sigma_{\mathrm{b}} \sim IW\left(2, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right)$ (See the Methods section in Study 1 for notation and interpretation details).

The R package “miscF” (Feng, 2016) was used for conducting data analysis with RJMCMC, and the package “DPpackage” (Jara, Hanson, Quintana, Mueller, & Rosner, 2017) was used in the DP conditions. In total, I requested 5000 samples for each chain, with the first 2500 iterations discarded as the burn-in phase for the RJMCMC and DP conditions. ML conditions were conducted using M*plus* 8, with 100 perturbations of user-specified staring values and 25 final stage optimizations. Likewise, the traditional Bayesian methods (with two sets of prior specifications) were also conducted using M*plus* 8, and I describe these details next.

4.2.2.2 Prior Specifications for Bayesian Estimation Methods

4.2.2.2.1 Diffuse Priors

I implemented the Bayesian estimation framework using two different sets of priors: diffuse priors and weakly informed priors. Conditions with both prior specifications were conducted in M*plus*. In the conditions implementing diffuse priors (B-Diff), prior distributions with less information were implemented in order to reflect uncertainty regarding the model and the population(s) in an exploratory study. The default diffuse prior specifications in M*plus* were implemented on model parameters in the conditions with diffuse priors (Muthén & Muthén, 1998-2017). Specifically, the prior distributions for the growth parameters and the latent class proportions were specified as follows.

- $\mu_c \sim N(0, 10^{10})$, which was implemented for the intercept and slope means for each latent class under the B-Diff condition. This prior specification provided very little information about the parameter.
- $\Sigma \sim IW(0, -4)$, which was implemented on the covariance matrix in the B-Diff condition.
- $\pi_c \sim Dirichlet(10,10)$, which was implemented for the class proportions in the B-Diff condition, indicating at minimum 10 cases in each class. This prior

specification was the default Dirichlet prior implemented in M*plus* so as to avoid the situations (such as when $\pi_c \sim Dirichlet(1,1)$) in which the formation of small or inadmissible class solutions might occur.

- $\sigma^2 \sim IG(-1,0)$, which was implemented on the residual variance with little information.

4.2.2.2.2 Weakly informed priors

In the Bayesian estimation conditions implementing weakly informed priors (B-Weak), prior distributions with a "moderate" amount (or weak amount) of information were implemented as to reflect some degree of certainty with regards to the model and the population(s).[8] The weakly informed priors were derived by combining the values of diffuse priors in B-Diff conditions and population values in the generative model. Specifically, the hyperparameters of means of the growth factors were derived from the population mean, while the other parameters, such as the Dirichlet priors, the covariance matrix and the residual variance were implemented using the prior specifications from B-Diff condition. The weakly informed priors were specified in such way that it allowed me to mimic the situation where the researcher is likely to take a "best guess" on some model parameters and use the "default" settings for others.

As mentioned earlier, the class proportions followed a Dirichlet distribution that was specified as a default diffuse prior in M*plus* with $\pi_c \sim D(10,10)$ for all simulation cells. With the same reason stated in the previous subsection, hyperparameters 10 and 10 were specified here as to prevent formations of small or inadmissible class solutions. The covariance matrix was also implemented with the default diffuse priors in M*plus*. It followed an inverse Wishart distribution with $\Sigma \sim IW(0,-4)$. Priors for the growth parameter means varied across simulation cells according to the different conditions. The mean hyperparameter in the normal prior distributions for the means of the growth factors was set at each parameter's population value as to center the prior distribution on the population value. The variance hyperparameter for the same parameters were fixed at 100, which suggests some variation around the center of the distribution but not completely diffuse, compared with the default diffuse specification in M*plus*, 10^{10}. This combination of mean and variance hyperparaemters for the mean of the growth factors suggested a normal prior distribution with some information about the center of the true value while still incorporated with some degree of uncertainty with its variation around the center.

For example, in the conditions of 2-class and MD = 1, the mean of the intercept in Class 1 followed a normal distribution with hyperparameters of 31.37 and 100 (see Row 3, Column 3 in Table 11). The mean of the slope in Class 1 in the same condition followed a normal distribution with hyperparameters of 1.802 and 100 (see Row 3, Column 6 in Table 11). Similarly, in the 3-class conditions where MD = 1 between Class 1 and Class 2 and MD = 3 between Class 2 and Class 3, the prior of the mean of the intercept was $\mathrm{N}(31.37,100)$ in Class 1, $\mathrm{N}(34.29,100)$ in Class 2 and $\mathrm{N}(43.01,100)$ in Class 3 (see Row 7, Columns 3-5 in Table 11). The details of the prior specification for the growth parameters in B-Weak conditions can be found in Table 11.

---

[8] Weakly informed priors are defined as priors that reflect less information than the researcher "actually has". (Gelman, Jakulin, Pittau, Pittau, & Su, 2008).

Table 11. Hyperparameter Values for Growth Parameters in Study 2

# of classes	MD	Mean-I			Mean-S		
		Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
2-Class	1	N(31.37,100)	N(34.29,100)		N(1.802,100)	N(2.834,100)	
	2	N(31.37,100)	N(37.182,100)		N(1.802,100)	N(3.856,100)	
	3	N(31.37,100)	N(40.09,100)		N(1.802,100)	N(4.884,100)	
3-Class	1/1	N(31.37,100)	N(34.29,100)	N(37.21,100)	N(1.802,100)	N(2.834,100)	N(3.866,100)
	1/3	N(31.37,100)	N(34.29,100)	N(43.01,100)	N(1.802,100)	N(2.834,100)	N(5.917,100)
	3/1	N(31.37,100)	N(40.09,100)	N(43.01,100)	N(1.802,100)	N(4.884,100)	N(5.917,100)
	3/3	N(31.37,100)	N(40.09,100)	N(48.81,100)	N(1.802,100)	N(4.884,100)	N(7.967,100)

4.3 Results of Study 2

In Study 2, I investigated the effect of class separation on the accuracy of class enumeration and model parameter estimation when implementing RJMCMC, DP, and two Bayesian estimation methods including diffuse priors (B-Diff) and weakly informed priors (B-Weak). I assessed how well the number of classes, the class proportions, and the growth parameters were recovered under various levels of class separation. Akin to Results from Study 1, I present the accuracy of the estimation in terms of percent bias and the MSE values for the growth parameters. All of the simulation results in Study 2 are listed in Tables 12-39. Tables for MSE values (Tables A3-A16) are presented in Appendix A.

4.3.1 Convergence

In Study 2, I used two forms of convergence diagnostics to assess the convergence of the sampling procedure within and between the Markov chains. The first form of diagnostic is the Geweke (1992) statistic, which was computed in R and used for RJMCMC and DP. The second form is the Potential Scale Reduction (PSR) factor provided in M*plus* for the B-diff and B-weak conditions (Muthén, 2010).

As described in Section 3.4.1, the Geweke diagnostic evaluates the convergence for a single Markov chain using a standard z-score. If the z-score is a small value (e.g., less than 2), then the chain can be considered reaching the convergence. The z-scores were less than 2.000 for all parameters in all simulation conditions in Study 2.

The PSR diagnostic is used when there is more than one Markov chain requested. It compares the parameter variation within each chain to that between chains, which is similar to a classical analysis of variance approach (Muthén, 2010). The PSR criterion requires the between-chain variation to be small relative to the total of between- and within-chain variation (Gelman & Rubin, 1992). In M*plus*, the default convergence criterion is that a PSR factor is 1.05 for each replication, which was used for the convergence diagnostic in this study. Any replication with a PSR value higher than 1.05

was marked as un-converged and was removed from the final estimation process. Therefore, The PSR values in the B-Diff and B-Weak conditions were all below 1.05 for all of the converged replications.

In addition to the Markov chain convergence being assessed for each parameter, I also monitored the model convergence of the total 2000 replications. All RJMCMC and DP conditions reached 100% model convergence, while the Bayesian estimation conditions achieved at least 85% convergence out of 2000 replications. All results that are presented here were calculated with the parameter estimates from the converged replications.

4.3.2 Class Enumeration and Parameter Estimates

4.3.2.1 2-Class Prop=50/50 MD=1 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 50/50 MD = 1 conditions are presented in Table 12 and the top section of Table A3. The number of classes was estimated accurately for RJMCMC and DP across sample sizes. The percentages of selecting a 2-class solution by RJMCMC were below 70%, which indicates that the confidence of a 2-class solution when MD is 1 was moderate based on the estimation of RJMCMC. The class proportions were recovered well for all estimation methods.

The intercept parameter means were recovered well for all estimation methods. However, the slope parameter means had relatively higher bias for B-Diff (around -15% to 23%) and B-Weak (around -14% to 23%), compared with those for DP (around 4% to 23%). RJMCMC yielded slightly high bias (around 13% to 26%) for the slope parameter means for the n = 200 condition and improve as the sample size increased (bias levels were around -5% to -8% when n = 600). The intercept parameter variance was estimated well, except for DP when n = 600 with a bias level around 23%. The slope variance was recovered well in B-Diff and B-Weak but was overestimated by RJMCMC and DP in the n = 200 condition and by DP in the n = 400 condition (bias levels were around 12% to 39%). The covariance parameter was consistently poorly recovered by all estimation methods across sample sizes (bias levels were around -54% to 280%). The MSE values are small for all parameters, except for the intercept variance in the DP condition for n = 600 (bias was around -14%).

Table 12. Parameter Estimates and Percent Bias for 2-Class 50/50 MD=1

	Pop	n=200 B-Diff	n=200 B-Weak	n=200 RJMCMC	n=200 DP	n=400 B-Diff	n=400 B-Weak	n=400 RJMCMC	n=400 DP	n=600 B-Diff	n=600 B-Weak	n=600 RJMCMC	n=600 DP
Parameter Estimates													
# classes	2.000	2.000	2.000	2.322 (0.643)	1.721	2.000	2.000	2.268 (0.675)	1.775	2.000	2.000	2.232 (0.672)	1.866
Prop C1		0.503	0.499	0.542	0.548	0.499	0.501	0.476	0.476	0.500	0.498	0.478	0.524
Prop C2		0.497	0.501	0.458	0.452	0.501	0.499	0.524	0.524	0.500	0.502	0.522	0.477
Mean I C1	31.370	31.488	31.492	29.266	30.264	31.569	31.550	32.612	32.010	31.598	31.591	29.716	32.408
Mean I C2	34.290	34.188	34.164	32.886	32.889	34.087	34.103	35.205	33.705	34.059	34.060	32.595	34.199
Mean S C1	1.802	2.228	2.217	2.277	1.890	2.216	2.219	1.992	1.955	2.206	2.215	1.710	1.914
Mean S C2	2.834	2.404	2.419	3.221	3.317	2.412	2.414	3.047	2.702	2.422	2.415	2.608	2.603
Var-I	16.000	16.233	16.283	15.089	17.520	16.304	16.258	14.972	17.534	16.364	16.366	17.367	19.737
Var-S	2.000	2.056	2.051	2.239	2.774	2.057	2.059	1.951	2.653	2.058	2.057	1.893	2.245
Cov	0.300	0.918	0.906	0.139	0.616	0.914	0.907	1.018	1.139	0.900	0.907	0.672	0.900
Residual	0.500	0.520	0.520	0.700	0.513	0.507	0.507	0.687	0.472	0.507	0.505	0.716	0.515
Percent Bias													
Prop C1		0.640	-0.244	8.333	9.666	-0.150	0.144	-4.833	-4.820	0.038	-0.410	-4.356	4.700
Prop C2		-0.640	0.244	-8.333	-9.666	0.150	-0.144	4.833	4.820	-0.038	0.410	4.356	-4.700
Mean I C1	31.370	0.376	0.389	-6.708	-3.524	0.636	0.575	3.958	2.039	0.727	0.704	-5.271	3.309
Mean I C2	34.290	-0.298	-0.369	-4.095	-4.087	-0.591	-0.546	2.667	-1.705	-0.674	-0.670	-4.944	-0.266
Mean S C1	1.802	**23.629**	**23.013**	**26.355**	4.894	**22.986**	**23.141**	**10.521**	8.474	**22.420**	**22.897**	-5.121	6.227
Mean S C2	2.834	**-15.187**	**-14.661**	**13.659**	**17.033**	**-14.905**	**-14.834**	7.514	-4.671	**-14.524**	**-14.795**	-7.981	-8.146
Var-I	16.000	1.454	1.766	-5.693	9.499	1.899	1.614	-6.425	9.585	2.275	2.288	8.542	**23.359**
Var-S	2.000	2.780	2.525	**11.949**	**38.678**	2.830	2.970	-2.467	**32.658**	2.875	2.870	-5.373	**12.255**
Cov	0.300	**205.867**	**201.900**	**-53.656**	**105.220**	**204.667**	**202.300**	**239.343**	**279.747**	**200.000**	**202.267**	**124.114**	**200.146**
Residual	0.500	3.920	4.080	**39.952**	2.606	1.380	1.400	**37.410**	-5.674	1.340	1.020	**43.218**	2.990

4.3.2.2 2-Class Prop=50/50 MD=2 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 50/50 MD = 2 conditions are presented in Table 13 and the top section of Table A4. The number of classes was estimated accurately by RJMCMC and DP. The percentages of selecting a 2-class solution by RJMCMC were all above 80%, which suggested a higher certainty about the 2-class solution compared to the conditions when MD = 1 (where the percentages were only around 64% to 67%). The class proportions were recovered well by all estimation methods across sample sizes.

The intercept parameter means were recovered with low bias levels (all below 10%) by all estimation methods for all sample sizes. The slope parameter means were estimated with slightly higher bias (around 12% to 22%) when the sample size was relatively small, but it was improved when n = 600 (bias was below 10%). The growth parameter variances were poorly recovered by B-Diff and B-Weak in the n = 200 and n = 400 conditions (bias levels were around 12% to 82%) and the bias levels decreased when n = 600 for both Bayesian estimation methods (bias was below 10%). The bias levels (around 52% to 82%) for the variance parameters were high for DP across sample sizes and held an inconsistent pattern for RJMCMC. The covariance was consistently poorly recovered by all estimation methods across sample sizes (bias levels were all above 80%). The MSE values were small for most parameters, although some intercept variances had relatively large MSE values (around 76).

Table 13. Parameter Estimates and Percent Bias for 2-Class 50/50 MD=2

			n=200				n=400				n=600			
			B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
		Pop	Parameter Estimates											
# classes		2.000	2.000	2.000	2.249 (0.811)	1.834	2.000	2.000	2.158 (0.863)	1.826	2.000	2.000	1.903 (0.851)	1.936
Prop	C1		0.502	0.501	0.452	0.520	0.502	0.503	0.506	0.526	0.500	0.499	0.467	0.513
	C2		0.498	0.499	0.548	0.480	0.498	0.497	0.494	0.474	0.500	0.501	0.533	0.487
Mean I C1		31.370	32.091	32.070	29.601	30.613	31.824	31.814	33.012	33.017	31.628	31.628	29.455	31.806
Mean I C2		37.182	36.480	36.497	39.187	35.173	36.742	36.752	39.097	38.090	36.917	36.915	39.087	35.092
Mean S C1		1.802	2.199	2.197	2.029	1.928	2.026	2.029	2.012	1.973	1.928	1.927	1.735	1.848
Mean S C2		3.857	3.459	3.463	4.200	4.100	3.631	3.629	4.127	3.626	3.722	3.723	3.557	3.724
Var-I		16.000	19.436	19.371	13.867	24.592	18.063	18.054	14.972	22.453	17.238	17.256	18.660	24.751
Var-S		2.000	2.440	2.438	1.600	3.633	2.278	2.276	1.951	3.183	2.175	2.170	1.993	3.044
Cov		0.300	1.803	1.793	0.835	2.027	1.218	1.212	1.018	2.737	0.861	0.854	0.057	3.075
Residual		0.500	0.521	0.520	0.692	0.509	0.508	0.506	0.687	0.482	0.507	0.506	0.716	0.535
			Percent Bias											
Prop	C1		0.480	0.264	-9.667	4.033	0.312	0.562	1.167	5.126	-0.052	-0.150	-6.556	2.554
	C2		-0.480	-0.264	9.667	-4.033	-0.312	-0.562	-1.167	-5.126	0.052	0.150	6.556	-2.554
Mean I C1		31.370	2.297	2.232	-5.640	-2.412	1.446	1.416	5.233	5.250	0.822	0.823	-6.103	1.390
Mean I C2		37.182	-1.889	-1.843	5.394	-5.403	-1.184	-1.158	5.149	2.442	-0.714	-0.719	5.122	-5.621
Mean S C1		1.802	**22.031**	**21.898**	**12.585**	6.986	**12.442**	**12.614**	**11.631**	9.503	6.992	6.915	-3.707	2.543
Mean S C2		3.857	**-10.309**	**-10.207**	8.882	6.309	-5.862	-5.911	7.005	-5.992	-3.495	-3.477	-7.770	-3.442
Var-I		16.000	**21.477**	**21.069**	**-13.333**	**53.702**	**12.896**	**12.838**	-6.425	**40.330**	7.736	7.849	**16.625**	**54.694**
Var-S		2.000	**21.995**	**21.910**	**-20.005**	**81.643**	**13.885**	**13.810**	-2.467	**59.152**	8.740	8.520	-0.346	**52.180**
Cov		0.300	**500.967**	**497.700**	**178.389**	**575.558**	**305.867**	**303.967**	**239.343**	**812.275**	**187.100**	**184.800**	**-81.020**	**925.094**
Residual		0.500	4.220	3.980	**38.369**	1.712	1.500	1.240	**37.410**	-3.589	1.320	1.100	**43.218**	6.963

4.3.2.3 2-Class Prop=50/50 MD=3 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 50/50 MD = 3 conditions are presented in Table 14 and the top section of Table A5. The number of classes was recovered accurately by RJMCMC and DP for all sample sizes. The percentages of selecting a 2-class solution by RJMCMC were relatively high, especially for larger sample sizes (around 85% to 93%). The class proportions were recovered well by all estimation methods for all sample sizes.

The growth parameter means were estimated very well with one exception for RJMCMC when n = 200, which had slightly higher bias (around -16%) for the slope in Class 2. The growth parameter variances were recovered well for B-Diff and B-Weak across sample sizes (bias levels were all below 10%). However, RJMCMC and DP produced higher bias levels (around -12% to 70%) for the variance parameters when n = 200. The bias decreased as the sample size increased (e.g., decreased to as low as 1.024% when n = 600). The covariance was consistently poorly estimated for all estimation methods for all sample sizes (around 15% to over 900%). The MSE values were small except for the variance of the intercept in the DP condition when n = 200 (MSE value was around 56).

Table 14. Parameter Estimates and Percent Bias for 2-Class 50/50 MD=3

			n=200				n=400				n=600			
			B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
		Pop	Parameter Estimates											
# classes		2.000	2.000	2.000	1.801 (0.853)	2.297	2.000	2.000	2.068 (0.933)	2.257	2.000	2.000	1.978 (0.940)	2.108
Prop	C1		0.500	0.500	0.482	0.484	0.501	0.501	0.486	0.507	0.501	0.501	0.497	0.494
	C2		0.500	0.500	0.518	0.516	0.499	0.499	0.514	0.493	0.499	0.499	0.503	0.506
Mean I C1		31.370	31.438	31.439	30.766	30.269	31.401	31.400	32.512	32.020	31.387	31.387	30.204	32.411
Mean I C2		40.090	40.020	40.019	38.686	38.682	40.071	40.071	40.055	41.996	40.076	40.076	39.995	39.994
Mean S C1		1.802	1.824	1.825	1.874	1.878	1.809	1.809	1.712	1.911	1.806	1.806	1.811	1.813
Mean S C2		4.885	4.859	4.857	4.098	5.170	4.876	4.876	5.255	5.256	4.878	4.877	4.870	4.976
Var-I		16.000	16.958	16.966	18.381	23.534	16.373	16.376	14.983	16.689	16.230	16.234	17.436	16.164
Var-S		2.000	2.132	2.132	1.748	3.394	2.059	2.060	2.066	2.156	2.037	2.037	1.732	2.374
Cov		0.300	0.518	0.525	0.491	3.194	0.376	0.377	0.564	0.554	0.346	0.346	0.357	0.396
Residual		0.500	0.519	0.522	0.697	0.505	0.506	0.507	0.665	0.477	0.506	0.506	0.626	0.520
			Percent Bias											
Prop	C1		-0.072	-0.066	-3.667	-3.171	0.186	0.176	-2.835	1.400	0.134	0.132	-0.556	-1.168
	C2		0.072	0.066	3.667	3.171	-0.186	-0.176	2.835	-1.400	-0.134	-0.132	0.556	1.168
Mean I C1		31.370	0.216	0.219	-1.926	-3.510	0.099	0.096	3.639	2.071	0.055	0.055	-3.718	3.319
Mean I C2		40.090	-0.174	-0.177	-3.503	-3.512	-0.047	-0.048	-0.088	4.754	-0.035	-0.035	-0.238	-0.239
Mean S C1		1.802	1.221	1.299	3.972	4.226	0.388	0.366	-5.017	6.045	0.228	0.222	0.519	0.587
Mean S C2		4.885	-0.540	-0.565	**-16.101**	5.842	-0.180	-0.184	7.580	7.596	-0.154	-0.156	-0.300	1.855
Var-I		16.000	5.989	6.039	**14.881**	**47.089**	2.329	2.349	-6.356	4.305	1.438	1.461	8.977	1.024
Var-S		2.000	6.575	6.620	**-12.591**	**69.687**	2.970	3.000	3.307	7.804	1.870	1.860	**-13.394**	**18.712**
Cov		0.300	**72.667**	**74.967**	**63.734**	**964.775**	**25.433**	**25.600**	**88.166**	**84.755**	**15.267**	**15.433**	**19.100**	**32.014**
Residual		0.500	3.880	4.320	**39.393**	1.015	1.260	1.320	**32.980**	-4.532	1.120	1.240	**25.187**	3.902

4.3.2.4 2-Class Prop=80/20 MD=1 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 80/20 MD = 1 conditions are presented in Table 15 and the bottom section of Table A3. The number of classes was estimated accurately by RJMCMC. However, when n = 200, DP was not able to correctly detect the number of classes and yielded a decimal number that was rounded to 1 (i.e., 1.376). The percentage of selecting a 2-class solution by RJMCMC was relatively low when the sample sizes were smaller (i.e., around 56% for n = 200 and 67% for n = 400). Combining the performance of DP and RJMCMC in estimating the number of classes when n = 200, it appeared that a 2-class solution might not be the optimal option based on the estimation of the two Bayesian non-parametric methods. It might indicate that the Bayesian non-parametric methods, especially DP, had difficulty in accurately enumerating the latent classes when there was a majority class and when the sample size was relatively small (and separation was poor with MD = 1). The class proportions were recovered with high bias levels (all above -34%) by B-Diff and B-Weak across sample sizes. RJMCMC yielded slight high bias for the minority class when the sample size was relatively small (bias the level was around -18%) but improved when the sample size reached 600 (bias level was below 10%). Overall, DP underestimated the size of the minority class; it was not able to produce the estimates for the class proportions for the n = 200 condition.

The intercept parameter means were estimated well for B-Diff, B-Weak, and RJMCMC. DP did not provide an estimate of the mean of a second class in the n = 200 condition. B-Diff and B-Weak yielded low bias for the slope parameter means in the majority class but underestimated the minority class for all sample sizes (bias levels were around -27%). RJMCMC produced slightly high bias levels (around 10% to 21%) for the slope parameter means in the minority class, and low bias (below 10%) for those in the majority class. DP did not provide an estimate for the slope parameter mean in the (minority) second class. The growth parameter variances were estimated well by B-Diff and B-Weak and had high bias levels (around 7% to -29%) in the RJMCMC and DP conditions for all sample sizes. The covariance parameter was poorly recovered by all estimation methods across sample sizes (bias levels were around 24% to over 500%). The MSE values were small for most mean and variance parameters, but they were higher for some of the intercept variances (e.g., MSE around 22).

Table 15. Parameter Estimates and Percent Bias for 2-Class 80/20 MD=1

			n=200				n=400				n=600			
			B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
		Pop	Parameter Estimates											
# classes		2.000	2.000	2.000	2.389 (0.560)	1.376	2.000	2.000	2.313 (0.667)	1.652	2.000	2.000	2.167 (0.757)	2.247
Prop	C1		0.513	0.514	0.837	-	0.512	0.521	0.824	0.844	0.525	0.525	0.788	0.834
	C2		0.487	0.486	0.163	-	0.488	0.479	0.176	0.156	0.475	0.475	0.212	0.166
Mean I C1		31.370	30.683	30.689	33.244	33.028	30.784	30.791	33.827	31.821	30.837	30.831	29.257	30.262
Mean I C2		34.290	33.282	33.292	36.570	-	33.210	33.213	34.654	32.844	33.189	33.191	32.662	33.644
Mean S C1		1.802	1.960	1.960	2.076	1.275	1.968	1.966	2.172	1.751	1.966	1.959	1.620	1.890
Mean S C2		2.834	2.065	2.065	3.103	-	2.064	2.068	3.104	3.193	2.076	2.079	2.604	2.503
Var-I		16.000	15.596	15.607	11.331	18.795	15.670	15.668	18.555	17.126	15.753	15.751	18.444	18.987
Var-S		2.000	1.979	1.984	1.483	2.438	1.977	1.984	2.299	2.437	1.988	1.989	2.224	2.328
Cov		0.300	0.697	0.693	0.228	0.762	0.704	0.706	1.886	1.012	0.699	0.698	0.603	0.816
Residual		0.500	0.521	0.523	0.733	0.520	0.509	0.508	0.737	0.473	0.507	0.507	0.519	0.516
			Percent Bias											
Prop	C1		**-35.871**	**-35.766**	4.582	-	**-36.006**	**-34.895**	2.962	5.478	**-34.409**	**-34.423**	-1.458	4.305
	C2		**143.485**	**143.065**	**-18.330**	-	**144.025**	**139.580**	**-11.850**	**-21.910**	**137.635**	**137.690**	5.833	**-17.220**
Mean I C1		31.370	-2.189	-2.172	5.973	5.286	-1.870	-1.845	7.831	1.438	-1.698	-1.719	-6.735	-3.531
Mean I C2		34.290	-2.940	-2.911	6.650		-3.149	-3.141	1.062	-4.218	-3.211	-3.206	-4.747	-1.885
Mean S C1		1.802	8.762	8.746	**15.222**	**-29.224**	9.206	9.123	**20.536**	-2.848	9.079	8.729	**-10.123**	4.880
Mean S C2		2.834	**-27.131**	**-27.131**	9.504	-	**-27.167**	**-27.018**	9.533	**12.657**	**-26.764**	**-26.630**	-8.110	**-11.664**
Var-I		16.000	-2.527	-2.459	**-29.181**	**17.467**	-2.063	-2.077	**15.969**	7.035	-1.542	-1.556	**15.273**	**18.667**
Var-S		2.000	-1.035	-0.805	**-25.842**	**21.899**	-1.130	-0.825	**14.932**	**21.858**	-0.620	-0.570	**11.208**	**16.421**
Cov		0.300	**132.467**	**130.933**	**-24.104**	**153.902**	**134.800**	**135.267**	**528.765**	**237.316**	**132.867**	**132.767**	**101.162**	**172.060**
Residual		0.500	4.100	4.640	**46.593**	4.021	1.760	1.660	**47.463**	-5.494	1.480	1.440	3.771	3.182

4.3.2.5 2-Class Prop=80/20 MD=2 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 80/20 MD = 2 conditions are presented in Table 16 and the bottom section of Table A4. Both RJMCMC and DP were able to accurately estimate the number of classes. The percentage of selecting a 2-class solution by RJMCMC was slightly higher when n = 600 (i.e., 81.6%) than when the sample sizes were 200 and 400 (i.e., 71.3% and 75%, respectively). The class proportions were recovered poorly by B-Diff and B-Weak when n = 200 (bias levels were around -25% to 101%), but the bias levels decreased as the sample size increased (around -8% to 31% when n = 600). RJMCMC tended to overestimate the minority class in the n = 200 and n = 400 conditions (bias levels were around 15% to 19%) but underestimated in the n = 600 condition (bias levels were around -27%). DP recovered the class proportions with relatively high bias when n = 200 (bias levels were around -14% to 60%) and improved as the sample size increased (bias levels were around 3% to -13% when n = 600).

The intercept parameter means were estimated with low bias by all estimation methods for all sample sizes. B-Diff, B-Weak, and RJMCMC yielded high bias levels (as high as -28%) for the slope parameter means when n = 200 and n = 400; estimation accuracy improve when the sample size reached 600 (bias levels were below 10%). DP produced slightly higher bias levels (around -17%) for the slope parameter means when the sample size was 400 and lower bias levels (below 10%) for other sample sizes. Variances were recovered with slightly higher bias levels (around 13% to 15%) by B-Diff and B-Weak when the sample size was 200 and was improved (bias levels decreased to below 10%) when the sample size increased to 400 and 600. DP consistently estimated the variances with high bias levels (around 12% to 58%) All estimation methods yielded high bias levels (around 15% to 590%) for the covariance parameter. The MSE values were small for most parameters, except for a few large values (around 16 to 34) for the variances of the intercept.

Table 16. Parameter Estimates and Percent Bias for 2-Class 80/20 MD=2

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
			Parameter Estimates											
# classes		2.000	2.000	2.000	2.369 (0.713)	1.674	2.000	2.000	2.212 (0.75)	1.772	2.000	2.000	2.17 (0.816)	2.234
Prop	C1		0.598	0.600	0.761	0.688	0.680	0.682	0.770	0.789	0.738	0.742	0.853	0.827
	C2		0.402	0.400	0.239	0.312	0.320	0.318	0.230	0.211	0.262	0.258	0.147	0.173
Mean I C1		31.370	31.120	31.102	29.023	31.030	31.155	31.151	33.672	32.826	31.175	31.176	28.344	31.260
Mean I C2		37.182	34.770	34.828	39.462	36.422	35.553	35.599	35.716	34.718	36.131	36.163	39.554	38.545
Mean S C1		1.802	1.934	1.940	1.736	1.975	1.838	1.835	1.884	1.774	1.783	1.780	1.682	1.720
Mean S C2		3.857	2.753	2.759	4.126	4.032	3.146	3.163	4.027	3.219	3.423	3.433	3.627	3.522
Var-I		16.000	18.132	18.047	13.258	21.845	17.049	17.003	17.638	19.942	16.464	16.430	16.434	19.280
Var-S		2.000	2.293	2.296	1.633	3.111	2.157	2.151	2.285	3.144	2.070	2.068	2.137	2.254
Cov		0.300	1.444	1.441	0.347	2.680	0.912	0.891	1.253	2.069	0.581	0.569	0.382	0.880
Residual		0.500	0.521	0.519	0.624	0.525	0.508	0.507	0.541	0.480	0.506	0.507	0.542	0.519
			Percent Bias											
Prop	C1		**-25.199**	**-24.969**	-4.863	**-13.963**	**-14.988**	**-14.701**	-3.780	-1.321	-7.704	-7.274	6.675	3.354
	C2		**100.795**	**99.875**	**19.450**	**55.850**	**59.950**	**58.805**	**15.120**	5.285	**30.815**	**29.095**	**-26.700**	**-13.415**
Mean I C1		31.370	-0.797	-0.855	-7.483	-1.085	-0.686	-0.698	7.338	4.640	-0.622	-0.617	-9.647	-0.349
Mean I C2		37.182	-6.488	-6.332	6.133	-2.043	-4.381	-4.259	-3.942	-6.627	-2.826	-2.742	6.381	3.666
Mean S C1		1.802	7.336	7.664	-3.650	9.625	1.970	1.831	4.523	-1.549	-1.065	-1.210	-6.640	-4.536
Mean S C2		3.857	**-28.618**	**-28.473**	6.970	4.527	**-18.424**	**-17.988**	4.398	**-16.535**	**-11.263**	**-10.990**	-5.972	-8.685
Var-I		16.000	**13.323**	**12.796**	**-17.137**	**36.529**	6.555	6.270	**10.238**	**24.636**	2.899	2.689	2.711	**20.502**
Var-S		2.000	**14.650**	**14.820**	**-18.328**	**55.529**	7.835	7.545	**14.273**	**57.213**	3.520	3.375	6.857	**12.703**
Cov		0.300	**381.167**	**380.467**	**15.578**	**793.345**	**203.933**	**196.867**	**317.511**	**589.591**	**93.633**	**89.500**	**27.437**	**193.244**
Residual		0.500	4.280	3.700	**24.708**	5.051	1.660	1.360	8.268	-4.079	1.240	1.480	8.454	3.870

4.3.2.6 2-Class Prop=80/20 MD=3 Conditions

Results for the parameter estimates and percent bias levels in the 2-class prop = 80/20 MD = 3 conditions are presented in Table 17 and the bottom section of Table A5. RJMCMC and DP correctly estimated the number of classes. The percentages of selecting a 2-class solution by RJMCMC were relatively high (i.e., around 75% to 86%), indicating a higher certainty compared lower degrees of class separation. The class proportions were estimated well for all estimation methods across all sample sizes. The two exceptions were in the n = 200 and n = 400 conditions where DP yielded bias levels that were slightly higher than 10%.

The means of the growth parameters were recovered well for all estimation methods for all sample sizes (bias levels were all below 10%). B-Diff and B-Weak yielded low bias (below 10%) for the variances of the growth parameters. The estimates from RJMCMC and DP were inconsistent across sample sizes for the growth parameter variances (bias levels were around 2% to 45%). The covariance was recovered with high bias level (above 30%) by all estimation methods when n = 200. The bias decreased to below 10% only for B-Diff and DP when the sample sizes increased. The MSE values were relatively small for all parameters except for the variance of the intercept in the RJMCMC condition when n = 600, which was around 11.

Table 17. Parameter Estimates and Percent Bias for 2-Class 80/20 MD=3

			n=200				n=400				n=600			
			B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
		Pop	Parameter Estimates											
# classes		2.000	2.000	2.000	1.851 (0.778)	2.353	2.000	2.000	2.155 (0.854)	2.234	2.000	2.000	1.935 (0.855)	2.301
Prop	C1		0.784	0.785	0.817	0.779	0.802	0.802	0.811	0.824	0.805	0.811	0.789	0.814
	C2		0.216	0.215	0.183	0.221	0.198	0.198	0.189	0.177	0.195	0.189	0.211	0.186
Mean I C1		31.370	31.264	31.264	30.043	30.023	31.320	31.391	32.584	32.827	31.336	31.383	30.486	30.257
Mean I C2		40.090	39.536	39.554	40.730	39.366	39.882	40.083	37.624	39.618	39.955	40.087	39.762	37.945
Mean S C1		1.802	1.770	1.769	1.704	1.977	1.781	1.806	1.947	1.773	1.788	1.805	1.639	1.721
Mean S C2		4.885	4.684	4.691	5.054	5.350	4.808	4.880	5.059	5.052	4.833	4.879	4.955	4.750
Var-I		16.000	16.661	16.633	17.877	13.816	16.189	16.342	14.102	16.669	16.135	16.234	19.372	16.669
Var-S		2.000	2.096	2.094	2.085	2.904	2.037	2.056	1.935	2.295	2.022	2.035	1.855	2.295
Cov		0.300	0.427	0.416	0.399	0.511	0.306	0.358	0.554	0.272	0.304	0.339	0.173	0.272
Residual		0.500	0.521	0.521	0.671	0.473	0.506	0.507	0.560	0.528	0.506	0.507	0.516	0.528
		Pop	Percent Bias											
Prop	C1		-2.051	-1.904	2.069	-2.668	0.266	0.267	1.375	2.938	0.626	1.420	-1.419	1.705
	C2		8.205	7.615	-8.278	**10.670**	-1.065	-1.070	-5.500	**-11.750**	-2.505	-5.680	5.675	-6.820
Mean I C1		31.370	-0.339	-0.337	-4.229	-4.295	-0.161	0.068	3.870	4.646	-0.109	0.040	-2.817	-3.547
Mean I C2		40.090	-1.382	-1.336	1.597	-1.806	-0.520	-0.017	-6.150	-1.177	-0.337	-0.007	-0.817	-5.351
Mean S C1		1.802	-1.804	-1.842	-5.444	9.725	-1.165	0.239	8.038	-1.616	-0.777	0.150	-9.073	-4.512
Mean S C2		4.885	-4.123	-3.980	3.459	9.510	-1.568	-0.098	3.569	3.421	-1.067	-0.123	1.429	-2.768
Var-I		16.000	4.131	3.957	**11.734**	**-13.648**	1.181	2.138	**-11.861**	4.181	0.843	1.464	**21.076**	4.181
Var-S		2.000	4.820	4.675	4.239	**45.200**	1.855	2.785	-3.226	**14.731**	1.095	1.745	-7.249	**14.731**
Cov		0.300	**42.167**	**38.633**	**33.038**	**70.323**	2.133	**19.167**	**84.741**	-9.226	1.467	**12.867**	**-42.365**	-9.226
Residual		0.500	4.260	4.240	**34.241**	-5.375	1.260	1.400	**12.053**	5.605	1.240	1.360	3.225	5.605

4.3.2.7 3-Class Prop=33/33/33 MD=1/1 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 33/33/33 MD = 1/1 conditions are presented in Tables 18, 19 and A6. RJMCMC and DP recovered the number of classes accurately. The percentages of selecting a 3-class solution by RJMCMC were below 70%, suggesting a moderate certainty of a 3-class solution. The class proportions were estimated well by B-Diff and B-Weak across sample sizes; the bias levels were slightly higher than 10% for RJMCMC and DP when n = 200 and n = 400, and they decreased to below 10% as the sample size reached 600.

The intercept means were estimated well by all methods for all sample sizes. B-Diff and B-Weak consistently overestimated Class 1 (bias levels were above 39%) and underestimated Class 3 (bias levels were above 19%) across sample sizes for the slope means. RJMCMC and DP yielded estimates of inconsistent bias levels for the slope parameter means (below 1% to above 40%). B-Diff and B-Weak produced low bias (all below or around 10%) for the growth parameter variances, while RJMCMC and DP yielded relatively high bias levels (up to 58%) for the same parameters. The covariance had high bias levels (around 30% to over 700%) for all estimation methods across sample sizes. The MSE values were relatively small for most parameters, but the variance of the intercept had large MSE values (as high as 54) in some conditions.

Table 18. Parameter Estimates for 3-Class 33/33/33 MD=1/1

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	3.2 (0.613)	2.834	3.000	3.000	3.23 (0.682)	2.760	3.000	3.000	3.256 (0.6333)	2.792
	C1		0.333	0.333	0.372	0.335	0.333	0.335	0.331	0.318	0.334	0.333	0.309	0.334
Prop	C2		0.334	0.334	0.355	0.301	0.334	0.332	0.280	0.329	0.333	0.332	0.353	0.331
	C3		0.334	0.333	0.273	0.364	0.333	0.333	0.389	0.353	0.334	0.335	0.338	0.335
Mean I														
C1		31.370	32.530	32.372	32.831	32.853	32.670	32.472	32.943	30.952	32.687	32.523	32.468	32.274
C2		34.290	34.309	34.589	36.924	35.420	34.227	34.561	36.985	33.758	34.221	34.590	35.447	34.447
C3		37.210	36.032	35.886	37.913	36.196	36.000	35.845	39.729	35.985	35.952	35.758	39.292	36.299
Mean S														
C1		1.802	2.585	2.551	1.672	1.996	2.793	2.523	2.047	1.784	2.558	2.500	1.736	1.957
C2		2.834	2.821	2.880	3.983	2.078	2.586	2.882	2.463	2.980	2.810	2.899	2.772	2.878
C3		3.867	3.086	3.058	3.900	3.579	3.115	3.080	4.127	3.549	3.120	3.091	4.645	3.972
Var-I		16.000	17.239	17.341	18.462	23.370	17.403	17.421	15.782	21.167	17.480	17.569	20.163	21.655
Var-S		2.000	2.179	2.175	1.685	3.173	2.202	2.197	1.909	2.928	2.201	2.205	1.703	2.919
Cov		0.300	1.663	1.670	0.701	2.663	1.600	1.595	1.235	2.401	1.551	1.554	0.391	2.596
Residual		0.500	0.522	0.521	0.715	0.480	0.509	0.510	0.653	0.475	0.510	0.509	0.717	0.517

Table 19. Percent Bias for 3-Class 33/33/33 MD=1/1

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	0.767	0.985	**12.727**	1.515	0.897	1.382	0.337	-3.535	1.106	0.897	-6.313	1.178
Prop	C2	1.106	1.067	7.576	-8.758	1.288	0.721	**-15.152**	-0.253	0.827	0.636	6.818	0.337
	C3	1.158	0.979	**-17.273**	**10.273**	0.845	0.927	**17.845**	6.818	1.097	1.497	2.525	1.515
Mean I													
C1	31.370	3.698	3.194	4.657	4.727	4.143	3.512	5.015	-1.332	4.198	3.676	3.501	2.882
C2	34.290	0.056	0.872	7.681	3.294	-0.183	0.790	7.861	-1.550	-0.201	0.876	3.373	0.458
C3	37.210	-3.165	-3.558	1.888	-2.726	-3.253	-3.669	6.771	-3.293	-3.380	-3.903	5.594	-2.448
Mean S													
C1	1.802	**43.441**	**41.543**	-7.194	**10.788**	**54.967**	**40.028**	**13.573**	-0.999	**41.942**	**38.713**	-3.679	8.615
C2	2.834	-0.455	1.606	**40.539**	**-26.670**	-8.761	1.690	**-13.106**	5.161	-0.840	2.279	-2.171	1.561
C3	3.867	**-20.202**	**-20.926**	0.856	-7.453	**-19.460**	**-20.354**	6.734	-8.232	**-19.320**	**-20.059**	**20.119**	2.710
Var-I	16.000	7.743	8.379	**15.386**	**46.061**	8.766	8.882	-1.365	**32.294**	9.247	9.807	**26.020**	**35.345**
Var-S	2.000	8.965	8.760	**-15.772**	**58.631**	**10.080**	9.835	-4.537	**46.382**	**10.050**	**10.245**	**-14.858**	**45.943**
Cov	0.300	**454.233**	**456.800**	**133.674**	**787.633**	**433.167**	**431.767**	**311.799**	**700.413**	**417.000**	**418.133**	**30.471**	**765.421**
Residual	0.500	4.340	4.180	**42.941**	-4.071	1.860	1.900	**30.533**	-4.926	2.000	1.820	**43.339**	3.496

4.3.2.8 3-Class Prop=33/33/33 MD=1/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 33/33/33 MD = 1/3 conditions are presented in Tables 20, 21 and A7. RJMCMC and DP correctly estimated the number of classes. The percentages of selecting a 3-class solution ranged from 65% to 75%, indicating a moderate certainty of the class solution. The class proportions were estimated well by all estimation methods except for RJMCMC in the n = 200 condition and for DP in the n = 600 condition, where the bias levels were slightly higher than 10%.

The intercept parameter means were recovered well by all estimation methods across sample sizes. The slope mean in Class 1 was estimated with slightly high bias (all above 18%) by all estimation methods in the n = 200 condition, and it was improved for RJMCMC and DP as the sample size increased (bias levels decreased to below or slightly higher than 10%). The bias for the slope means in Class 2 and Class 3 was relatively low (below 10%) for all estimation methods, although a few conditions in B-Diff had bias levels that were slightly higher than 10%. The variances of the growth parameters were estimated with low bias level (below 10%) for B-Diff and B-Weak and with high bias (as high as 47%) for RJMCMC and DP across sample sizes. The covariance was recovered poorly by all estimation methods for all sample sizes (bias levels were all above ±60%). The MSE values were small for all parameters except for the intercept variance for DP (around 40).

Table 20. Parameter Estimates for 3-Class 33/33/33 MD=1/3

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	2.963 (0.713)	3.061	3.000	3.000	3.215 (0.648)	3.165	3.000	3.000	2.838 (0.7443)	3.350
	C1		0.335	0.342	0.298	0.335	0.330	0.338	0.324	0.318	0.329	0.334	0.356	0.304
Prop	C2		0.334	0.327	0.364	0.320	0.331	0.323	0.339	0.329	0.330	0.326	0.329	0.303
	C3		0.331	0.331	0.338	0.345	0.339	0.339	0.337	0.353	0.341	0.340	0.315	0.393
Mean I														
C1		31.370	32.766	31.607	29.386	30.854	32.765	31.572	31.943	32.951	32.752	31.569	33.069	32.075
C2		34.290	33.559	35.234	32.955	32.416	32.958	34.328	35.205	34.969	32.885	34.164	32.356	32.451
C3		43.010	42.410	41.986	42.894	41.697	42.833	42.748	44.573	42.072	42.896	42.851	41.579	41.094
Mean S														
C1		1.802	2.311	2.131	2.172	2.268	2.296	2.060	1.927	1.783	2.304	2.035	1.646	2.005
C2		2.834	2.580	2.931	3.027	2.679	2.359	2.653	2.871	2.936	2.327	2.626	2.575	2.476
C3		5.917	5.684	5.539	6.131	6.330	5.849	5.821	6.285	6.105	5.872	5.856	6.046	6.047
Var-I		16.000	16.624	16.627	14.363	17.847	16.392	16.362	16.456	17.484	16.344	16.300	18.880	22.369
Var-S		2.000	2.095	2.093	2.211	2.932	2.065	2.064	2.010	2.276	2.056	2.055	2.358	2.275
Cov		0.300	0.732	0.724	0.097	1.330	0.609	0.605	1.031	0.994	0.576	0.566	0.118	1.046
Residual		0.500	0.520	0.520	0.513	0.510	0.507	0.505	0.469	0.454	0.507	0.506	0.713	0.519

Table 21. Percent Bias for 3-Class 33/33/33 MD=1/3

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	1.445	3.694	-9.697	1.515	0.073	2.406	-1.818	-3.535	-0.200	1.070	7.879	-7.912
Prop	C2	1.197	-0.900	**10.303**	-3.030	0.170	-2.055	2.727	-0.253	-0.033	-1.185	-0.303	-8.148
	C3	0.388	0.236	2.424	4.545	2.788	2.679	2.121	6.818	3.264	3.145	-4.545	**19.091**
Mean I													
C1	31.370	4.451	0.755	-6.323	-1.644	4.448	0.643	1.827	5.040	4.406	0.635	5.417	2.248
C2	34.290	-2.133	2.753	-3.894	-5.466	-3.885	0.112	2.668	1.980	-4.097	-0.367	-5.641	-5.364
C3	43.010	-1.396	-2.380	-0.271	-3.053	-0.411	-0.609	3.635	-2.180	-0.266	-0.370	-3.327	-4.456
Mean S													
C1	1.802	**28.219**	**18.241**	**20.559**	**25.836**	**27.397**	**14.295**	6.923	-1.035	**27.869**	**12.925**	-8.675	**11.281**
C2	2.834	-8.948	3.423	6.806	-5.465	**-16.775**	-6.390	1.291	3.599	**-17.904**	-7.332	-9.154	**-12.640**
C3	5.917	-3.936	-6.393	3.618	6.983	-1.153	-1.629	6.212	3.176	-0.767	-1.033	2.172	2.197
Var-I	16.000	3.898	3.919	**-10.230**	**11.544**	2.452	2.265	2.851	9.274	2.147	1.877	**18.002**	**39.808**
Var-S	2.000	4.770	4.670	**10.540**	**46.599**	3.235	3.220	0.490	**13.821**	2.790	2.755	**17.892**	**13.762**
Cov	0.300	**144.133**	**141.367**	**-67.628**	**343.473**	**102.933**	**101.700**	**243.789**	**231.176**	**91.967**	**88.600**	**-60.750**	**248.537**
Residual	0.500	3.920	4.060	2.500	1.980	1.320	1.080	-6.124	-9.147	1.300	1.120	**42.522**	3.825

4.3.2.9 3-Class Prop=33/33/33 MD=3/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 33/33/33 MD = 3/3 conditions are presented in Tables 22, 23 and A8. RJMCMC and DP correctly estimated the number of classes. The percentages (i.e., 82.9% to 97.3%) of selecting a 3-class solution show higher confidence than the percentages when the MD values were 1/1 and 1/3, which were around 60% to 71%. The class proportions were recovered well by all estimation methods for all sample sizes except for RJMCMC, which yielded a bias level that was slightly higher than 10% when n = 200.

The intercept means were estimated well for all estimation methods for all sample sizes. The slope means in Class 1 and Class 2 were recovered with moderate-to-high bias levels (around -14% to 54%) by B-Diff and were improved as the sample size increased (bias levels decreased to below 1%). RJMCMC and DP yielded relatively high bias levels (around 26% to 32%) for the slope mean in Class 1, and the bias levels decreased to below 10% as the sample size increased. The slope parameter mean in Class 3 was estimated well by all estimation methods. The bias for the growth parameter variances was relatively high for B-Diff and B-Weak (bias levels were above 14%) and decreased to below 10% as the sample size increased. The bias levels for RJMCMC were relatively low across sample sizes (all below or slightly higher than 10%). DP yielded the estimates with inconsistent bias levels (around 3% to 92%) across sample sizes. The covariance parameter had very high bias levels for all estimation methods. The MSE values were relatively small for all parameters.

Table 22. Parameter Estimates for 3-Class 33/33/33 MD=3/3

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	2.964 (0.829)	3.201	3.000	3.000	2.946 (0.970)	3.131	3.000	3.000	2.959 (0.973)	3.270
	C1		0.333	0.341	0.298	0.335	0.336	0.315	0.356	0.308	0.335	0.331	0.311	0.314
Prop	C2		0.329	0.324	0.312	0.322	0.330	0.360	0.308	0.349	0.331	0.340	0.360	0.331
	C3		0.338	0.335	0.390	0.343	0.335	0.325	0.336	0.343	0.334	0.330	0.329	0.355
Mean I														
C1		31.370	34.080	31.735	33.706	32.844	31.616	30.355	32.466	32.954	31.418	31.287	33.675	31.070
C2		40.090	38.179	41.285	41.243	38.228	39.949	40.118	41.055	42.565	40.101	40.115	41.680	38.257
C3		48.810	47.988	47.334	47.525	46.499	48.728	48.002	49.235	50.576	48.773	48.758	48.781	48.643
Mean S														
C1		1.802	2.767	1.942	2.276	2.371	1.885	1.802	2.147	1.981	1.820	1.801	1.646	1.836
C2		4.885	4.201	5.300	5.062	5.227	4.832	4.890	4.957	4.988	4.836	4.890	4.824	4.825
C3		7.968	7.670	7.440	8.054	8.380	7.938	7.729	8.236	8.155	7.951	7.946	8.004	7.698
Var-I		16.000	18.360	18.445	17.497	23.072	16.718	17.406	16.096	16.536	16.398	16.396	17.961	17.692
Var-S		2.000	2.294	2.305	2.142	3.836	2.099	2.184	2.105	2.162	2.060	2.061	2.095	2.381
Cov		0.300	1.032	1.065	0.132	3.368	0.497	0.751	0.856	0.523	0.411	0.414	0.212	0.558
Residual		0.500	0.521	0.520	0.544	0.472	0.508	0.505	0.454	0.470	0.506	0.507	0.521	0.509

Table 23. Percent Bias for 3-Class 33/33/33 MD=3/3

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	0.942	3.336	-9.788	1.515	1.679	-4.488	7.855	-6.566	1.633	0.182	-5.758	-4.882
Prop	C2	-0.297	-1.764	-5.455	-2.424	-0.058	8.964	-6.748	5.808	0.309	2.952	9.091	0.337
	C3	2.385	1.458	**18.273**	3.939	1.409	-1.445	1.924	3.788	1.088	-0.103	-0.303	7.576
Mean I													
C1	31.370	8.639	1.164	7.446	4.697	0.783	-3.236	3.493	5.050	0.152	-0.264	7.347	-0.956
C2	40.090	-4.766	2.980	2.876	-4.645	-0.351	0.069	2.406	6.175	0.028	0.063	3.965	-4.571
C3	48.810	-1.684	-3.025	-2.633	-4.735	-0.169	-1.655	0.870	3.617	-0.076	-0.107	-0.060	-0.342
Mean S													
C1	1.802	**53.546**	7.769	**26.282**	**31.577**	4.606	0.000	**19.129**	9.955	0.988	-0.039	-8.675	1.910
C2	4.885	**-14.008**	8.504	3.632	6.991	-1.089	0.092	1.474	2.118	-1.013	0.106	-1.257	-1.226
C3	7.968	-3.746	-6.628	1.077	5.168	-0.380	-2.999	3.360	2.343	-0.208	-0.281	0.447	-3.389
Var-I	16.000	**14.752**	**15.284**	9.357	**44.198**	4.490	8.789	0.601	3.349	2.489	2.475	**12.257**	**10.575**
Var-S	2.000	**14.685**	**15.235**	7.099	**91.791**	4.965	9.175	5.237	8.091	2.990	3.050	4.746	**19.027**
Cov	0.300	**244.100**	**254.867**	**-56.158**	**1022.698**	**65.533**	**150.433**	**185.441**	**74.435**	**37.133**	**38.067**	**-29.188**	**86.152**
Residual	0.500	4.180	3.980	8.845	-5.625	1.580	1.060	-9.185	-6.100	1.200	1.300	4.155	1.792

4.3.2.10 3-Class Prop=45/45/10 MD=1/1 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 45/45/10 MD = 1/1 conditions are presented in Tables 24, 25 and A9. RJMCMC and DP correctly estimated the number of classes, although the percentage of selecting a 3-class solution by RJMCMC was relatively lower (all below 70%). The class proportions were recovered poorly by B-Diff and B-Weak across all sample sizes where the class size of the minority class was largely overestimated (bias levels were above 224%). RJMCMC also tended to overestimate the class size of the minority class when the sample size was small but the bias level decreased as the sample size increased; however, bias levels were always above 10%. DP yielded slightly high bias levels for the class size of the minority class when n = 200 and n = 400 (bias level were around -16% to 28%), but levels decreased for n = 600 (bias level decreased to 13%).

The intercept means were estimated well by all estimation methods for all sample sizes, although B-Diff slightly underestimated the intercept mean of the minority class (bias levels were all below or slightly higher than 10%). B-Diff and B-Weak consistently underestimated the slope mean in the minority class (bias levels were around -32% to -36%), and it overestimated the slope mean in Class 1 (bias levels were around 30%) across sample sizes. RJMCMC yielded inconsistent bias for the slope means (bias levels were around 2% to 12%). DP produced slightly high bias levels (above 10%) for the slope parameter means, but the bias levels decreased to below 10% when n = 600. The growth parameter variances were recovered well by B-Diff and B-Weak (bias levels were all below 10%) but were recovered relatively poorly by RJMCMC and DP (bias levels were around -9% to 56%). The covariance had very high bias levels (ranging up to 501%) for all estimation methods for all sample sizes. The MSE values are small for most parameters, although RJMCMC and DP yielded a few high MSE values (around 15 to 54) for the intercept variance under some conditions.

Table 24. Parameter Estimates for 3-Class 45/45/10 MD=1/1

			n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes	3.000	3.000	3.000	3.194 (0.680)	2.676	3.000	3.000	3.318 (0.644)	2.861	3.000	3.000	2.752 (0.678)	2.825
Prop C1		0.339	0.340	0.494	0.444	0.342	0.342	0.480	0.433	0.342	0.344	0.477	0.439
Prop C2		0.334	0.328	0.387	0.428	0.331	0.326	0.405	0.483	0.334	0.326	0.410	0.467
Prop C3		0.328	0.332	0.119	0.129	0.327	0.332	0.115	0.084	0.324	0.330	0.113	0.095
Mean I													
C1	31.370	31.720	31.592	30.468	31.581	31.824	31.742	33.136	32.122	31.857	31.802	29.458	33.304
C2	34.290	34.902	33.556	35.357	32.601	33.353	33.496	34.357	33.206	33.325	33.534	36.047	34.045
C3	37.210	33.289	34.763	38.946	38.637	34.753	34.703	36.486	35.934	34.757	34.621	36.357	38.839
Mean S													
C1	1.802	2.352	2.331	1.625	1.987	2.351	2.329	1.714	1.955	2.341	2.334	1.724	1.931
C2	2.834	2.620	2.516	3.147	3.187	2.612	2.524	2.773	2.909	2.627	2.507	2.635	2.818
C3	3.867	2.471	2.596	4.152	4.271	2.485	2.594	4.053	4.125	2.492	2.620	3.147	3.490
Var-I	16.000	15.977	16.088	18.217	21.839	16.249	16.245	13.215	19.014	16.401	16.493	19.842	21.125
Var-S	2.000	2.015	2.017	1.767	2.577	2.062	2.059	3.117	2.898	2.086	2.086	1.823	2.395
Cov	0.300	1.252	1.268	0.442	1.806	1.274	1.265	1.369	1.789	1.248	1.265	0.239	1.352
Residual	0.500	0.522	0.521	0.704	0.541	0.509	0.508	0.672	0.455	0.504	0.505	0.708	0.507

Table 25. Percent Bias for 3-Class 45/45/10 MD=1/1

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-24.744**	**-24.511**	9.716	-1.444	**-24.053**	**-23.929**	6.667	-3.756	**-23.987**	**-23.587**	6.044	-2.556
Prop	C2	**-25.876**	**-27.011**	**-14.044**	-4.956	**-26.344**	**-27.649**	-10.000	7.356	**-25.851**	**-27.562**	-8.933	3.778
	C3	**227.790**	**231.850**	**19.480**	**28.800**	**226.790**	**232.100**	**15.000**	**-16.200**	**224.270**	**230.170**	**13.000**	-5.500
Mean I													
C1	31.370	1.115	0.707	-2.874	0.672	1.449	1.186	5.629	2.396	1.551	1.376	-6.096	6.164
C2	34.290	1.784	-2.141	3.113	-4.925	-2.734	-2.316	0.196	-3.162	-2.814	-2.205	5.123	-0.714
C3	37.210	**-10.537**	-6.576	4.666	3.834	-6.603	-6.736	-1.947	-3.429	-6.594	-6.959	-2.292	4.377
Mean S													
C1	1.802	**30.544**	**29.345**	-9.839	**10.286**	**30.461**	**29.245**	-4.907	8.467	**29.900**	**29.495**	-4.352	7.179
C2	2.834	-7.562	**-11.221**	**11.039**	**12.445**	-7.823	**-10.946**	-2.136	2.650	-7.322	**-11.524**	-7.039	-0.564
C3	3.867	**-36.108**	**-32.881**	7.382	**10.438**	**-35.738**	**-32.925**	4.810	6.661	**-35.552**	**-32.258**	**-18.623**	-9.758
Var-I	16.000	-0.145	0.550	**13.856**	**36.492**	1.557	1.533	**-17.404**	**18.837**	2.505	3.082	**24.011**	**32.034**
Var-S	2.000	0.760	0.835	**-11.662**	**28.843**	3.075	2.950	**55.848**	**44.890**	4.290	4.305	-8.842	**19.748**
Cov	0.300	**317.433**	**322.800**	**47.354**	**501.902**	**324.567**	**321.533**	**356.310**	**496.167**	**316.067**	**321.600**	**-20.296**	**350.718**
Residual	0.500	4.320	4.180	**40.753**	8.101	1.780	1.640	**34.455**	-8.923	0.820	0.900	**41.600**	1.347

4.3.2.11 3-Class Prop=45/45/10 MD=1/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 45/45/10 MD = 1/3 conditions are presented in Tables 26, 27 and A10. The number of classes estimated by RJMCMC and DP was correct for all sample sizes. The percentage of selecting a 3-class solution by RJMCMC ranged around 65% - 75%, which suggests moderate certainty about the 3-class solution. The class size in the minority class was overestimated by all estimation methods (bias levels were all above 10%) except for RJMCMC, which yielded a relatively low bias level (i.e., -5.57%). The bias for the class proportions decreased as the sample size increased, especially for B-Diff and B-Weak, whose bias levels decrease to below 10%.

The intercept means were recovered well for all estimation methods across sample sizes. B-Diff and B-Weak yielded relatively higher bias levels (above 23%) for the slope means in the minority class and Class 1, compared with RJMCMC and DP (bias levels were below 18%) for the same parameters. B-Diff and B-Weak recovered the growth parameter variances well (bias levels were all below 10%), and RJMCMC and DP recovered them relatively poorly (bias levels ranging up to 37%). All estimation methods yielded high bias levels (ranging above 400%) for the covariance parameter. The MSE values were small for all parameters except for the intercept variance under DP (which was around 18).

Table 26. Parameter Estimates for 3-Class 45/45/10 MD=1/3

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	2.596 (0.750)	3.324	3.000	3.000	3.14 (0.642)	2.977	3.000	3.000	2.783 (0.758)	3.308
Prop	C1		0.429	0.428	0.467	0.503	0.436	0.438	0.430	0.446	0.439	0.436	0.420	0.475
	C2		0.391	0.408	0.439	0.418	0.449	0.447	0.458	0.429	0.452	0.454	0.468	0.413
	C3		0.181	0.164	0.094	0.078	0.116	0.115	0.112	0.125	0.110	0.110	0.112	0.111
Mean I														
C1		31.370	31.679	31.587	30.485	29.278	31.638	31.635	33.357	33.123	31.681	31.681	33.357	33.299
C2		34.290	34.847	34.439	32.536	31.590	33.796	33.803	34.236	34.082	33.788	33.790	33.466	35.034
C3		43.010	40.374	40.892	41.835	45.000	42.256	42.291	41.573	41.716	42.469	42.473	43.357	42.210
Mean S														
C1		1.802	2.289	2.289	1.984	1.888	2.251	2.248	2.076	2.114	2.226	2.225	1.885	2.032
C2		2.834	2.674	2.492	3.047	3.090	2.302	2.304	2.906	2.968	2.336	2.339	2.946	2.572
C3		5.917	4.970	5.153	6.205	6.201	5.656	5.660	6.125	6.183	5.726	5.726	5.536	5.517
Var-I		16.000	16.825	16.793	15.072	17.317	16.590	16.556	13.983	16.351	16.594	16.602	18.353	20.325
Var-S		2.000	2.116	2.113	2.519	2.624	2.084	2.086	2.539	2.746	2.092	2.090	1.937	2.441
Cov		0.300	1.070	1.057	0.208	1.504	0.901	0.898	0.518	1.360	0.871	0.873	0.198	1.073
Residual		0.500	0.523	0.521	0.697	0.495	0.509	0.510	0.574	0.472	0.507	0.506	0.727	0.518

Table 27. Percent Bias for 3-Class 45/45/10 MD=1/3

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	-4.722	-4.911	3.727	**11.867**	-3.184	-2.611	-4.460	-0.989	-2.480	-3.129	-6.778	5.580
Prop	C2	**-13.180**	-9.336	-2.489	-7.044	-0.269	-0.711	1.778	-4.591	0.347	0.978	4.044	-8.133
	C3	**80.560**	**64.110**	-5.570	**-21.700**	**15.540**	**14.950**	**12.070**	**25.110**	9.600	9.680	**12.300**	**11.490**
Mean I													
C1	31.370	0.986	0.691	-2.823	-6.668	0.856	0.844	6.335	5.588	0.991	0.990	6.335	6.151
C2	34.290	1.623	0.435	-5.116	-7.874	-1.440	-1.420	-0.158	-0.606	-1.464	-1.459	-2.404	2.170
C3	43.010	-6.130	-4.925	-2.733	4.627	-1.753	-1.673	-3.340	-3.008	-1.257	-1.249	0.807	-1.861
Mean S													
C1	1.802	**26.998**	**27.026**	**10.086**	4.764	**24.917**	**24.761**	**15.183**	**17.309**	**23.535**	**23.496**	4.584	**12.787**
C2	2.834	-5.635	**-12.082**	7.510	9.038	**-18.776**	**-18.705**	2.535	4.717	**-17.558**	**-17.470**	3.940	-9.251
C3	5.917	**-16.008**	**-12.912**	4.862	4.797	-4.409	-4.348	3.509	4.496	-3.223	-3.225	-6.444	-6.762
Var-I	16.000	5.153	4.955	-5.799	8.234	3.690	3.478	**-12.603**	2.195	3.714	3.760	**14.704**	**27.032**
Var-S	2.000	5.775	5.625	**25.975**	**31.178**	4.215	4.295	**26.953**	**37.324**	4.575	4.515	-3.162	**22.074**
Cov	0.300	**256.700**	**252.200**	**-30.531**	**401.293**	**200.167**	**199.233**	**72.667**	**353.431**	**190.233**	**191.100**	**-34.107**	**257.619**
Residual	0.500	4.500	4.280	**39.360**	-1.085	1.840	1.940	**14.865**	-5.529	1.380	1.220	**45.361**	3.512

4.3.2.12 3-Class Prop=45/45/10 MD=3/1 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 45/45/10 MD = 3/1 conditions are presented in Tables 28, 29 and A11. RJMCMC and DP correctly estimated the number of classes. The percentages of selecting a 3-class solution by RJMCMC were around 74% to 78%. The class size in the minority class was consistently overestimated by B-Diff, B-Weak, and DP (bias levels were above 200%) and underestimated by RJMCMC (bias levels were -19% to -24%) across sample sizes. Both B-Diff and B-Weak yielded high levels of bias (above 200%) for the class proportions when n = 200.

The intercept mean in the minority class were estimated well by all estimation methods. B-Diff and B-Weak produced high levels of bias (above 10%) for the slope means in the minority class and in Class 1; the bias levels for Class 1 decreased to below 10% as the sample size increased. The slope means were recovered well by RJMCMC and DP. The growth parameter variances were recovered well by B-Diff and were recovered with high bias levels (around 33%) by B-Weak when n = 400. Both RJMCMC and DP yielded high bias levels (around 10% to 85%) for the growth parameter variances when n = 200; the bias levels decreased slightly (around 7% to 34%) when the sample size increased. All estimation methods produced very high bias levels (ranging -23% to over 1200%) for the covariance parameter. DP yielded a few high MSE values (i.e., 159.517) for the variance parameters when n = 200, and other parameters all had relatively small MSE values.

Table 28. Parameter Estimates for 3-Class 45/45/10 MD=3/1

		Pop	n=200 B-Diff	n=200 B-Weak	n=200 RJMCMC	n=200 DP	n=400 B-Diff	n=400 B-Weak	n=400 RJMCMC	n=400 DP	n=600 B-Diff	n=600 B-Weak	n=600 RJMCMC	n=600 DP
# classes		3.000	3.000	3.000	2.599 (0.741)	3.133	3.000	3.000	3.115 (0.782)	3.108	3.000	3.000	2.955 (0.737)	3.248
Prop	C1		0.356	0.377	0.432	0.396	0.387	0.401	0.421	0.438	0.411	0.420	0.439	0.416
	C2		0.328	0.307	0.492	0.476	0.302	0.295	0.498	0.437	0.299	0.286	0.478	0.468
	C3		0.316	0.316	0.076	0.129	0.311	0.303	0.081	0.125	0.290	0.295	0.083	0.116
Mean I														
C1		31.370	32.446	31.742	32.247	31.286	31.902	31.572	33.356	32.733	31.573	31.448	28.468	32.049
C2		40.090	37.658	38.971	39.468	38.394	38.858	39.594	38.875	38.996	39.997	39.997	39.364	38.844
C3		43.010	40.088	39.860	45.798	43.929	39.976	39.877	42.076	42.032	39.648	39.996	45.212	44.141
Mean S														
C1		1.802	2.252	1.997	1.835	1.885	2.025	1.905	1.985	2.110	1.895	1.855	1.864	1.830
C2		4.885	4.814	4.446	5.148	5.140	4.806	4.698	4.975	5.292	4.729	4.830	5.024	5.019
C3		5.917	4.019	4.768	5.236	6.226	4.439	4.777	6.285	6.177	4.821	4.842	5.524	5.542
Var-I		16.000	16.453	16.466	17.744	28.630	15.959	21.184	14.877	18.245	15.983	16.008	18.686	18.932
Var-S		2.000	2.060	2.062	2.533	3.707	2.011	2.676	1.672	2.765	2.007	2.005	1.768	2.465
Cov		0.300	0.663	0.661	0.209	4.008	0.460	0.462	0.523	1.156	0.436	0.433	0.229	0.828
Residual		0.500	0.523	0.521	0.529	0.475	0.507	0.509	0.608	0.483	0.504	0.506	0.524	0.509

Table 29. Percent Bias for 3-Class 45/45/10 MD=3/1

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-20.942**	**-16.289**	-4.000	**-12.067**	**-14.060**	**-10.829**	-6.444	-2.569	-8.693	-6.773	-2.549	-7.460
Prop	C2	**-27.109**	**-31.776**	9.333	5.689	**-32.811**	**-34.378**	**10.667**	-2.927	**-33.624**	**-36.527**	6.311	3.933
	C3	**216.230**	**216.290**	**-24.000**	**28.700**	**210.920**	**203.430**	**-19.000**	**24.730**	**190.430**	**194.850**	**-16.930**	**15.870**
Mean I													
C1	31.370	3.431	1.185	2.795	-0.267	1.697	0.643	6.332	4.345	0.648	0.248	-9.250	2.165
C2	40.090	-6.068	-2.791	-1.551	-4.231	-3.074	-1.236	-3.032	-2.730	-0.231	-0.231	-1.811	-3.108
C3	43.010	-6.793	-7.325	6.482	2.138	-7.053	-7.285	-2.171	-2.274	-7.816	-7.008	5.120	2.629
Mean S													
C1	1.802	**24.961**	**10.810**	1.807	4.591	**12.353**	5.716	**10.133**	**17.094**	5.172	2.919	3.457	1.555
C2	4.885	-1.455	-8.979	5.374	5.226	-1.609	-3.820	1.849	8.339	-3.204	-1.128	2.837	2.752
C3	5.917	**-32.079**	**-19.417**	**-11.514**	5.221	**-24.979**	**-19.267**	6.214	4.393	**-18.521**	**-18.163**	-6.649	-6.342
Var-I	16.000	2.829	2.911	**10.898**	**78.939**	-0.256	**32.397**	-7.020	**14.034**	-0.109	0.048	**16.784**	**18.328**
Var-S	2.000	3.015	3.090	**26.637**	**85.327**	0.540	**33.780**	**-16.380**	**38.250**	0.325	0.245	**-11.596**	**23.272**
Cov	0.300	**120.933**	**120.333**	**-30.496**	**1236.163**	**53.200**	**53.900**	74.378	**285.239**	**45.267**	**44.167**	**-23.643**	**176.070**
Residual	0.500	4.540	4.120	5.847	-5.043	1.320	1.800	**21.530**	-3.349	0.860	1.220	4.766	1.810

4.3.2.13 3-Class Prop=45/45/10 MD=3/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 45/45/10 MD = 3/3 conditions are presented in Tables 30, 31 and A12. RJMCMC and DP correctly estimated the number of classes. The percentages of selecting a 3-class solution by RJMCMC were relatively high (i.e., 96.7% to 99.5%) across sample sizes, compared with the percentages in other conditions with prop = 45/45/10 (around 64% to 78%). The class sizes of the minority class and Class 2 were estimated inaccurately by B-Diff and B-Weak when n = 200 and n = 400 (bias levels were mostly above 10%) and the bias levels decreased to below 10% when n = 600. RJMCMC and DP yielded some levels of bias (around 13% to 25%) for the class size of the minority class, but improved as the sample size increased (bias levels decreased to below 10% when n = 600).

The intercept means were estimated well by all estimation methods; the bias levels for B-Diff and B-Weak in the n = 200 condition were slightly high (around -11% to -15%), but they decreased to below 10% as the sample size increased. RJMCMC and DP correctly recovered the intercept means. B-Diff and B-Weak yielded high bias levels (around -26% to 46%) for the slope means in the minority class and Class 1 when n = 200; the bias levels decreased to below or slightly higher than 10% as the sample size increased. RJMCMC and DP produced relatively low bias levels (all below or slightly higher than 10%) for the slope parameter means. B-Diff and B-Weak yielded high bias levels (all above 54%) for the growth parameter variances when n = 200 and improved as the sample size increased (bias levels decreased to below or slightly higher than 10%). RJMCMC estimated the growth parameter variances well (bias levels were all below or slightly higher than 10%) and DP had more inaccuracy (bias levels were mostly around 23% to 184%) across sample sizes. The covariance parameter was recovered with very high bias levels (mostly around 34% to over 3100%) by all estimation methods. Most parameters had small MSE values except for the growth parameter variances (around 76 to 261).

Table 30. Parameter Estimates for 3-Class 45/45/10 MD=3/3

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes	3.000	3.000	3.000	2.981 (0.967)	3.294	3.000	3.000	2.981 (0.995)	3.138	3.000	3.000	2.982 (0.973)	3.044
Prop	C1	0.391	0.405	0.433	0.439	0.441	0.443	0.458	0.438	0.447	0.448	0.434	0.451
	C2	0.313	0.343	0.442	0.448	0.404	0.425	0.421	0.454	0.443	0.448	0.462	0.439
	C3	0.296	0.252	0.125	0.113	0.155	0.132	0.121	0.108	0.110	0.104	0.104	0.111
Mean I													
C1	31.370	33.433	32.445	29.357	32.285	31.837	31.524	32.536	33.132	31.418	31.383	32.537	30.304
C2	40.090	39.936	39.771	38.345	38.391	40.375	39.920	39.225	39.199	40.059	39.961	38.575	38.828
C3	48.810	41.424	43.031	50.468	50.757	46.629	47.509	46.246	47.513	48.228	48.400	50.467	50.012
Mean S													
C1	1.802	2.637	2.339	1.546	1.986	1.983	1.874	1.736	1.951	1.821	1.801	1.847	1.841
C2	4.885	5.258	4.668	5.247	5.141	4.962	4.800	4.724	4.822	4.866	4.875	5.047	4.925
C3	7.968	4.847	5.885	8.236	8.267	7.197	7.507	8.236	8.234	7.761	7.974	7.525	7.266
Var-I	16.000	25.123	24.758	15.902	24.013	17.637	17.453	17.285	32.179	16.396	16.440	16.990	16.697
Var-S	2.000	3.150	3.103	1.761	5.698	2.210	2.194	1.836	4.350	2.059	2.062	2.040	2.466
Cov	0.300	3.895	3.749	0.281	9.679	0.886	0.820	2.133	6.136	0.422	0.429	0.197	0.357
Residual	0.500	0.526	0.524	0.614	0.472	0.508	0.506	0.513	0.471	0.506	0.506	0.613	0.501

Table 31. Percent Bias for 3-Class 45/45/10 MD=3/3

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-13.009**	-9.896	-3.711	-2.444	-2.000	-1.642	1.778	-2.593	-0.620	-0.464	-3.578	0.156
Prop	C2	**-30.456**	**-23.822**	-1.867	-0.533	**-10.264**	-5.502	-6.444	0.926	-1.553	-0.376	2.667	-2.556
	C3	**195.590**	**151.730**	**25.100**	**13.400**	**55.190**	**32.150**	**21.000**	7.500	9.780	3.780	4.100	**10.800**
Mean I													
C1	31.370	6.577	3.426	-6.416	2.917	1.490	0.492	3.717	5.617	0.154	0.042	3.721	-3.399
C2	40.090	-0.384	-0.795	-4.352	-4.239	0.711	-0.425	-2.158	-2.223	-0.077	-0.322	-3.780	-3.148
C3	48.810	**-15.133**	**-11.839**	3.397	3.989	-4.469	-2.665	-5.253	-2.657	-1.193	-0.840	3.396	2.463
Mean S													
C1	1.802	**46.310**	**29.817**	**-14.227**	**10.193**	**10.039**	3.973	-3.678	8.266	1.049	-0.033	2.484	2.146
C2	4.885	7.627	-4.442	7.415	5.232	1.570	-1.746	-3.289	-1.299	-0.381	-0.205	3.316	0.822
C3	7.968	**-39.174**	**-26.141**	3.362	3.751	-9.680	-5.789	3.369	3.337	-2.598	0.078	-5.566	-8.805
Var-I	16.000	**57.018**	**54.735**	-0.610	**50.080**	**10.232**	9.081	8.033	**101.121**	2.473	2.753	6.185	4.357
Var-S	2.000	**57.505**	**55.155**	**-11.953**	**184.877**	**10.520**	9.690	-8.178	**117.486**	2.965	3.090	1.987	**23.282**
Cov	0.300	**1198.400**	**1149.800**	-6.365	**3126.469**	**195.233**	**173.267**	**610.877**	**1945.467**	**40.533**	**42.933**	**-34.169**	**19.009**
Residual	0.500	5.160	4.740	**22.768**	-5.682	1.540	1.120	2.683	-5.817	1.240	1.120	**22.505**	0.222

4.3.2.14 3-Class Prop=70/20/10 MD=1/1 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 70/20/10 MD = 1/1 conditions are presented in Tables 32, 33 and A13. RJMCMC and DP correctly estimated the number of classes. The percentage of selecting a 3-class solution by RJMCMC was higher when n = 600 (i.e., 86.7%) than when the sample sizes were smaller (i.e., below 65%). The class proportions were estimated poorly by all estimation methods across sample sizes (bias levels were mostly above 10%). B-Diff and B-Weak tended to overestimate the class size in the minority class (bias levels were around -47% to 230%), while underestimating it in the majority class (bias levels were around 47% to 50%). The bias for the minority class size under DP (bias levels were below or slightly higher than 10%) decreased when the sample size increased, but it was consistently high for B-Diff and B-Weak across all sample sizes (bias levels were all above -46%). RJMCMC tended to underestimate the class size in the minority class (bias levels were all above -37%), and it overestimated the majority class slightly (bias levels were slightly higher than 10%).

The intercept means were recovered well by all estimation methods for all sample sizes. B-Diff and B-Weak yielded relatively high bias level (around 12% to -36%) for the slope means, while RJMCMC and DP estimated these parameters with relatively low bias levels (below or slightly higher than 10%). Overall, B-Diff and B-Weak estimated the growth parameter variances well except for B-Weak, which produced slightly higher bias levels (around 23%) when n = 200. RJMCMC and DP consistently yielded relatively high bias levels (around 11% to 44%) for the variance parameters across all sample sizes. All estimation methods yielded high bias (around 15% to 425%) for the covariance parameter. The MSE values are small for most parameters. The variance of the intercept had relatively high MSE values (around 10 to 24) in some conditions.

Table 32. Parameter Estimates for 3-Class 70/20/10 MD=1/1

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	3.338 (0.649)	2.656	3.000	3.000	3.392 (0.621)	2.656	3.000	3.000	2.721 (0.867)	3.309
	C1		0.351	0.352	0.788	0.627	0.358	0.362	0.812	0.636	0.369	0.371	0.758	0.653
Prop	C2		0.334	0.317	0.174	0.232	0.338	0.312	0.177	0.238	0.344	0.315	0.180	0.254
	C3		0.315	0.331	0.039	0.142	0.304	0.326	0.011	0.125	0.287	0.314	0.063	0.093
Mean I														
C1		31.370	31.021	30.932	32.048	30.054	31.165	31.046	31.996	30.899	31.222	31.041	29.468	32.292
C2		34.290	32.572	32.872	33.357	32.722	32.586	32.994	32.735	34.932	32.490	33.409	32.780	33.840
C3		37.210	34.365	34.180	36.358	39.143	34.439	34.147	33.547	35.929	34.727	33.989	38.736	38.744
Mean S														
C1		1.802	2.074	2.076	2.046	1.971	2.223	2.055	2.059	1.907	2.203	2.024	1.784	1.919
C2		2.834	2.236	2.278	2.954	2.962	2.041	2.325	2.704	2.696	1.996	2.425	3.094	2.926
C3		3.867	2.443	2.400	4.135	4.167	2.567	2.465	4.237	4.125	2.728	2.470	3.249	3.486
Var-I		16.000	15.903	19.781	12.378	18.876	15.893	15.908	13.858	20.418	15.760	15.831	19.189	20.903
Var-S		2.000	2.007	2.495	1.761	2.723	2.007	2.011	2.897	2.647	1.996	2.006	1.857	2.579
Cov		0.300	1.173	1.576	1.181	1.187	1.039	1.472	0.682	1.757	0.905	0.919	0.253	1.383
Residual		0.500	0.523	0.521	0.727	0.470	0.509	0.510	0.479	0.461	0.507	0.507	0.713	0.509

Table 33. Percent Bias for 3-Class 70/20/10 MD=1/1

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-49.826**	**-49.701**	**12.529**	**-10.476**	**-48.859**	**-48.294**	**16.000**	-9.096	**-47.336**	**-46.997**	8.238	-6.714
Prop	C2	**67.020**	**58.705**	**-13.250**	**15.900**	**69.145**	**56.180**	**-11.300**	**19.160**	**72.165**	**57.400**	**-10.250**	**27.000**
	C3	**214.740**	**230.500**	**-61.200**	**41.530**	**203.720**	**225.700**	**-89.400**	**25.350**	**187.020**	**214.180**	**-37.167**	-7.000
Mean I													
C1	31.370	-1.112	-1.396	2.160	-4.194	-0.653	-1.032	1.995	-1.501	-0.473	-1.048	-6.062	2.938
C2	34.290	-5.010	-4.135	-2.720	-4.573	-4.971	-3.780	-4.536	1.872	-5.250	-2.570	-4.403	-1.311
C3	37.210	-7.646	-8.144	-2.290	5.196	-7.447	-8.231	-9.845	-3.442	-6.674	-8.657	4.100	4.123
Mean S													
C1	1.802	**15.089**	**15.228**	**13.526**	9.378	**23.352**	**14.018**	**14.240**	5.852	**22.259**	**12.342**	-1.027	6.480
C2	2.834	**-21.090**	**-19.633**	4.225	4.518	**-27.985**	**-17.960**	-4.605	-4.855	**-29.587**	**-14.450**	9.185	3.252
C3	3.867	**-36.827**	**-37.934**	6.921	7.746	**-33.618**	**-36.261**	9.563	6.682	**-29.452**	**-36.137**	**-15.985**	-9.846
Var-I	16.000	-0.607	**23.629**	**-22.636**	**17.975**	-0.670	-0.577	**-13.387**	**27.613**	-1.500	-1.056	**19.928**	**30.646**
Var-S	2.000	0.330	**24.760**	**-11.953**	**36.170**	0.360	0.525	**44.835**	**32.345**	-0.180	0.315	-7.145	**28.941**
Cov	0.300	**290.867**	**425.333**	**293.635**	**295.829**	**246.167**	**390.700**	**127.467**	**485.725**	**201.700**	**206.333**	**-15.700**	**361.096**
Residual	0.500	4.520	4.180	**45.343**	-6.044	1.780	1.900	-4.140	-7.734	1.340	1.360	**42.508**	1.716

4.3.2.15 3-Class Prop=70/20/10 MD=1/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 70/20/10 MD = 1/3 conditions are presented in Tables 34, 35 and A14. RJMCMC and DP correctly estimated the number of classes. The percentage of selecting a 3-class solution by RJMCMC (around 63% to 79%) indicates a moderate confidence of the 3-class solution. The bias for the class proportions yielded by B-Diff and B-Weak were relatively high (mostly above 22%). The recovery of the class size of the minority class by B-Diff improved as the sample size increased (bias levels decreased below 10%) while the bias for the class sizes in the other two classes were consistently high. B-Weak produced high bias levels (all above 11%) for the class proportions for all sample sizes. RJMCMC consistently underestimated the class size of the minority class (bias levels were above -30%), while DP consistently overestimated the same parameter (bias levels were above 40%).

The intercept means were estimated accurately by all estimation methods for all sample sizes. The slope means in Class 2 were underestimated by B-Diff and B-Weak (bias levels were around -28%). RJMCMC also yielded slightly higher bias levels (above 10%) when $n = 400$ and $n = 600$. B-Diff and B-Weak recovered the growth parameter variances well (bias levels were below 10%), and RJMCMC and DP yielded some higher bias levels (above 10%). The covariance parameter was recovered with high bias (around 110% to 159%) by all estimation methods. The MSE values were quite small for all of the parameters.

Table 34. Parameter Estimates for 3-Class 70/20/10 MD=1/3

	Pop	n=200				n=400				n=600			
		B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes	3.000	3.000	3.000	2.756 (0.794)	3.287	3.000	3.000	2.845 (0.724)	3.333	3.000	3.000	2.884 (0.634)	2.862
Prop C1		0.441	0.443	0.728	0.621	0.447	0.446	0.720	0.632	0.452	0.443	0.745	0.649
Prop C2		0.433	0.434	0.217	0.224	0.443	0.442	0.233	0.218	0.440	0.434	0.186	0.210
Prop C3		0.126	0.123	0.055	0.156	0.109	0.112	0.047	0.150	0.107	0.123	0.070	0.141
Mean I													
C1	31.370	30.773	30.764	29.846	29.054	30.897	30.891	33.268	32.201	30.940	30.764	31.066	30.287
C2	34.290	33.254	33.177	33.246	32.722	32.990	33.003	35.027	32.937	32.986	33.177	35.536	32.829
C3	43.010	42.036	42.155	41.936	45.015	42.574	42.581	42.794	42.509	42.659	42.155	42.248	45.208
Mean S													
C1	1.802	2.003	2.004	1.936	1.972	1.985	1.977	2.105	1.906	1.963	2.004	2.182	1.840
C2	2.834	2.051	2.018	3.074	2.968	2.015	2.023	2.376	2.694	2.052	2.018	3.146	2.930
C3	5.917	5.568	5.611	6.236	6.197	5.761	5.761	6.174	6.287	5.790	5.611	5.521	5.518
Var-I	16.000	15.892	15.892	14.726	15.921	15.915	15.893	14.302	18.845	15.935	15.892	17.534	17.888
Var-S	2.000	2.004	2.000	2.618	3.092	2.008	2.006	2.172	2.858	2.009	2.000	2.345	2.538
Cov	0.300	0.697	0.696	0.779	0.771	0.648	0.639	0.713	1.065	0.631	0.696	0.671	0.764
Residual	0.500	0.520	0.522	0.537	0.478	0.508	0.508	0.704	0.475	0.506	0.522	0.525	0.507

Table 35. Percent Bias for 3-Class 70/20/10 MD=1/3

	Pop	n=200				n=400				n=600			
		B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-36.989**	**-36.740**	4.000	**-11.314**	**-36.106**	**-36.296**	2.886	-9.684	**-35.417**	**-36.740**	6.357	-7.301
Prop	C2	**116.365**	**117.180**	8.500	**11.800**	**121.725**	**121.220**	**16.500**	8.890	**120.245**	**117.180**	-7.050	5.030
	C3	**26.190**	**22.820**	**-45.000**	**55.600**	9.290	**11.630**	**-53.200**	**50.010**	7.430	**22.820**	**-30.400**	**41.050**
Mean I													
C1	31.370	-1.902	-1.931	-4.858	-7.382	-1.509	-1.526	6.050	2.650	-1.372	-1.931	-0.968	-3.453
C2	34.290	-3.022	-3.245	-3.044	-4.573	-3.790	-3.753	2.149	-3.945	-3.803	-3.245	3.634	-4.260
C3	43.010	-2.264	-1.987	-2.498	4.661	-1.014	-0.998	-0.503	-1.164	-0.817	-1.987	-1.771	5.109
Mean S													
C1	1.802	**11.138**	**11.204**	7.421	9.415	**10.161**	9.734	**16.831**	5.777	8.957	**11.204**	**21.069**	2.100
C2	2.834	**-27.625**	**-28.807**	8.451	4.726	**-28.913**	**-28.603**	**-16.161**	-4.924	**-27.608**	**-28.807**	**10.996**	3.386
C3	5.917	-5.895	-5.168	5.384	4.736	-2.633	-2.636	4.336	6.258	-2.153	-5.168	-6.688	-6.743
Var-I	16.000	-0.678	-0.676	-7.964	-0.495	-0.530	-0.669	**-10.615**	**17.782**	-0.404	-0.676	9.585	**11.802**
Var-S	2.000	0.195	0.000	**30.900**	**54.586**	0.380	0.315	8.589	**42.900**	0.470	0.000	**17.265**	**26.881**
Cov	0.300	**132.167**	**131.867**	**159.700**	**157.109**	**116.133**	**112.900**	**137.635**	**254.983**	**110.300**	**131.867**	**123.634**	**154.731**
Residual	0.500	4.080	4.380	7.340	-4.415	1.540	1.560	**40.900**	-4.992	1.220	4.380	5.080	1.456

4.3.2.16 3-Class Prop=70/20/10 MD=3/1 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 70/20/10 MD = 3/1 conditions are presented in Tables 36, 37 and A15. The number of classes was estimated correctly by RJMCMC for all sample sizes. However, DP yielded an estimate that was rounded to 2 (i.e., 2.448) for the number of classes in the n = 200 condition. This indicates that DP selected a 2-class solution when the sample size was 200, which is not accurate. It appears that DP was unable to find one of the minority classes in this condition. The percentages of the selecting a 3-class solution by RJMCMC suggest moderate certainty (ranging between 75% and 82.6%). The class proportions were recovered inaccurately by B-Diff and B-Weak, which consistently overestimated the class size of the smallest minority class (bias levels were around 200%) and underestimated the class size of the majority class (bias levels were around -46%). RJMCMC tended to underestimate the class size of the minority class (bias level were around -20% to -56%) while producing low bias levels for majority class size (bias levels were all below 10%). DP was not able to yield accurate estimates for the class proportions for n = 200, and it overestimated the class size of the minority class when n = 400 and n = 600 (bias levels were over 30%).

The intercept means were estimated well by all estimation methods, although B-Diff, B-Weak, and RJMCMC yielded some bias levels that were slightly higher than 10%. DP was not able to provide the estimate for the intercept mean in the third class because it could not find that small minority class. B-Diff and B-Weak yielded relatively high bias levels (around -26% to -40%) for the slope parameter means in Class 2 and the smallest minority class (Class 3). RJMCMC produced low bias levels (all below or slightly higher than 10%) for all the slope parameter means. DP failed to provide an estimate of the slope mean in the third class for the same reasons stated above. The covariance parameter was recovered with very high bias levels (around 23% to 159%) by all estimation methods. The MSE values were relatively high for some intercept means (around 18 to 50) and low for other parameters (below 10).

Table 36. Parameter Estimates and Percent Bias for 3-Class 70/20/10 MD=3/1

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	2.737 (0.787)	2.448	3.000	3.000	3.205 (0.826)	3.263	3.000	3.000	2.894 (0.713)	3.327
	C1		0.371	0.377	0.724	0.684	0.375	0.376	0.729	0.654	0.375	0.380	0.713	0.677
Prop	C2		0.328	0.326	0.233	0.316	0.325	0.325	0.220	0.215	0.324	0.321	0.207	0.189
	C3		0.300	0.297	0.043	-	0.300	0.298	0.051	0.131	0.301	0.298	0.080	0.134
Mean I														
C1		31.370	30.809	30.706	33.469	34.054	30.879	30.801	27.984	32.004	30.922	30.875	29.253	32.292
C2		40.090	35.580	35.430	40.345	42.601	35.316	35.198	39.789	40.719	35.264	35.280	39.980	38.646
C3		43.010	38.271	38.652	50.096	-	38.458	38.687	41.953	41.726	38.522	38.629	40.264	45.142
Mean S														
C1		1.802	1.867	1.849	1.690	1.271	1.853	1.841	1.952	1.805	1.856	1.844	1.754	1.878
C2		4.885	3.014	2.957	5.155	3.013	2.963	2.917	5.174	5.251	2.956	2.966	5.169	4.976
C3		5.917	4.243	4.350	6.299	-	4.305	4.365	6.285	6.180	4.331	4.350	5.250	5.542
Var-I		16.000	16.135	16.103	14.726	18.288	16.022	16.007	17.893	15.175	16.052	16.033	18.936	19.464
Var-S		2.000	2.030	2.028	2.618	2.678	2.014	2.015	2.214	2.301	2.016	2.016	2.331	2.299
Cov		0.300	0.628	0.624	0.779	0.110	0.551	0.548	0.689	0.684	0.540	0.537	0.133	0.371
Residual		0.500	0.519	0.519	0.537	0.518	0.507	0.508	0.529	0.525	0.505	0.506	0.536	0.523

Table 37. Percent Bias for 3-Class 70/20/10 MD=3/1

		n=200				n=400				n=600			
	Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1	**-46.934**	**-46.097**	3.400	-2.286	**-46.467**	**-46.237**	4.114	-6.571	**-46.457**	**-45.677**	1.905	-3.271
Prop	C2	**64.045**	**62.960**	**16.500**	**58.000**	**62.535**	**62.605**	**10.165**	7.500	**62.240**	**60.655**	3.333	-5.650
	C3	**200.450**	**196.760**	**-56.800**		**200.200**	**198.450**	**-49.130**	**31.000**	**200.720**	**198.430**	**-20.001**	**34.200**
Mean I													
C1	31.370	-1.788	-2.118	6.690	8.556	-1.565	-1.815	**-10.792**	2.020	-1.429	-1.577	-6.748	2.938
C2	40.090	**-11.249**	**-11.624**	0.637	6.264	**-11.909**	**-12.202**	-0.750	1.568	**-12.038**	**-11.998**	-0.274	-3.603
C3	43.010	**-11.019**	**-10.132**	**16.474**	-	**-10.585**	**-10.052**	-2.458	-2.985	**-10.436**	**-10.186**	-6.385	4.956
Mean S													
C1	1.802	3.579	2.597	-6.232	**-29.461**	2.841	2.142	8.349	0.189	2.969	2.314	-2.680	4.217
C2	4.885	**-38.307**	**-39.470**	5.521	**-38.331**	**-39.353**	**-40.287**	5.909	7.498	**-39.490**	**-39.286**	5.810	1.858
C3	5.917	**-28.290**	**-26.488**	6.449	-	**-27.249**	**-26.224**	6.226	4.440	**-26.806**	**-26.483**	**-11.267**	-6.345
Var-I	16.000	0.841	0.642	-7.964	**14.301**	0.137	0.046	**11.833**	-5.159	0.324	0.206	**18.349**	**21.648**
Var-S	2.000	1.480	1.395	**30.900**	**33.904**	0.720	0.740	**10.694**	**15.043**	0.810	0.815	**16.538**	**14.927**
Cov	0.300	**109.467**	**107.867**	**159.700**	**-63.344**	**83.500**	**82.500**	**129.590**	**128.068**	**80.000**	**79.033**	**-55.614**	**23.659**
Residual	0.500	3.740	3.860	7.340	3.537	1.320	1.560	5.748	4.942	1.000	1.140	7.146	4.533

4.3.2.17 3-Class Prop=70/20/10 MD=3/3 Conditions

Results for the parameter estimates and percent bias levels in the 3-class prop = 70/20/10 MD = 3/3 conditions are presented in Tables 38, 39 and A16. The number of the classes estimated by RJMCMC and DP were accurate. The percentages of selecting a 3-class solution by RJMCMC were relatively high (all above 90%), suggesting a high certainty of the 3-class solution. The class proportions were recovered with relatively high bias levels (around -25% to 69%) by B-Diff and B-Weak when n = 200, and the bias levels decreased (to below 30%) when the sample size increased. RJMCMC and DP yielded bias levels for the class proportions that were inconsistent (around -2% to 41%) across the sample sizes. In general, the estimates were improved for all estimation methods when n = 600 (bias levels were mostly below or slightly higher than 10%).

The intercept means were estimated well for all estimation methods across sample sizes. The bias levels for the slope means in Class 2 were relatively high (slightly higher than 10%) for B-Diff and B-Weak in the n = 200 condition; however, they decreased to below 10% as the sample size became larger. RJMCMC and DP yielded relatively low bias levels (below 10%) for the slope means for almost all sample sizes with the exception that DP produced a moderate bias level (around -46%) for the slope mean in the majority class when n = 200. The bias levels for the growth parameter variances were high (around 20% to 70%) for B-Diff, B-Weak, and DP when n = 200 and decreased to below 10% or slightly higher than 10 as the sample size became larger. RJMCMC estimated the variances of the growth parameters with relatively low bias levels (below or slightly higher than 10%) across sample sizes. The covariance was estimated with very high bias levels (mostly around 15% to over 400%) by all estimation methods. The MSE values were relatively high (e.g., around 20 to 49 for the intercept means and around 5-8 for the intercept variances) for some mean and variance parameters and low for other parameters (around 10 to 21).

Table 38. Parameter Estimates for 3-Class 70/20/10 MD=3/3

			n=200				n=400				n=600			
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
# classes		3.000	3.000	3.000	2.987 (0.943)	2.828	3.000	3.000	2.989 (0.935)	3.263	3.000	3.000	2.975 (0.967)	3.048
	C1		0.506	0.523	0.742	0.666	0.633	0.658	0.632	0.672	0.677	0.692	0.684	0.655
Prop	C2		0.325	0.313	0.183	0.243	0.240	0.224	0.226	0.230	0.212	0.210	0.220	0.217
	C3		0.169	0.164	0.074	0.091	0.127	0.118	0.141	0.098	0.112	0.098	0.096	0.128
Mean I														
C1		31.370	31.157	31.147	32.457	31.054	31.270	31.276	33.277	33.904	31.313	31.067	33.043	30.291
C2		40.090	35.424	35.804	38.574	37.521	38.158	38.745	37.543	38.722	39.184	39.888	37.568	38.614
C3		48.810	46.487	46.637	50.469	49.784	47.852	48.126	46.946	47.517	48.332	48.762	51.003	50.005
Mean S														
C1		1.802	1.997	1.969	2.037	0.971	1.849	1.816	1.926	1.935	1.812	1.780	1.895	1.868
C2		4.885	2.939	3.103	5.158	5.019	4.107	4.353	4.934	4.849	4.529	4.811	5.007	5.187
C3		7.968	7.148	7.193	8.247	8.260	7.627	7.722	8.284	8.235	7.798	7.957	7.548	7.571
Var-I		16.000	19.285	18.998	17.277	20.492	17.125	16.768	15.835	16.744	16.516	16.258	16.784	16.342
Var-S		2.000	2.412	2.386	1.908	3.401	2.154	2.108	1.928	2.279	2.071	2.037	2.251	2.344
Cov		0.300	1.688	1.580	0.373	2.203	0.738	0.576	0.664	0.767	0.475	0.353	0.274	0.345
Residual		0.500	0.520	0.520	0.564	0.488	0.506	0.508	0.521	0.490	0.506	0.506	0.524	0.525

Table 39. Percent Bias for 3-Class 70/20/10 MD=3/3

			n=200				n=400				n=600		
		Pop	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
	C1		**-27.767**	**-25.354**	6.043	-4.857	-9.641	-5.936	-9.644	-4.000	-3.344	-1.073	-2.356	-6.400
Prop	C2		**62.470**	**56.620**	-8.335	**21.500**	**20.155**	**11.905**	**13.180**	**15.036**	5.910	4.895	**10.100**	8.500
	C3		**69.430**	**64.240**	**-25.630**	-9.000	**27.180**	**17.740**	**41.150**	-2.071	**11.590**	-2.280	-3.710	**27.800**
Mean I														
C1		31.370	-0.681	-0.712	3.465	-1.007	-0.320	-0.299	6.079	**8.079**	-0.181	-0.965	5.332	-3.440
C2		40.090	**-11.638**	**-10.690**	-3.782	-6.408	-4.820	-3.356	-6.354	-3.412	-2.259	-0.503	-6.290	-3.681
C3		48.810	-4.759	-4.453	3.398	1.995	-1.964	-1.402	-3.818	-2.650	-0.980	-0.098	4.493	2.449
Mean S														
C1		1.802	**10.794**	9.262	**13.015**	**-46.089**	2.630	0.755	6.903	7.394	0.538	-1.232	5.144	3.678
C2		4.885	**-39.840**	**-36.477**	5.585	2.748	**-15.930**	**-10.890**	0.995	-0.730	-7.296	-1.523	2.491	**6.178**
C3		7.968	**-10.292**	-9.733	3.499	3.660	-4.277	-3.089	3.962	3.356	-2.132	-0.142	-5.268	-4.983
Var-I		16.000	**20.534**	**18.739**	7.983	**28.075**	7.033	4.800	-1.029	4.653	3.223	1.610	4.897	2.135
Var-S		2.000	**20.585**	**19.275**	-4.586	**70.038**	7.680	5.405	-3.614	**13.972**	3.570	1.865	**12.554**	**17.219**
Cov		0.300	**462.700**	**426.667**	**24.476**	**634.234**	**145.833**	**92.000**	**121.443**	**155.780**	**58.367**	**17.767**	-8.776	**15.075**
Residual		0.500	4.020	4.000	**12.768**	-2.320	1.280	1.540	4.148	-2.075	1.220	1.160	4.709	5.016

4.4 Study 2 Discussion

In Study 2, RJMCMC, DP, and the Bayesian estimation method with diffuse and weakly informed priors were examined through various levels of class separation, class proportions and sample size. The primary goal was to investigate the performance of RJMCMC, DP, and the two Bayesian methods under different class separation conditions, when crossed with other influencing factors. The following are some conclusions of Study 2.

4.4.1 Estimation Methods

In general, RJMCMC and DP performed comparably well and sometimes better than the Bayesian estimators with diffuse and weakly informed priors in recovering the model parameters. As detailed in the Study 1 Discussion (Section 3.6.4), RJMCMC and DP were able to provide the estimates for the number of classes without comparing across competing models with different class structures. This made RJMCMC and DP more efficient than Bayes, although DP failed to yield the correct class solution under two conditions (i.e., prop = 80/20, MD = 1 and prop = 70/20/10, MD = 3/1). The percentage of selecting a class solution by RJMCMC also helped with the decision making on the number of classes. Within the Bayesian estimation methods, B-Weak performed better than B-Diff under some conditions and both Bayesian estimators yielded close outcomes under other conditions.

4.4.2 Model Parameter Recovery

Akin to the results of the simulation in Study 1, the accuracy of the recovery of model parameters varied under different conditions. The number of classes was extracted correctly by RJMCMC under all conditions. However, DP was not able to detect the correct number of classes for conditions prop = 80/20, MD = 1 and prop = 70/20/10, MD = 3/1. Specifically, DP extracted one class less for each of these two conditions. This could be due to the sensitivity of the decision on the number of classes to the decimal values of the estimate, as detailed in Section 3.6.3. This also suggests that when there was a minority class, and when the class separation between this minority class and its adjacent class was relatively small, DP could have difficulty correctly extracting the number of classes.

The percentage of selecting a class solution by RJMCMC became higher as the MD values increased under the 2-class model conditions. It was more complicated for 3-class models. In general, the percentages were the lowest under the MD = 1/1 conditions and highest under the MD = 3/3 conditions for all class proportions and samples sizes. Conditions where MD = 1/3 and MD = 3/1 had the moderate magnitude of the percentages. This indicates that class separation, and the location of the separation, could affect the percentage of selecting a class solution. The percentage did not vary much between evenly split class proportions and unevenly split class proportions for 2-class and 3-class models, or across sample size.

In general, the class proportions were recovered better under RJMCMC and DP than under the Bayesian conditions. Within the Bayesian conditions, the performance of B-Weak and B-Diff were comparable in estimating the class proportions under most conditions. The class proportions were recovered better when the classes were evenly split than when they were unevenly split for 2-class and 3-class models. The Bayesian estimators tended to overestimate the minority class while underestimating the majority class when classes were unevenly split. Within the non/semi-parametric methods conditions, RJMCMC tended to underestimate the minority class while DP overestimated the minority class under most conditions where the classes were unevenly split. The recovery of the class proportions was best in conditions where the MD values were the highest (e.g., MD = 3 or MD = 3/3) and were the worst when the MD values were the lowest (e.g., MD = 1 or MD = 1/1). The class size of the minority class were recovered better when the class separation between the minority class and its adjacent class were higher (e.g., MD = 1/3) and vice versa. The recovery of the class proportions usually improved as the sample size increased in general.

Overall, the intercept parameter means were recovery better than other growth parameters; they were recovered comparably well under all four estimation methods. The slope parameter means were recovery better under RJMCMC and DP than the B-Diff and B-Weak under most conditions. The recovery of the growth parameter means generally improved when the MD values were higher and when the sample size became larger. The location of class separation did not affect the recovery of the growth parameter means very much.

The covariance structure of the growth parameters was estimated worse than other parameters. The growth parameter variances were recovered better under B-Diff and B-Weak than under RJMCMC and DP. The growth parameter variances were recovered the worst when MD = 2 for 2-class models when MD = 3/3 for 3-class models under B-Diff and B-Weak. The recovery for the growth parameter variances was not consistent across MD values under RJMCMC and DP. Larger sample sizes did not improve the recovery of the growth parameter variances under all B-Diff and B-Weak and showed inconsistent effects on RJMCMC and DP. The covariance was recovered poorly under all conditions for all estimation methods while the residual variances were recovered well under all conditions (with only a few exceptions).

4.4.3 Implications

The findings of Study 2 have several implications. First of all, RJMCMC and DP, as non/semi-parametric methods, proved to perform comparably to the Bayesian conditions with diffuse and weakly informed priors. These approaches can be used as alternatives to the traditional Bayesian estimation framework, especially because they eliminate the need for estimating multiple competing models and using model fit indices or information criteria to aid in model selection.

Second, between the two Bayesian estimation methods, B-Diff and B-Weak did not show much difference in the recovery of model parameters. Considering the prior specifications implemented for these two estimators, B-Weak differed from B-Diff only in the growth parameter means. Specifically, the priors on the growth parameter means in for B-Weak used the population values as the hyperparameter means and 100 as the hyperparameter variances; all other prior specifications remained the same for both estimators. The fact that the performance of B-Weak and B-Diff were quite comparable under the same conditions suggests that this level of informativeness of the growth parameter mean priors might not impact the parameter recovery.

Third, the covariance structure was recovered poorly under DP, as well as RJMCMC at times. For DP and RJMCMC conditions, the prior specifications implemented on the covariance matrix were quite diffuse and contained very little information. The inaccurate recovery of the covariance structure might be due to the uninformative prior specifications, which is consistent with the previous findings in the SEM literature (see e.g., Depaoli, 2013).

Chapter 5

Main Discussion

In this section, I discuss the main conclusions and implications of the two studies in this dissertation. I also include the contributions and limitations of the model estimation methods introduced in this dissertation and discuss some future directions for methodological research.

5.1 Contributions

This dissertation introduced two Bayesian non/semi-parametric methods, RJMMC and DP, as alternative estimation techniques for LGMMs in the SEM framework. The two simulation studies showed valid and comparable results regarding RJMCMC and DP, and they compared these methods with traditional ML and Bayesian estimation approaches. An empirical example was also included as an illustration of the application of RJMCMC and DP on a longitudinal dataset, and the substantive interpretation of the results was highlighted.

The major advantage of RJMCMC and DP as detailed in Sections 3.6.1 and 4.4.1, is that they provide the number of classes as a parameter estimate, without the presumption and model comparison used in the traditional approaches (e.g., ML estimation). This feature can make the modeling process more efficient and straightforward since the traditional model comparison approaches are dependent on statistical tests and model fit indices (and the indices often disagree!). With these non/semi-parametric methods, researchers may use the numeric values directly calculated by the algorithms as the final estimate of the number of latent classes. They can also use the percentage of selecting a certain number of classes provided by the algorithm as a reference or guideline for decision making in the case of using RJMCMC. The percentage of iterations aligning with a certain class solution can also be interpreted as the degree of (un)certainty in that class solution. The non/semi-parametric approach not only avoids multiple model fitting and comparison processes, it also circumvents the potential contradiction in the model comparison measures. For example, Tofight and Enders (2008) and Nylund et al. (2010) found conflicting evidence in support of model fit indices and information criteria. This contradiction makes it difficult for applied researchers to know which measures to trust (and which not to). Despite being compared with the traditional model comparison methods in this dissertation, the non/semi-parametric methods are not designed to be a replacement of the traditional approach. Instead, they provide cross validation that is complementary to the traditional methods. Researchers should not be confined with the use of only one approach. In conclusion, RJMCMC and DP can benefit the SEM framework by simplifying the modeling process and quantifying the certainty of selecting a particular number of classes for LGMMs. They can also be easier to interpret than some of the model comparison measures traditionally used.

Another advantage of RJMCMC and DP is that they performed better than ML or Bayesian estimators under certain conditions. For example, according to the findings of Study 1, DP performed well in recovering the class proportions when ML and RJMCMC tended to overestimate or underestimate the class size of the minority class when classes were unevenly split. Another example is that RJMCMC and DP performed better than the Bayesian estimation methods with diffuse and weakly informed priors in recovering the slope parameter means.

5.2 Limitations, Suggestions, and Future Directions

The number of latent classes is on a categorical scale, while the numeric value of the estimate calculated by RJMCMC and DP is on a continuous scale. Deciding on the number of classes based on the non/semi-parametric methods is essentially turning a continuous scale into a categorical scale. Therefore, one of the limitations of RJMCMC and DP is that the numeric value can be sensitive to a cutoff threshold. As I have discussed in Section 3.6.3, a value at the middle point between two integers can be rounded/truncated to either end, and the consequent model and the interpretation of the parameter estimates can be completely different in each scenario. Therefore, purely relying on the model estimate to decide on the number of classes can be quite dangerous, especially for applied researchers who might have very little knowledge about the non/semi-parametric modeling process.

This limitation of RJMCMC and DP requires researchers to be very careful when interpreting the results, especially the estimate for the number of classes when they decide to use the non/semi-parametric methods. Therefore, my suggestion for applied researchers is to always make a decision in conjunction with statistical evidence *and* the substantive meaning of the dataset. When the number of classes calculated by an algorithm falls within more than one category (i.e., between two class solutions), then researchers should use their substantive knowledge to help them decide which class solution would better explain the phenomenon being studied.

As discussed in Sections 3.6.4 and 4.4.3, the current model and prior specifications of RJMCMC and DP in this dissertation did not solve the estimation issues linked to the covariance structure for LGMMs. These issues of poor parameter recovery are prominent in ML and the Bayesian framework as well (see e.g., Depaoli, 2013). One important future direction is to identify prior specifications that improve the estimation the covariance structure. Sensitivity analyses can be done through simulation studies to examine how different priors may affect the covariance structure, as well as other parameters. Specifically, different prior implementations should be examined for the inverse Wishart prior, which is directly related to the covariance matrix parameters for both RJMCMC and DP. Another prior that is specific for DP is the dispersion parameter, a, in $DP(a, N(\mu, \Sigma))$. Previous studies indicated that this dispersion parameter affects the number of mixture components (i.e., latent classes) in DP mixture models (e.g., Teh, 2010; Gelman, et al., 2014). However, there is no investigation about the effect of the dispersion parameter on the covariance structure in LGMMs. Simulation studies should be done to examine the recovery of the covariance matrix with different specifications of a in a DP distribution. If this matrix can be recovered more accurately, then many of the problems plaguing the estimation of LGMMs will have been solved.

Overall, RJMCMC and DP appear to be interesting and viable approaches for detecting the number of latent classes in an LGMM. More research on the particular settings that should (or should not) be used within these approaches will help to further illuminate their utility within the field.

References

[NCES]. (2001). *Early childhood longitudinal study: Kindergarten class of 1998-99: Base year public-use data files user's manual (nces 2001-029).* Wahsington, DC.: National Center for Educatio Statistics.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Andrieu , C., de Freitas , N., & Doucet, A. (2001). Robust Full Bayesian Learning for Radial Basis Networks. *Neural Computation*, 2359-2407.

Andrieu, C., & Doucet, A. (1999). Joint Bayesian detection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing, 47:10*, 2667-2676.

Andrieu, C., & Doucet, A. (1999). Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing*, 2667 - 2676.

Andrieu, C., de Freitas, N., & Doucet, A. (2001). Robust full Bayesia learning for radial basis networks. *Neural Computation, 13*, 2359-2407.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications of overextraction of latent trajectory classes. *Psychological Methods, 8*, 338-363.

Blackwell, Blackwell, D., & MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, 353-355.

Blei, D. M., & Jordan, M. I. (2006). Variational methods for the Dirichlet process. *Bayesian Analysis, 1*, 121-144.

Bollen, K. A. (1989). *Structural squations with latent variables.* Hoboken, NJ, USA.: John Wiley & Sons, Inc.,.

Bollen, K. A., & Curran, P. J. (2005). *Latent curve models: A structural equation perspective.* NY: Wiley: Hoboken.

Boys, R. J., & Henderson, D. A. (2001). A Comparison of Reversible-Jump MCMC Algorithms for DNA Sequence Segmentation Using Hidden Markov Models. *Computer Science and Statistics*, 33, 35-49.

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

Carota, C., Filippone, M., & Polettini, S. (2018). Assessing Bayesian Nonparametric Log-Linear Models: an application to Disclosure Risk estimation. *ArXiv e-prints*.

Celeux, G. H., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of American Statistical Association*.

Chung, H., Loken, E., & Schafer, J. (2004). Difficulties in Drawing Inferences with Finite-Mixture Models: A Simple Example with a Simple Solution. *The American Statistician*, 152-158.

Dellaportas, P., & Papageorgiou, I. (2006). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing, 16*, 57-68.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, 1-38.

Depaoli, S. (2013). Mixutre class recovery in gmm under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*, 186-291.

Depaoli, S., & Boyajian, J. (2014). Linear and nonlinear growth models: Describing a new Bayesian perspective. *Journal of Consulting and Clinical Psychology*, 82, 784-802.

Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian statistics to model uncertainty in mixture models: A sensitivity analysis of priors. *Structural Equation Modeling: A Multidisciplinary Journal*, 198-215.

Escobar, M. D., & West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 577-588.

Fan, Y., & Sisson, S. A. (2013). Reversible Jump Markov chain Monte. In S. Brooks, A. Gelman, G. Jones, & X. Meng, *Handbook of Markov Chain Monte.* Chapman & Hall/CRC.

Feng, D. (2016). *'miscF': Miscellaneous Functions.* CRAN.

Gelfand, A. E., Kottas, A., & MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association, 100:471*, 1021-1035.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.

Gelman, A., Garlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. (2014). *Bayesian Data Analysis.* Boca Raton, FL: CRC Press.

Gershmana, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 1-12.

Green, P. J. (1994). Discussion on representations of knowledge in complex systems. *Journal of the Royal Statistical Society (Series B), 56*, 589-590.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrics, 82*, 711-732.

Hastie, D. I., & Green, P. J. (2012). Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica*, 309–338.

Hastie, D., Liverani, S., & Richardson, S. (2017). *PReMiuM: Dirichlet Process Bayesian Clustering, Profile Regression.* CRAN.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57(1)*, 97-109.

Hipp, J., Bauer, D., & Bollen, K. (2005). Conducting Tetrad Tests of Model Fit and Contrasts of Tetrad-Nested Models: A New SAS Macro. *Structural Equation Modeling*, 12(1), 76–93.

Ho, R. K., & Hu, I. (2008). Flexible modelling of random effects in linear mixed models - A Bayesian approach. *Computational Statistics & Data Analysis, 52*, 1347-1361.

Holmes, C. C., & Mallick, B. K. (1998). Bayesian radial basis functions of variable dimension. *Neural Computation, 10*, 1217-1233.

Huelsenbeck, J. P., Larget, B., & Alfaro, M. E. (2004). Bayesian Phylogenetic Model Selection Using Reversible Jump Markov Chain Monte Carlo. *Molecular Biology and Evolution*, 1123-1133.

Hurn, M., Justel, A., & Robert, C. (2003). Estimating Mixtures of Regressions. *Journal of Computational and Graphical Statistics*, 55-79.

Jara, A., Hanson, T., Quintana, F., Mueller, P., & Rosner, G. (2017). *'DPpackage': Bayesian Nonparametric Modeling in R.* CRAN.

Jiang, K., Kulis, B., & Jordan, M. I. (2012). Small-variance asymptotics for exponential family Dirichlet process mixture models. *Advances in Neural Information Processing Systems* (pp. 1-9). Lake Tahoe: Neural Information Processing Systems Foundation, Inc.

Kaplan, D. (2002). Methodological advancesin the analysis of individual growth with relevance to education policy. *Peabody Journal of Education, 77*, 189-215.

Kim, S., Tadesse, M. G., & Vannucci, M. (2006). Variable selection in clustering via Dirichlet process mixture models. *Biometrika*, *93*, 877–893.

Larocque, J.-R., & Reilly, J. P. (2002). Reversible jump MCMC for joint detection and estimation of sources in colored noise. *IEEE Transactions on Signal Processing, 50*, 231-240.

Lee, S. Y., & Song, X. Y. (2003). Model Comparison of nonlinear structural equation. *Psychometrika*, 68, 27-47.

Liu, H., & Song, X. (2017). Bayesian Analysis of Mixture Structural Equation Models With an Unknown Number of Components. *Structural Equation Modeling: A Multidisciplinary Journal*, 25:1 41-55.

Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics*, 351-357.

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

Maceachern, S. N., & Muller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics, 7:2*, 223-238.

Marrs, A. D. (1998). An application of reversible-jump MCMC to multivariate spherical Gaussian mixtures. *Advances in Neural Information Precessin Systems* (pp. 577-583). Denver, CO: Powell's Books.

McLachlan, G. J., & Peel, D. (2000). *Finite mixure models.* New York, NY, USA: John Wiley & Sons, Inc.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*, 1087-1092.

Miller, J. W., & Harrison, M. T. (2016). Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 1-50.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides, & R. E. Schumacker, *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical. In R. Nelson, L. Collins, & A. Sayer, *New methods for the analysis of change. Decade of behavior* (pp. 299-322). Washington, DC: American Psychological Association.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*, 369-377.

Muthén, B. (2010). *Bayesian Analysis In Mplus: A Brief Introduction. (Incomplete Draft, Version 3).*

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*, 882-891.

Muthén, L. K., & Muthen, B. (1998-2017). *Mplus (version 8).* Los Angeles, CA: Muthen & Muthen.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition.* Los Angeles, CA: Muthén & Muthén.

Neal, R. M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 249-265.

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

Papastamoulis, P., & Iliopoulos, G. (2005). Reversible Jump MCMC in mixtures of normal distributions with the same component means. *Computational Statistics & Data Analysis*, 900-911.

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

Roberts, G. O., Petrone, S., & Rosenthal, J. S. (2000). Rates of convergence for Markov chains associated with Dirichlet processes. *Far East Journal of Theoretical Statisticas*, 4(2): 207-236.

Roeder, K., & Wasserman, L. (1997). Practical Bayesian densit estimation using mixtures of normals. *Journal of the American Statistical Association, 92:439*, 894-902.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometricka, 52*, 33-343.

Sisson, S. A., & Hurn, M. A. (2004). Bayesian Point Estimation of Quantitative Trait Loci. *Biometrics*, 60-68.

Stephens, M. (2002). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B*, 795 - 809.

Teh, Y. W. (2011). Dirichlet Process. *Encyclopedia of Machine Learning*, 280-287.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2012). Hierarchical Dirichlet processes. *Journal of the American Statistical Association, 101:476*, 1566-1581.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock, & K. M. Samuelson, *Advances in latent variable mixture models* (pp. 317-341). Charlotte, NC: Information age Publishing.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal, 17*, 165-192.

van de Shoot, R., Sijbrandij, M., Depaoli, S., Winter, S., Olff, M., & Loey, N. (2018). Bayesian PTSD-tranectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, online first publication.

Wang, L., & Dunson, D. B. (2011). Fast Bayesian Inference in Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 196-216.

Zhang, Z., Chan, K. L., Wu, Y., & Chen, C. (2004). Learning a multivariate Gaussian mixture model with the reversible jump MCMC algorithm. *Statistics and Computing, 14*, 343-355.

Appendix A

A.1 Additional Tables for Study 1 and Study 2

Appendix A contains the additional tables for Study 1 and Study 2. Tables A1 and A2 contain MSE values for the parameter estimate for Study 1. Tables A3 – A16 contain MSE values for the parameter estimate for Study 2.

Table A1. Mean Square Errors of the Parameters for GCM, 2-Class 50/50, and 80/20

	n=200			n=400			n=600		
	ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
					GCM				
Mean-I	0.000	0.312	0.011	0.000	0.025	0.067	0.000	0.036	0.001
Mean-S	0.000	0.026	0.004	0.000	0.000	0.003	0.000	0.000	0.014
Var-I	0.010	0.158	3.218	0.000	0.000	2.409	0.000	0.062	1.560
Var-S	0.000	0.428	0.073	0.000	0.016	0.001	0.000	0.035	0.000
Cov	0.000	0.960	0.048	0.000	0.132	0.254	0.000	0.000	0.024
Residual	0.000	0.000	0.065	0.000	0.038	0.035	0.000	0.000	0.039
					2-Class 50/50				
Mean-I C1	0.071	0.552	0.340	0.014	0.131	0.124	0.003	0.142	0.053
Mean-I C2	0.092	1.355	1.548	0.062	0.018	0.022	0.027	1.530	1.075
Mean-S C1	0.221	0.009	0.037	0.187	0.058	0.026	0.091	0.241	0.013
Mean-S C2	0.252	0.090	0.002	0.194	0.084	0.069	0.116	0.011	0.005
Var-I	0.003	1.390	10.732	0.108	0.608	22.288	0.089	0.026	18.054
Var-S	0.001	0.004	0.843	0.001	0.018	0.623	0.000	0.014	0.376
Cov	0.778	0.280	1.464	0.598	0.027	1.343	0.381	0.118	0.884
Residual	0.000	0.045	0.000	0.000	0.042	0.000	0.000	0.040	0.000
					2-Class 80/20				
Mean-I C1	0.613	0.289	0.080	0.244	0.052	0.050	0.094	0.168	1.208
Mean-I C2	0.007	0.449	0.080	0.011	0.143	0.156	0.015	0.011	0.444
Mean-S C1	0.082	0.010	0.118	0.037	0.037	0.023	0.016	0.399	0.026
Mean-S C2	0.327	0.004	0.000	0.118	0.094	0.016	0.067	0.006	0.091
Var-I	0.767	5.058	23.088	0.146	0.637	6.482	0.089	0.064	12.681
Var-S	0.017	0.012	0.980	0.000	0.007	0.601	0.002	0.025	0.245
Cov	0.268	0.084	1.221	0.141	0.076	2.250	0.086	0.107	1.506
Residual	0.000	0.056	0.000	0.000	0.053	0.000	0.000	0.048	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A2. Mean Square Errors of the Parameters for GMM, 3-Class 33/33/33, 45/45/10, and 70/20/10

	n=200			n=400			n=600		
	ML	RJMCMC	DP	ML	RJMCMC	DP	ML	RJMCMC	DP
				3-Class 33/33/33					
Mean-I C1	0.124	0.130	0.557	0.154	0.661	0.094	0.153	0.169	0.388
Mean-I C2	0.017	2.749	26.071	0.027	1.126	0.572	0.017	1.217	0.759
Mean-I C3	0.671	3.640	3.077	0.759	0.794	0.810	0.734	0.333	1.208
Mean-S C1	0.042	0.006	0.007	0.005	0.017	0.016	0.000	0.022	0.024
Mean-S C2	0.018	0.002	0.048	0.026	0.002	0.002	0.026	0.052	0.048
Mean-S C3	0.060	0.420	0.284	0.023	0.004	0.004	0.012	0.003	1.721
Var-I	0.171	1.515	67.289	0.097	7.640	49.435	0.069	1.130	0.974
Var-S	0.027	0.003	1.932	0.010	0.018	1.272	0.005	0.055	0.284
Cov	0.570	0.426	9.163	0.155	0.010	5.900	0.071	0.046	0.738
Residual	0.000	0.057	0.000	0.000	0.042	0.000	0.000	0.038	0.000
				3-Class 45/45/10					
Mean-I C1	0.349	1.266	3.557	0.180	1.350	1.306	0.106	0.249	0.487
Mean-I C2	0.835	1.346	9.290	0.496	0.658	0.783	0.365	0.814	0.372
Mean-I C3	0.118	3.073	0.023	0.014	1.464	3.201	0.002	0.110	0.040
Mean-S C1	0.053	0.018	0.002	0.033	0.047	0.032	0.023	0.029	0.031
Mean-S C2	0.130	0.072	0.178	0.072	0.106	0.068	0.052	0.009	0.008
Mean-S C3	0.423	0.312	0.309	0.257	0.076	0.173	0.177	0.002	0.082
Var-I	0.500	3.115	112.615	0.025	1.369	117.592	0.007	0.026	10.426
Var-S	0.075	0.002	2.592	0.023	0.003	2.924	0.010	0.027	0.548
Cov	0.740	0.295	14.738	0.527	0.048	12.173	0.371	0.153	1.348
Residual	0.000	0.057	0.000	0.000	0.049	0.000	0.000	0.039	0.000
				3-Class 70/20/10					
Mean-I C1	0.968	2.135	1.513	0.650	0.376	0.371	0.504	0.127	4.088
Mean-I C2	1.404	0.865	7.513	0.869	0.064	1.067	0.667	16.032	4.024
Mean-I C3	0.215	1.801	1.362	0.360	1.232	2.917	0.444	2.074	2.025
Mean-S C1	0.009	0.030	0.002	0.012	0.005	0.005	0.002	0.003	0.003
Mean-S C2	0.294	0.471	0.034	2.746	0.091	0.052	0.073	0.001	0.001
Mean-S C2	0.019	0.408	0.132	0.011	0.084	0.050	0.012	0.325	0.174
Var-I	3.112	3.077	124.189	1.353	0.972	82.828	0.637	64.048	0.306
Var-S	0.115	0.007	2.836	0.051	0.001	1.685	0.028	0.000	1.563
Cov	0.006	0.055	14.033	0.000	0.071	6.492	0.000	0.011	6.776
Residual	0.000	0.046	0.000	0.000	0.052	0.000	0.000	0.044	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A 3. Mean Square Errors of the Parameters for 2-Class 50/50 and 80/20, MD=1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
					2-Class 50/50							
Mean-I C1	0.014	0.015	4.427	1.223	0.040	0.032	1.543	0.410	0.052	0.049	2.736	1.077
Mean-I C2	0.010	0.016	1.971	1.963	0.041	7.508	0.837	0.342	0.053	0.053	2.873	0.008
Mean-S C1	0.181	0.172	0.226	0.008	0.171	0.174	0.036	0.023	0.163	0.171	0.008	0.013
Mean-S C2	0.185	0.172	0.150	0.233	0.178	0.176	0.045	0.017	0.170	0.176	0.051	0.053
Var-I	0.054	0.080	0.830	2.310	0.092	0.067	1.057	2.353	0.132	0.134	1.869	13.965
Var-S	0.003	0.003	0.057	0.599	0.003	0.003	0.002	0.426	0.003	0.003	0.011	0.060
Cov	0.382	0.367	0.026	0.100	0.377	0.368	0.516	0.704	0.360	0.368	0.138	0.360
Residual	0.000	0.000	0.040	0.000	0.000	0.000	0.035	0.001	0.000	0.000	0.047	0.000
					2-Class 80/20							
Mean-I C1	0.472	0.464	3.512	2.749	0.343	0.335	6.037	0.203	0.284	0.291	4.465	1.228
Mean-I C2	1.016	0.996	5.198	-	1.166	1.160	0.132	2.091	1.212	1.208	2.650	0.417
Mean-S C1	0.025	0.025	0.075	0.278	0.028	0.027	0.137	0.003	0.027	0.025	0.033	0.008
Mean-S C2	0.591	0.591	0.072	-	0.593	0.587	0.073	0.129	0.575	0.570	0.053	0.110
Var-I	0.163	0.154	21.800	7.812	0.109	0.110	6.528	1.268	0.061	0.062	5.973	8.922
Var-S	0.000	0.000	0.267	0.192	0.001	0.000	0.089	0.191	0.000	0.000	0.050	0.108
Cov	0.158	0.154	0.005	0.213	0.163	0.165	2.515	0.507	0.159	0.158	0.092	0.266
Residual	0.000	0.001	0.054	0.000	0.000	0.000	0.056	0.001	0.000	0.000	0.000	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance. "-" indicates there is no estimate for this parameter.

Table A4. Mean Square Errors of the Parameters for 2-Class 50/50 and 80/20, MD=2

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
					2-Class 50/50							
Mean-I C1	0.520	0.490	3.129	0.573	0.206	0.197	2.696	2.713	0.067	0.067	3.667	0.190
Mean-I C2	0.493	0.469	4.020	4.036	0.194	0.185	3.667	0.824	0.071	0.071	3.629	4.368
Mean-S C1	0.158	0.156	0.052	0.016	0.050	0.052	0.044	0.029	0.016	0.016	0.004	0.002
Mean-S C2	0.158	0.155	3.342	0.059	0.051	0.052	0.073	0.053	0.018	0.018	0.090	0.018
Var-I	11.806	11.364	4.550	73.822	4.256	4.219	1.057	41.641	1.533	1.578	7.076	76.580
Var-S	0.194	0.192	0.160	2.667	0.077	0.076	0.002	1.399	0.031	0.029	0.000	1.090
Cov	2.259	2.229	0.286	11.109	0.843	0.832	0.516	5.939	0.315	0.307	0.059	7.701
Residual	0.000	0.000	0.037	0.000	0.000	0.000	0.268	0.000	0.000	0.000	0.047	0.001
					2-Class 80/20							
Mean-I C1	0.063	0.072	5.508	0.116	0.046	0.048	5.299	2.120	0.038	0.038	9.157	0.012
Mean-I C2	5.818	5.541	5.198	0.578	2.654	2.506	2.149	6.071	1.105	1.038	5.626	1.858
Mean-S C1	0.017	0.019	0.004	0.030	0.001	0.001	0.007	0.001	0.000	0.000	0.014	0.007
Mean-S C2	1.219	1.206	0.072	0.031	0.506	0.482	0.029	0.407	0.188	0.180	0.053	0.112
Var-I	4.545	4.190	7.519	34.164	1.100	1.006	2.683	15.539	0.215	0.185	0.188	10.758
Var-S	0.086	0.088	0.135	1.234	0.025	0.023	0.081	1.309	0.005	0.005	0.019	0.065
Cov	1.309	1.302	0.002	5.664	0.375	0.349	0.908	3.129	0.079	0.072	0.007	0.336
Residual	0.000	0.000	0.015	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.002	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A5. Mean Square Errors of the Parameters for 2-Class 50/50 and 80/20, MD=3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
					2-Class 50/50							
Mean-I C1	0.005	0.005	0.365	1.212	0.001	0.001	1.304	0.423	0.000	0.000	1.360	1.084
Mean-I C2	0.005	0.005	1.971	1.982	0.000	0.000	0.001	3.633	0.000	0.000	0.009	0.009
Mean-S C1	0.000	0.001	0.005	0.006	0.000	0.000	0.008	0.012	0.000	0.000	0.000	0.000
Mean-S C2	0.001	0.001	0.619	0.081	0.000	0.000	0.137	0.138	0.000	0.000	0.000	0.000
Var-I	0.918	0.933	5.669	56.761	0.139	0.141	1.034	0.475	0.053	0.053	2.062	0.027
Var-S	0.017	0.017	0.064	1.943	0.003	0.004	0.004	0.024	0.001	0.001	0.072	0.140
Cov	0.048	0.051	0.036	8.375	0.006	0.006	0.070	0.065	0.002	0.002	0.003	0.009
Residual	0.000	0.000	0.039	0.000	0.000	0.000	0.027	0.001	0.000	0.000	0.016	0.000
					2-Class 80/20							
Mean-I C1	0.011	0.011	1.761	2.123	0.003	0.000	1.474	2.123	0.001	0.000	0.781	1.239
Mean-I C2	0.307	0.287	0.410	0.223	0.043	0.000	6.081	0.223	0.018	0.000	0.108	4.601
Mean-S C1	0.001	0.001	0.010	0.001	0.000	0.000	0.021	0.001	0.000	0.000	0.027	0.007
Mean-S C2	0.040	0.038	0.029	0.028	0.006	0.000	0.030	0.028	0.003	0.000	0.005	0.018
Var-I	0.437	0.401	3.523	0.448	0.036	0.117	3.602	0.448	0.018	0.055	11.370	0.448
Var-S	0.009	0.009	0.007	0.087	0.001	0.003	0.004	0.087	0.000	0.001	0.021	0.087
Cov	0.016	0.013	0.010	0.001	0.000	0.003	0.065	0.001	0.000	0.002	0.016	0.001
Residual	0.000	0.000	0.029	0.001	0.000	0.000	0.004	0.001	0.000	0.000	0.000	0.001

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A6. Mean Square Errors of the Parameters for 3-Class 33/33/33, MD=1/1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	1.346	1.004	2.135	2.199	1.690	1.214	2.474	0.175	1.734	1.329	1.206	0.817
Mean-I C2	0.000	0.089	6.938	1.277	0.004	0.073	7.263	0.283	0.005	0.090	1.339	0.025
Mean-I C3	8.416	1.753	0.494	1.028	1.464	1.863	6.345	1.501	1.583	2.108	4.335	0.830
Mean-S C1	0.613	0.561	0.017	0.038	0.982	0.520	0.060	0.000	0.572	0.487	0.004	0.024
Mean-S C2	0.000	0.002	1.320	0.572	0.062	0.002	0.138	0.021	0.001	0.004	0.004	0.002
Mean-S C3	0.610	0.654	0.001	0.083	0.566	0.619	0.068	0.101	0.558	0.602	1.199	0.011
Var-I	1.535	1.798	6.061	54.317	1.968	2.019	0.048	26.698	2.190	2.462	17.331	31.979
Var-S	0.032	0.031	0.099	1.376	0.041	0.039	0.008	0.861	0.040	0.042	0.088	0.845
Cov	1.858	1.877	0.161	5.584	1.690	1.677	0.874	4.414	1.565	1.573	0.008	5.272
Residual	0.000	0.000	0.046	0.000	0.000	0.000	0.023	0.001	0.000	0.000	0.047	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A7. Mean Square Errors of the Parameters for 3-Class 33/33/33, MD=1/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	1.949	0.056	3.936	0.266	1.946	0.041	0.328	2.500	1.910	0.040	2.887	0.497
Mean-I C2	0.534	0.891	1.782	3.512	1.774	0.001	0.837	0.461	1.974	0.016	3.740	3.382
Mean-I C3	0.360	1.049	0.013	1.724	0.031	0.069	2.443	0.880	0.013	0.025	2.048	3.671
Mean-S C1	0.259	0.108	0.137	0.217	0.244	0.067	0.016	0.000	0.252	0.054	0.024	0.041
Mean-S C2	0.065	0.009	0.037	0.320	0.226	0.033	0.001	0.010	0.257	0.043	0.067	0.128
Mean-S C3	0.054	0.143	0.046	0.171	0.005	0.009	0.135	0.035	0.002	0.004	0.017	0.017
Var-I	0.389	0.393	2.680	3.411	0.154	0.131	0.208	2.202	0.118	0.090	8.294	40.564
Var-S	0.009	0.009	0.045	0.869	0.004	0.004	0.000	0.076	0.003	0.003	0.128	0.076
Cov	0.187	0.180	0.041	1.061	0.095	0.093	0.534	0.482	0.076	0.071	0.033	0.557
Residual	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.045	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A8. Mean Square Errors of the Parameters for 3-Class 33/33/33, MD=3/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	7.344	0.133	5.457	2.173	0.061	1.030	1.201	2.509	0.002	0.007	5.313	0.090
Mean-I C2	3.652	1.428	1.329	3.467	0.020	0.001	0.931	6.126	0.000	0.001	2.528	3.360
Mean-I C3	0.676	2.179	1.651	5.341	0.007	0.653	0.181	3.119	0.001	0.003	0.001	0.028
Mean-S C1	0.931	0.020	0.225	0.324	0.007	0.000	0.119	0.032	0.000	0.000	0.024	0.001
Mean-S C2	0.468	0.172	0.031	0.117	0.003	0.000	0.005	0.011	0.002	0.000	0.004	0.004
Mean-S C3	0.089	0.279	0.007	0.170	0.001	0.057	0.072	0.035	0.000	0.000	0.001	0.073
Var-I	5.570	5.978	2.241	50.013	0.516	1.977	0.009	0.287	0.158	0.157	3.846	2.863
Var-S	0.086	0.093	0.020	3.371	0.010	0.034	0.011	0.026	0.004	0.004	0.009	0.145
Cov	0.536	0.585	0.028	9.413	0.039	0.203	0.309	0.050	0.012	0.013	0.008	0.067
Residual	0.000	0.000	0.002	0.001	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A9. Mean Square Errors of the Parameters for 3-Class 45/45/10, MD=1/1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.123	0.049	2.135	2.199	0.206	0.138	3.119	0.566	0.237	0.187	3.656	3.740
Mean-I C2	0.375	0.539	6.938	1.277	0.878	0.630	0.004	1.175	0.931	0.572	3.087	0.060
Mean-I C3	15.374	5.988	0.494	1.028	6.037	6.285	0.524	1.628	6.017	6.703	0.728	2.654
Mean-S C1	0.303	0.280	0.017	0.038	0.301	0.278	0.008	0.023	0.291	0.283	0.006	0.017
Mean-S C2	0.046	0.101	1.320	0.572	0.049	0.096	0.004	0.006	0.043	0.107	0.040	0.000
Mean-S C3	1.949	1.615	0.001	0.083	1.910	1.621	0.035	0.067	1.891	1.555	0.518	0.142
Var-I	0.001	0.008	6.061	54.317	0.062	0.060	7.756	9.084	0.161	0.243	14.761	26.266
Var-S	0.000	0.000	0.099	1.376	0.004	0.003	1.248	0.806	0.007	0.007	0.031	0.156
Cov	0.906	0.937	0.161	5.584	0.949	0.931	1.143	2.217	0.899	0.931	0.004	1.107
Residual	0.000	0.000	0.046	0.000	0.000	0.000	0.030	0.002	0.000	0.000	0.043	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A10. Mean Square Errors of the Parameters for 3-Class 45/45/10, MD=1/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.095	0.047	0.783	4.376	0.072	0.070	3.948	3.073	0.097	0.097	3.948	0.721
Mean-I C2	0.310	0.022	3.077	7.290	0.244	0.237	0.003	0.043	0.252	0.250	0.679	0.554
Mean-I C3	6.948	4.486	1.381	3.960	0.569	0.517	2.065	1.674	0.293	0.288	0.120	0.640
Mean-S C1	0.237	0.237	0.033	0.007	0.202	0.199	0.075	0.097	0.180	0.179	0.007	0.053
Mean-S C2	0.026	0.117	0.045	0.066	0.283	0.281	0.005	0.018	0.248	0.245	0.013	0.069
Mean-S C3	0.897	0.584	0.083	0.081	0.068	0.066	0.043	0.071	0.036	0.036	0.145	0.160
Var-I	0.681	0.629	0.861	1.734	0.348	0.309	4.068	0.123	0.353	0.362	5.537	18.706
Var-S	0.013	0.013	0.269	0.389	0.007	0.007	0.291	0.557	0.008	0.008	0.004	0.194
Cov	0.593	0.573	0.008	1.450	0.361	0.358	0.048	1.124	0.326	0.328	0.010	0.598
Residual	0.001	0.000	0.039	0.000	0.000	0.000	0.005	0.001	0.000	0.000	0.052	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A11. Mean Square Errors of the Parameters for 3-Class 45/45/10 MD=3/1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	1.158	0.138	0.769	0.007	0.283	0.041	3.944	1.858	0.041	0.006	8.422	1.136
Mean-I C2	5.915	1.252	0.387	2.876	1.518	0.246	1.476	1.197	0.009	0.009	0.527	1.593
Mean-I C3	8.538	9.923	7.773	0.845	9.205	9.816	0.872	0.956	11.303	9.084	4.849	49.028
Mean-S C1	0.203	0.038	0.001	0.007	0.050	0.011	0.033	0.095	0.009	0.003	0.004	0.002
Mean-S C2	0.005	0.193	0.069	0.065	0.006	0.035	0.008	0.166	0.024	0.003	0.019	0.002
Mean-S C3	3.602	1.320	0.591	0.095	2.184	1.300	0.135	0.068	1.201	1.156	0.797	1.820
Var-I	0.205	0.217	3.042	159.517	0.002	26.874	1.261	5.040	0.000	0.000	7.215	0.486
Var-S	0.004	0.004	0.284	2.914	0.000	0.457	0.108	0.585	0.000	0.000	0.054	0.217
Cov	0.132	0.130	0.008	13.749	0.026	0.026	0.050	0.733	0.018	0.018	0.005	0.003
Residual	0.001	0.000	0.001	0.001	0.000	0.000	0.012	0.000	0.000	0.000	0.001	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A12. Mean Square Errors of the Parameters for 3-Class 45/45/10, MD=3/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	4.256	1.156	4.052	0.837	0.218	1.156	1.360	3.105	0.002	0.000	1.362	1.136
Mean-I C2	0.024	0.102	3.045	2.887	0.081	0.102	0.748	0.794	0.001	0.017	2.295	1.593
Mean-I C3	54.553	33.397	2.749	3.791	4.757	33.397	6.574	1.682	0.339	0.168	2.746	1.445
Mean-S C1	0.697	0.288	0.066	0.034	0.033	0.288	0.004	0.022	0.000	0.000	0.002	0.002
Mean-S C2	0.139	0.047	0.131	0.066	0.006	0.047	0.026	0.004	0.000	0.000	0.026	0.002
Mean-S C3	9.741	4.339	0.072	0.089	0.594	4.339	0.072	0.071	0.043	0.000	0.196	0.493
Var-I	83.229	76.703	0.010	64.208	2.680	76.703	1.651	261.760	0.157	0.194	0.980	0.486
Var-S	1.323	1.217	0.057	13.675	0.044	1.217	0.027	5.523	0.003	0.004	0.002	0.217
Cov	12.924	11.896	0.000	87.966	0.343	11.896	3.360	34.059	0.015	0.017	0.011	0.003
Residual	0.001	0.001	0.013	0.001	0.000	0.001	0.000	0.001	0.000	0.000	0.013	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A13. Mean Square Errors of the Parameters for 3-Class 70/20/10, MD=1/1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.122	0.192	0.460	1.732	0.042	0.105	0.392	0.222	0.022	0.108	3.618	0.850
Mean-I C2	2.952	2.011	0.870	2.459	2.904	1.680	2.418	0.412	3.240	0.776	2.280	0.203
Mean-I C3	8.094	9.181	0.726	3.736	7.678	9.382	13.418	1.641	6.165	10.375	2.329	2.353
Mean-S C1	0.074	0.075	0.060	0.029	0.177	0.064	0.066	0.011	0.161	0.049	0.000	0.014
Mean-S C2	0.358	0.309	0.014	0.016	0.629	0.259	0.017	0.019	0.702	0.167	0.068	0.008
Mean-S C3	2.028	2.152	0.072	0.090	1.690	1.966	0.137	0.067	1.297	1.952	0.382	0.145
Var-I	0.009	14.296	13.119	8.271	0.011	0.008	4.588	19.519	0.058	0.029	10.170	24.039
Var-S	0.000	0.245	0.057	0.523	0.000	0.000	0.805	0.419	0.000	0.000	0.020	0.335
Cov	0.762	1.628	0.776	0.787	0.546	1.374	0.146	2.123	0.366	0.383	0.002	1.173
Residual	0.001	0.000	0.052	0.001	0.000	0.000	0.000	0.002	0.000	0.000	0.045	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A14. Mean Square Errors of the Parameters for 3-Class 70/20/10, MD=1/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.356	0.367	2.323	5.364	0.224	0.229	3.602	0.691	0.185	0.367	0.092	1.173
Mean-I C2	1.073	1.239	1.090	2.459	1.690	1.656	0.543	1.831	1.700	1.239	1.553	2.135
Mean-I C3	0.949	0.731	1.153	4.020	0.190	0.184	0.047	0.251	0.123	0.731	0.581	4.831
Mean-S C1	0.040	0.041	0.018	0.029	0.033	0.031	0.092	0.011	0.026	0.041	0.144	0.001
Mean-S C2	0.613	0.666	0.058	0.018	0.671	0.658	0.210	0.020	0.612	0.666	0.097	0.009
Mean-S C3	0.122	0.094	0.102	0.078	0.024	0.024	0.066	0.137	0.016	0.094	0.157	0.159
Var-I	0.012	0.012	1.623	0.006	0.007	0.011	2.883	8.094	0.004	0.012	2.353	3.565
Var-S	0.000	0.000	0.382	1.192	0.000	0.000	0.030	0.736	0.000	0.000	0.119	0.289
Cov	0.158	0.157	0.229	0.222	0.121	0.115	0.171	0.585	0.110	0.157	0.138	0.215
Residual	0.000	0.000	0.001	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.001	0.000

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A15. Mean Square Errors of the Parameters for 3-Class 70/20/10, MD=3/1

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.315	0.441	4.406	7.204	0.241	0.324	11.465	0.402	0.201	0.245	4.482	1.164
Mean-I C2	20.340	21.716	0.065	6.305	22.791	23.932	0.091	0.396	23.290	23.136	0.012	2.179
Mean-I C3	22.458	18.992	50.211	-	20.721	18.688	1.117	1.649	20.142	19.193	7.541	48.930
Mean-S C1	0.004	0.002	0.013	0.282	0.003	0.002	0.023	0.000	0.003	0.002	0.002	0.004
Mean-S C2	3.501	3.717	0.073	3.504	3.694	3.873	0.084	0.134	3.721	3.683	0.081	0.091
Mean-S C3	2.802	2.455	0.146	-	2.599	2.409	0.135	0.069	2.515	2.455	0.445	2.736
Var-I	0.018	0.011	1.623	5.235	0.000	0.000	3.583	0.681	0.003	0.001	8.620	0.117
Var-S	0.001	0.001	0.382	0.460	0.000	0.000	0.046	0.091	0.000	0.000	0.110	0.118
Cov	0.108	0.105	0.229	0.036	0.063	0.062	0.151	0.147	0.058	0.056	0.028	0.002
Residual	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.

Table A16. Mean Square Errors of the Parameters for 3-Class 70/20/10, MD=3/3

	n=200				n=400				n=600			
	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP	B-Diff	B-Weak	RJMCMC	DP
Mean-I C1	0.045	0.050	1.182	0.100	0.010	0.009	3.637	6.421	0.003	0.092	2.799	1.164
Mean-I C2	21.772	18.370	2.298	6.600	3.733	1.809	6.487	1.871	0.821	0.041	6.360	2.179
Mean-I C3	5.396	4.722	2.752	0.949	0.918	0.468	3.474	1.672	0.228	0.002	4.809	1.428
Mean-S C1	0.038	0.028	0.055	0.691	0.002	0.000	0.015	0.018	0.000	0.000	0.009	0.004
Mean-S C2	3.787	3.176	0.075	0.018	0.605	0.283	0.002	0.001	0.127	0.005	0.015	0.091
Mean-S C3	0.672	0.601	0.078	0.085	0.116	0.061	0.100	0.071	0.029	0.000	0.176	0.158
Var-I	10.791	8.988	1.631	20.178	1.266	0.590	0.027	0.554	0.266	0.067	0.615	0.117
Var-S	0.170	0.149	0.008	1.963	0.024	0.012	0.005	0.078	0.005	0.001	0.063	0.118
Cov	1.927	1.638	0.005	3.621	0.192	0.076	0.132	0.218	0.031	0.003	0.001	0.002
Residual	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001

Note. Mean-I = mean of the intercept and Mean-S = mean of the slope; C1 = latent Class 1, C2 = latent Class 2, and C3 = latent Class 3; Var-I = the variance of the intercept, Var-S = the variance of the slope, and Cov = the covariance between the intercept and the slope; the variances and the covariance were held equal across latent classes; Residual = residual variance.