ABSTRACT

Title of the Dissertation: ESTIMATION AND MODEL SELECTION FOR FINITE MIXTURES OF LATENT INTERACTION MODELS

Jui-Chen Hsu, Doctor of Philosophy, 2011

Directed by: Professor Gregory R. Hancock, Department of Measurement, Statistics and Evaluation

Professor Jeffrey R. Harring, Department of Measurement, Statistics and Evaluation

Latent interaction models and mixture models have received considerable attention in social science research recently, but little is known about how to handle if unobserved population heterogeneity exists in the endogenous latent variables of the nonlinear structural equation models. The current study estimates a mixture of latent interaction models using an unconstrained product indicator approach. It also investigates the performance of the method in parameter recovery, classification quality, and identification of the correct number of latent classes using data simulated under a variety of conditions.

The major findings of this study are (1) class separation in the factor means is a critical factor influencing parameter recovery and classification results; (2) the relative biases (or bias) for the class-specific interaction effects are larger than those

of other parameters; (3) the precision of all structural parameter estimates for the conditions with larger separation are satisfactory; (4) entropy values, correct assignment probabilities, and convergence rates are improved for larger separation models; (5) AIC, BIC, and ABIC all support the two-class models over one-class models regardless data are generated from one- or two-class latent interaction models.

ESTIMATION AND MODEL SELECTION FOR FINITE MIXTURES OF LATENT INTERACTION MODELS

By

Jui-Chen Hsu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Gregory R. Hancock, Co-Chair
Professor Jeffrey Harring, Co-Chair
Professor George Macready
Professor Hong Jiao
Professor Paul J. Hanges

DEDICATION

To my dearest parents,
Tien-Liang Hsu and I-Jen Wang Hsu.

謹將此論文獻給我的爸爸媽媽, 許添良、許王一媜。

ACKNOWLEDGEMENTS

I am indebted to many people for their help, support, and guidance which made this dissertation possible.

First, I would like to express my deepest gratitude to my advisors, Dr. Hancock and Dr. Harring, whose research insight, enthusiasm, and guidance have illuminated me while conducting this study. I am very fortunate to have two mentors, rather than one, whose immeasurable trust, patience, and support gave me the drive to complete the dissertation. Their perfect mentoring styles have made all the impossible become possible.

I would like to thank my committee members, Dr. Macready, Dr. Jiao, and Dr. Hanges, for their valuable input on the dissertation. I would also like to thank Dr. Mislevy for his magnificent instructions and help. I would like to extend my gratitude to all the teachers who have helped me in my academic career over the years, especially Dr. Wu, Shan-Hwa and my previous advisor, Dr. Anderson.

I appreciate Dr. Hsieh and Dr. Lu for their constructive suggestions. I am grateful to Dr. Ye for her computational help. I am grateful to Amy, Yanmei, Wilbur, Chuan, and John for giving me valuable feedback on the draft.

I would like to extend a heartfelt thanks to Joyce for her valuable suggestions and timely help which were critical to me in accomplishing my goals for the graduate

program. Thanks to Marian, who is like a local parent, understanding how I feel and encouraging me to move forward. I thank Christina for her thoughtful suggestions and help. I would like to thank sisters Yang, Bijou, Helen and brothers Hu, Wu, Kuan, and Peng of the Lamrim class for their kindly help and support. Thanks to Dolly, Tina, Larry, Vivian, Marcella, Dan and Joanne for their encouragement and help. Thanks to my colleagues in Tainan- S. H. Tao, M. H. Wu., T. H. Wu., T. L. Pon, and Z. Lee for their continuous support. Thanks to Yuan, Ying and Brandi for their help and suggestions on my dissertation.

Finally, but most of all, I would like to thank my parents for their unconditional love and endless support which gave me the courage and persistence to reach every milestone in my doctoral program. I would also like to thank my brothers, Jovan and Jason, and their families for being there whenever I needed. I thank Buddhas, Bodhisattvas, and dharma masters for blessing me with the ability to complete the dissertation and for the precious experiences that I have had along my graduate study.

我由衷地感謝我的爸爸、媽媽對我無止盡的愛護及支持，讓我有勇氣有毅力達成博士學習過程中的每一個里程碑！也衷心地感謝我的弟弟- 嘉榮、嘉新及他們的家人對我的支持，更感恩諸佛菩薩及師長的慈悲加持，讓我在博士學習過程中得到這麼多貴人的幫忙及這麼難得的寶貴經驗，讓我有能力、有勇氣得以實現自己的夢想!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Structural equation modeling (SEM) has been increasingly utilized in behavioral, social, and psychological research due to the rapid growth in the development of estimation methods and specialized SEM programs such as LISREL, Mplus, EQS, AMOS, and Mx (e.g., Hancock & Mueller, 2006; Lee, 2007). Traditional SEM examines linear relations between latent variables and observed variables (indicators) as well as relations among latent variables. The linearity assumption in the structural equation allows for a straightforward interpretation but restricts the ability of the model to explain many phenomena found in practice, which posit more complex associations. For example, for elementary school students the relation between latent motivation and latent achievement may systematically depend on a third variable – latent parental support. In this context, parental support would be seen as moderating the relation between motivation and achievement. From a modeling perspective, studying moderation requires adding a cross-product term to the structural model which only has additive effects.

In contrast to traditional linear SEM, nonlinear SEM has been recognized as providing a more accurate and meaningful explanation for data with increased complexity (Lee, Song, & Tang, 2007). Adding nonlinear exogenous terms to a structural equation often permits better prediction of the dependent variable. Although

general nonlinear latent variable models represent a very broad class (Wall & Amemiya, 2007), a considerable amount of attention has been dedicated to polynomial models and more specifically, to quadratic models which include second-order terms as well as the product among the latent variables representing interaction effects. This added complexity does not come without some consequence. Wall (2009), among others, have pointed out that maximum likelihood estimation of nonlinear structural models assuming a normal distribution for the latent variables is more complicated than in standard structural equation models with only linear relations.

1.1 Previous Work

Nonlinear SEM typically assumes that the model holds for a single population. However, a single homogeneous population assumption may not be always feasible in practice. In the previous example which hypothesized that parental support moderated the linear relation between motivation and achievement, heterogeneity in the population may be evident through the moderating effect. In other words, moderation may well be stronger for some children than others. For children with greater parental support, for example, the linear relation between motivation and achievement may be quite strong; for children with less parental support, on the other hand, the linear relation between motivation and achievement may be quite weak. As Muthén (1989)

stated, “Data are frequently analyzed as if they were obtained from single populations although it is often unlikely that all individuals in our sample have the same set of parameter values (p. 558).” The results of the analysis of SEM can be considerably distorted when heterogeneity embedded within the data is not taken into account (Muthén, 1989; Yung, 1997). If heterogeneity is observed and known a priori, then multiple-group approaches can be applied to the analysis. For example, Rigdon, Schumacker, and Wothke (1998) used a multi-sample approach to test the hypothesis of whether two groups have equivalent latent interaction effects. Under the conditions of equal group means and unequal group means, they tested the hypothesis with a chi-square difference test. Incorporating known population heterogeneity can also be accomplished in the structural model itself. Inclusion of a categorical variable with known group membership, like gender, could be integrated into a latent interaction model by creating, and eventually testing, a three-way interaction term (analogous to a three-way interaction in between-subjects analysis of variance (ANOVA) with three factors). Despite the multi-group and direct modeling approaches in handling observed population heterogeneity, little is known about how to handle unobservable heterogeneity in the dependent latent variable of a nonlinear structural equation model. Thus, there is a need for research to determine whether interaction effects in nonlinear

structural equation models can be uncovered and estimated when unobserved population heterogeneity exists.

1.2 Purpose of Research

While other types of nonlinearities could be used as the focus of this study (see e.g., Wall, 2009), the latent interaction model is substantively compelling and is more commonly used in practice than models with higher order terms (terms of degree ≥ 2) since it plays the most important role in the existing literature of nonlinear latent variables (Cudeck, Harring, & du Toit, 2009). The objectives of the current study are (1) to estimate a mixture latent variable interaction model, (2) to evaluate the performance the performance of the estimation method, and (3) to identify the correct number of latent classes in the data generated from mixture latent interaction models.

To accomplish the objectives, specific research questions are as follows.

1. Can a mixture of latent interaction models be estimated using an unconstrained product indicator approach?
2. How well does the estimation method for two-class latent variable interaction models perform in recovering the parameters of interest under various conditions?
3. How well does this estimation method classify the observations from a two-class latent variable interaction model?

4. What is the probability to correctly select models from one-, two- and three-class latent interaction models for data generated from a population with two-class latent interaction effects using model fit statistics?

These research questions will be investigated by two Monte Carlo simulation studies which generate data from mixture latent variable interaction models with pre-specified parameter values under a variety of manipulated conditions.

Chapter 2: Literature Review

This chapter will explore the basic idea of interactions among continuous observed variables to interactions among continuous latent variables. The proposed nonlinear structural model with latent interaction will include latent classes where the relation between the endogenous latent variable and an exogenous latent variable is moderated by a second exogenous latent variable, and the magnitude of this moderation is allowed to vary across unobservable groups. These unobservable latent classes can be thought of as a finite mixture. Thus, the theoretical foundations of mixture modeling will also be subsequently reviewed in the second section. Finally, estimation of the mixture latent interaction model will be outlined and discussed.

2.1 Interactions among Observed Variables

Two continuous variables interact with each other if the relation between the outcome variable and one predictor variable depends on the values of the third variable. To understand this dependency, consider the following regression equation which describes how motivation (X) and parental support (Z) impact achievement (Y):

$$\begin{aligned} \hat{Y} &= \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ \\ &= (\beta_0 + \beta_2 Z) + (\beta_1 + \beta_3 Z) X \end{aligned} \tag{1}$$

where $\hat{Y}$ is the estimated value for the outcome variable Y, X and Z are the

predictors, XZ is the interaction between predictors, β_0 is the intercept , and β_1, β_2, β_3 are the regression coefficients. It can be clearly seen from Equation 1 that the regression between Y and X (including both intercept and slope) is a function of Z . *Z* is sometimes called the *moderator* of the relation between *X* and *Y* because the nature of the relation depends on (is moderated by) values of *Z*. Figure 1 illustrates a moderated regression model such that the relation between Y and X depends on different levels of Z (i.e., Z_{high}, Z_{med}, and Z_{low}). In other words, the slopes of achievement on motivation are different for students with high, medium, and low levels of parental support.

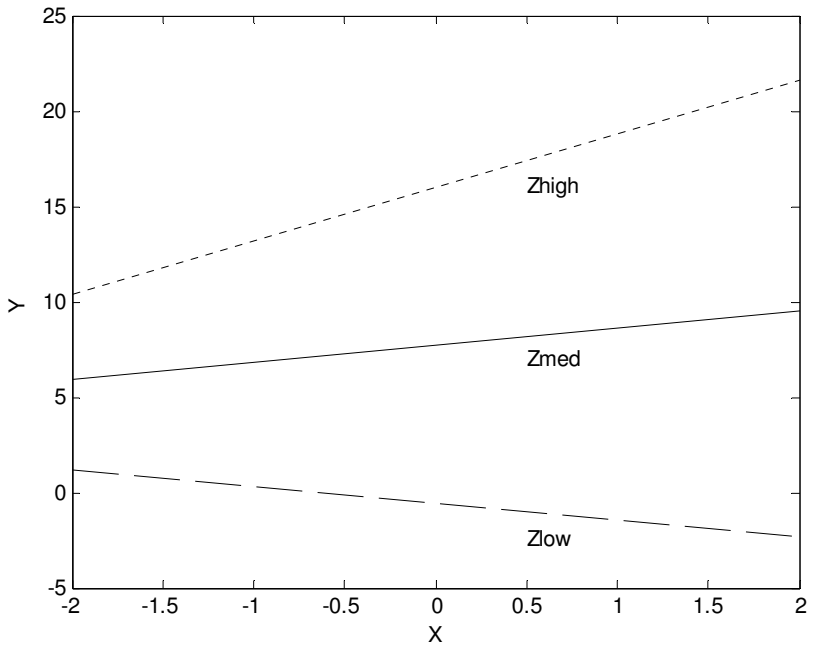


Figure 1. Two-way interaction: depicting how variable Z moderates the relation between Y and X .

Compared with two-way interactions, three-way interactions are less commonly

seen in the social science literature. However, if one predictor is a categorical variable, then the three-way interactions can be reduced to a two-way interaction model.

Suppose

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW .$$

If W, for example, is a dichotomous variable ($W = 0, 1$), then

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_4 XZ \qquad \text{when } W = 0$$

and $$\hat{Y} = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) X + (\beta_2 + \beta_6) Z + (\beta_4 + \beta_7) XZ \quad \text{when } W = 1 .$$

The caveat is that the interaction between X and Z is allowed to be different at the two levels of W. For $W = 0$, the slopes and intercepts of achievement on motivation for children with high and low levels of parental support are different from those in $W = 1$. Figure 2 illustrates that the regression of Y on X depends not only on the levels of Z but also on the group membership (i.e., $W = 0, 1$). In other words, the interaction patterns differ across groups.

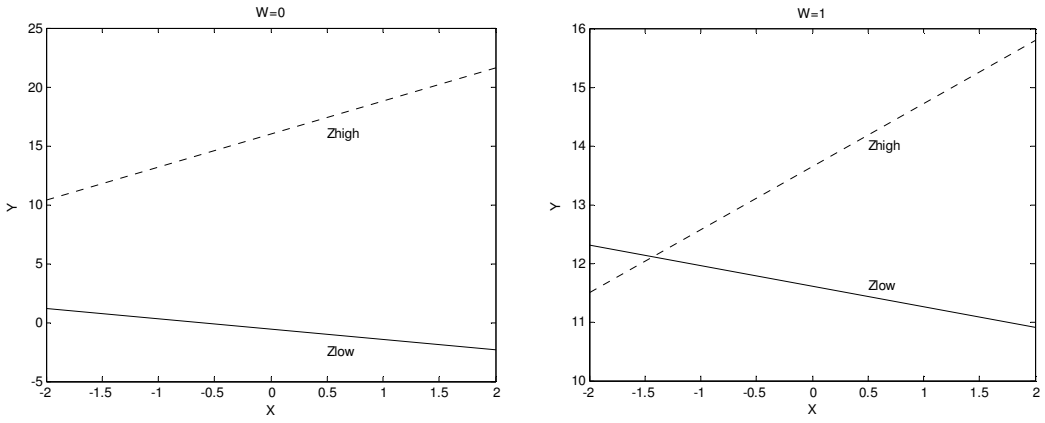


Figure 2. Three-way interaction depicting how variable Z moderates the relation between Y and X differently at two levels of a dichotomous variable, W.

For methods falling under the general linear model (GLM) umbrella, such as multiple regression or ANOVA, it is common to mean-center variables. Mean-centering is a linear transformation of a variable so that its mean is zero. Note that the variability of the variable is invariant to this type of transformation. In other words, the location of the variable is changed whereas the standard deviation remains unchanged. It is recommended to mean-center a continuous predictor before regression analyses with interactions except when it has a meaningful zero point and not to center the dependent variable in order to keep its origin scale for prediction (Cohen, Cohen, West, & Aiken, 2003). It can be demonstrated that when mean-centering the predictors in a regression equation with interaction, all regression coefficients change except that of the highest order term(s) (Harring, 2010 unpublished manuscript). Mean-centering has the advantages of providing meaningful interpretations and reducing the multicollinearity among the predictors that might cause problems (e.g., large sampling variability) in estimating and making inferences about the regression coefficients (Aiken & West, 1991; Neter, Kutner, Nachtsheim, & Wasserman, 1996).

It is assumed in multiple regression or ANOVA that the independent variables are measured without errors. However, ignoring measurement error may lead to bias in parameter estimation and/or lower power for testing interactions (Jaccard & Wan,

1995; Moulder & Algina, 2002). Ideally, observed measures indicating the presence of an underlying construct and the underlying latent construct itself will be perfectly or nearly perfectly correlated. Often however, this is not true because measurement error in the observed variables keep them far from being perfectly correlated (Bollen, 1989). Furthermore, measurement errors are compounded due to the multiplicative nature of the interaction term and therefore lead to weakened latent interaction effects (Jaccard & Wan, 1995). In other words, the interaction effects should have larger estimates at the latent variable level (Jaccard & Wan, 1995). Therefore, it is more appropriate to use latent variables that separate the measurement error from the observed variables than the traditional observed variables to conduct the interaction analyses.

2.2 Latent Interaction Models

Structural equation models allow researchers to estimate and test simultaneously a series of regression equations with observed variables and latent variables. For example, the LISREL model (Jöreskog & Sörbom, 1996) consists of a measurement equation relating the observed variables to the latent variables and a structural equation relating the endogenous latent variables to other endogenous and exogenous latent variables. The LISREL (Jöreskog & Sörbom, 1996) specification for the structural equation of a model with two exogenous variables interacting with each other can be expressed as

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \varsigma \quad (2)$$

where η represents a continuous endogenous latent variable, ξ_1 and ξ_2 represent continuous first-order (linear) exogenous latent variables, the product term, $\xi_1\xi_2$, represent an interaction between the two exogenous latent variables, α represents the intercept of η when the exogenous predictors are zero; γ_1, γ_2, and γ_3 represent the direct path estimates of the two first-order latent variables and the latent variable interaction, respectively, and ς represents the regression residual.

Latent variable interaction models have become increasingly popular in social science research. Since the seminal work of Kenny and Judd (1984), considerable effort has been devoted to the estimation of these cross-product effects. The estimation approaches for latent interaction models can be categorized into: (1) latent variable score approaches (e.g., Cohen, Cohen, West, & Aiken, 2003; Schumacker, 2002); (2) moment-based approach (Wall & Amemiya, 2003); (3) Maximum likelihood estimation (a) linearization approaches: product indicator approaches (e.g., Bollen, 1995, 1996; Hayduk, 1987; Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Marsh, Wen, & Hau, 2004; Wall & Amemiya, 2001); (b) exact maximum likelihood approaches (e.g., Cudeck, et al., 2009; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Lee & Zhu, 2002); and (4) Bayesian estimation (e.g., Arminger & Muthén, 1999; Lee, 2007). While a variety of estimation methods have been developed for

latent variable interaction models, the current paper limits the discussion to product indicator approaches in that they linearize the latent interaction factor which make them accessible through commonly accessible software (e.g., Mplus).

2.2.1 Product indicator approaches. The indicators of the latent interaction variable for these approaches were formed by taking the products of the indicators of the linear factors. This implies that the variances and the loadings of the product indicators are functions of those of the indicators of the linear factors. Similarly, the mean and variance for the latent interaction variable can also be derived from the means and variances of the linear factors, respectively. Therefore, the class of "constrained approaches" consists of those created based on the Kenny-Judd model (K-J) which required nonlinear parameter constraints (Kelava, Moosbrugger, Dimitruk, & Schermelleh-Engel, 2008; Marsh et al., 2004). The partially constrained approach, which is between constrained and unconstrained approaches, constrained parameters except those in factor covariance matrix (Wall & Amemiya, 2001) allowing for the potential that the exogenous factors are not normally-distributed. The class of "unconstrained approaches" is made up of those that have released these nonlinear constraints (Little et al., 2006; Marsh et al., 2004).

2.2.1.1 Constrained approach. Kenny and Judd (1984) pioneered the analysis of simple latent nonlinear models using product indicators. In this model, they used two

indicators for each of the latent exogenous variables ξ_1 and ξ_2 and a single manifest variable, y, as the criterion. For illustration purposes, the current paper will use three indicators for both linear factors. They suggested mean-centering all observed variables and using all possible products of the observed variables. Therefore, with three indicators per latent variable, there are nine possible product indicators for the Kenny-Judd model. For the purpose of simplicity and consistency, the current study will adopt the matched-paired product indicators strategy proposed by Marsh and colleagues (2004). This approach pairs the indicators from each of the first-order latent variables accordingly if the number of indicators for each first-order latent variable is equal. Marsh et al. (2004) suggested that the precision of the match-pair estimation was systematically better than that of possible products (Kenny & Judd, 1984) or one pair product (Jöreskog & Yang, 1996). There are nonlinear constraints to specify the factor loadings and variance related to the latent interaction term. The LISREL specification of the Kenny-Judd model can be written as

$$\eta = y = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \varsigma$$

$$= \Gamma\xi + \varsigma$$

where $\xi = (\xi_1, \xi_2, \xi_1\xi_2)'$, and $\Gamma = (\gamma_1, \gamma_2, \gamma_3)$.

$$\begin{pmatrix} y-\mu_y \\ x_1-\mu_{x_1} \\ x_2-\mu_{x_2} \\ x_3-\mu_{x_3} \\ x_4-\mu_{x_4} \\ x_5-\mu_{x_5} \\ x_6-\mu_{x_6} \\ (x_1-\mu_{x_1})(x_4-\mu_{x_4}) \\ (x_2-\mu_{x_2})(x_5-\mu_{x_5}) \\ (x_3-\mu_{x_3})(x_6-\mu_{x_6}) \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \lambda_1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ 0 & 0 & \lambda_7 \\ 0 & 0 & \lambda_8 \\ 0 & 0 & \lambda_9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \varsigma \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{pmatrix},$$

$$\Phi = \text{cov}(\xi) = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix},$$

where $\gamma_1, \gamma_2,$ and γ_3 are the direct path estimates of the first-order exogenous latent variables ξ_1, ξ_2, and latent interaction term $\xi_1\xi_2$, respectively, μ_y and the μ_{x_i} terms are the means of y and x variables, and $\lambda_i,\ i = 1,\cdots,9$ are the structural coefficients relating each latent variables in ξ to its measured variable indicators. The assumptions for $\xi_1,\ \xi_2,\ \varsigma$ and $\delta_i,\ i = 1,\cdots,9$ are as follows.

1. ξ_1, ξ_2, ς, and δ_i are mutually uncorrelated.
2. δ_i and δ_j are mutually uncorrelated.
3. All ξ_1, ξ_2, ς, and δ_i are normally distributed. That is,

$$(\xi_1, \xi_2) \sim \text{MVN}(\mathbf{0}, \begin{pmatrix} \phi_{11} & \\ \phi_{21} & \phi_{22} \end{pmatrix}),\ \varsigma \sim \text{N}(0, \sigma_\varsigma^2), \text{ and } \delta_i \sim \text{N}(0, \theta_{\delta i}),\ i = 1,\cdots,9.$$

Following from the above distributional assumptions, Kenny and Judd constrained the model in the following ways. The first constraint for the loadings of the indicators

on the latent interaction term was set to be equal to the product of the loadings for the two indicators which form the product indicator. That is,

$$\lambda_7 = \lambda_1\lambda_4, \quad \lambda_8 = \lambda_2\lambda_5, \quad \lambda_9 = \lambda_3\lambda_6.$$

The second constraint for the variance of the latent interaction variable was to constrain to be the sum of the product of the variances of ξ_1 and ξ_2 and the squared covariance between ξ_1 and ξ_2, which is

$$\phi_{33} = \phi_{11}\phi_{22} + \phi_{21}^2.$$

The third constraint is the covariance between the latent interaction variable and each of the first-order latent variables ξ_1 and ξ_2 is set to be zero based on the assumption of normality (Algina & Moulder, 2001; Jöreskog & Yang, 1996; Kenny & Judd, 1984). This means that

$$\phi_{31} = 0 \quad \text{and} \quad \phi_{32} = 0.$$

The last constraints for the error variances of the product indicators for the latent interaction term were constrained to be as follows.

$$\begin{aligned}
\theta_{\delta 7} &= \phi_{11}\theta_{\delta 4} + \phi_{22}\theta_{\delta 1} + \theta_{\delta 1}\theta_{\delta 4}, \\
\theta_{\delta 8} &= \lambda_2^2\phi_{11}\theta_{\delta 5} + \lambda_5^2\phi_{22}\theta_{\delta 2} + \theta_{\delta 2}\theta_{\delta 5}, \\
\theta_{\delta 9} &= \lambda_3^2\phi_{11}\theta_{\delta 6} + \lambda_6^2\phi_{22}\theta_{\delta 3} + \theta_{\delta 3}\theta_{\delta 6}.
\end{aligned} \tag{3}$$

Jöreskog and Yang (1996) proposed a general latent interaction model in a matrix form, which can be implemented in LISREL 8 (Jöreskog & Sörbom, 1996). They

suggested using uncentered variables and using only one product of the observed variables since it is sufficient to identify the model. The differences between the Jöreskog-Yang model and the Kenny-Judd model are the product indicators used for the latent variable interaction and the inclusion of the intercept term in the structural equation. Unlike the K-J model, the J-Y model did include the intercept term in the structural equation. Furthermore, the dependent variable in the structural equation in the Jöreskog-Yang model is latent whereas that in the Kenny-Judd model is observed. The assumptions for the Jöreskog-Yang model are the same as those for the Kenny-Judd model. It is worth noting that the intercept α in the structural equation and those τ_y terms in the measurement equations cannot be estimated simultaneously due to identification constraints (Algina & Moulder, 2001; Jöreskog & Yang, 1996; Marsh, Wen, & Hau, 2006). Therefore, the authors set $\alpha = 0$ in Equation 4 and estimated the intercept terms τ_{yi} in Equation 5. The LISREL specification for Jöreskog-Yang model can be written as follows.

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \varsigma\,, \tag{4}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_{y_1} \\ \tau_{y_2} \\ \tau_{y_3} \end{pmatrix} + \begin{pmatrix} 1 \\ \lambda_{y_2} \\ \lambda_{y_3} \end{pmatrix} \eta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}, \tag{5}$$

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y\eta + \boldsymbol{\varepsilon}\,,$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_1x_4 \\ x_2x_5 \\ x_3x_6 \end{pmatrix} = \begin{pmatrix} \tau_{x1} \\ \tau_{x2} \\ \tau_{x3} \\ \tau_{x4} \\ \tau_{x5} \\ \tau_{x6} \\ \tau_{x1}\tau_{x4} \\ \tau_{x2}\tau_{x5} \\ \tau_{x3}\tau_{x6} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ \lambda_1\tau_{x4} & \lambda_4\tau_{x1} & \lambda_7 \\ \tau_{x5}\lambda_2 & \tau_{x2}\lambda_5 & \lambda_8 \\ \tau_{x6}\lambda_3 & \tau_{x3}\lambda_6 & \lambda_9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{pmatrix}, \qquad (6)$$

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta},$$

$$\mathbf{K} = \mathrm{E}(\boldsymbol{\xi}) = \begin{pmatrix} 0 \\ 0 \\ \kappa_3 \end{pmatrix}, \quad \boldsymbol{\Phi} = \mathrm{cov}(\boldsymbol{\xi}) = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix},$$

$$\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} = diag(\theta_{\varepsilon1}, \theta_{\varepsilon2}, \theta_{\varepsilon3}),$$

$$\boldsymbol{\Theta}_{\boldsymbol{\delta}} = \begin{pmatrix} \theta_{\delta1} & & & & & & & & \\ 0 & \theta_{\delta2} & & & & & & & \\ 0 & 0 & \theta_{\delta3} & & & & & & \\ 0 & 0 & 0 & \theta_{\delta4} & & & & & \\ 0 & 0 & 0 & 0 & \theta_{\delta5} & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta6} & & & \\ \tau_{x_4}\theta_{\delta1} & 0 & 0 & \tau_{x_1}\theta_{\delta4} & 0 & 0 & \theta_{\delta7} & & \\ 0 & \tau_{x_5}\theta_{\delta2} & 0 & 0 & \tau_{x_2}\theta_{\delta5} & 0 & 0 & \theta_{\delta8} & \\ 0 & 0 & \tau_{x_6}\theta_{\delta3} & 0 & 0 & \tau_{x_3}\theta_{\delta6} & 0 & 0 & \theta_{\delta9} \end{pmatrix},$$

where $diag(\theta_{\varepsilon1}, \theta_{\varepsilon2}, \theta_{\varepsilon3})$ denotes a square matrix (3×3) with diagonal elements $\theta_{\varepsilon1}, \theta_{\varepsilon2}, \theta_{\varepsilon3}$ and off-diagonal elements 0. The derivation of $\boldsymbol{\Theta}_{\boldsymbol{\delta}}$ is included in Appendix A. If the intercept terms τ_{x_i} are equal to zero, then $\boldsymbol{\Theta}_{\boldsymbol{\delta}}$ will be reduced to a diagonal matrix. The nonlinear constraints for the associated parameters and normality constraints are as follows.

(1) loadings: $\lambda_7 = \lambda_1\lambda_4$, $\lambda_8 = \lambda_2\lambda_5$, $\lambda_9 = \lambda_3\lambda_6$.

(2) error variances: $\theta_{\delta 7} = \tau_1^2\theta_{\delta 4} + \tau_4^2\theta_{\delta 1} + \phi_{11}\theta_{\delta 4} + \phi_{22}\theta_{\delta 1} + \theta_{\delta 1}\theta_{\delta 4}$,

$$\theta_{\delta 8} = \tau_2^2\theta_{\delta 5} + \tau_5^2\theta_{\delta 2} + \lambda_2^2\phi_{11}\theta_{\delta 5} + \lambda_5^2\phi_{22}\theta_{\delta 2} + \theta_{\delta 2}\theta_{\delta 5},$$

$$\theta_{\delta 9} = \tau_3^2\theta_{\delta 6} + \tau_6^2\theta_{\delta 3} + \lambda_3^2\phi_{11}\theta_{\delta 6} + \lambda_6^2\phi_{22}\theta_{\delta 3} + \theta_{\delta 3}\theta_{\delta 6}.$$

(3) variance of $\xi_1\xi_2$: $\phi_{33} = \phi_{11}\phi_{22} + \phi_{21}^2$.

(4) normality: $\phi_{31} = 0, \ \phi_{32} = 0$.

(5) mean structure: $\kappa_3 = \phi_{21}$.

Jöreskog and Yang (1996) argued that the intercept α will not be zero even when $\xi_1, \xi_2,$ and ς have zero means. The rationale is based on the fact that the mean of the latent interaction variable, $\xi_1\xi_2$ (i.e., ϕ_{21}) will not typically have zero mean even when the indicators of ξ_1 and ξ_2 are mean-centered. Therefore, the mean structure is necessary unless $\hat{\phi}_{21}$ is very close to zero.

Algina and Moulder (2001) revised the Jöreskog-Yang model by mean-centering the indicators of the linear factors ξ_1 and ξ_2 before computing the product indicators, simplifying the nonlinear constraints since the intercept terms in the right-hand side of Equation 6 are all zero. Furthermore, the Algina-Moulder (A-M) model had a higher convergence rate than the Jöreskog-Yang model. The LISREL specification for the Algina-Moulder model is

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \varsigma,$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_{y_1} \\ \tau_{y_2} \\ \tau_{y_3} \end{pmatrix} + \begin{pmatrix} 1 \\ \lambda_{y_2} \\ \lambda_{y_3} \end{pmatrix} \eta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix},$$

$$\begin{pmatrix} x_1 - \tau_{x_1} \\ x_2 - \tau_{x_2} \\ x_3 - \tau_{x_3} \\ x_4 - \tau_{x_4} \\ x_5 - \tau_{x_5} \\ x_6 - \tau_{x_6} \\ (x_1 - \tau_{x_1})(x_4 - \tau_{x_4}) \\ (x_2 - \tau_{x_2})(x_5 - \tau_{x_5}) \\ (x_3 - \tau_{x_3})(x_6 - \tau_{x_6}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_8 \\ 0 & 0 & \lambda_9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{pmatrix},$$

$$\mathrm{K} = \mathrm{E}(\boldsymbol{\xi}) = \begin{pmatrix} 0 \\ 0 \\ \kappa_3 \end{pmatrix}, \quad \Phi = \mathrm{cov}(\boldsymbol{\xi}) = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix},$$

$$\Theta_\varepsilon = diag(\theta_{\varepsilon 1}, \theta_{\varepsilon 2}, \theta_{\varepsilon 3}),$$

$$\Theta_\delta = diag(\theta_{\delta 1}, \theta_{\delta 2}, \theta_{\delta 3}, \theta_{\delta 4}, \theta_{\delta 5}, \theta_{\delta 6}, \theta_{\delta 7}, \theta_{\delta 8}, \theta_{\delta 9}).$$

The constraints of the A-M model are the same as those for the Jöreskog-Yang model except those for error variances, which are the same as those for the Kenny-Judd model in Equation 3.

2.2.1.2 Unconstrained approach. Marsh et al. (2004) revised the constrained approach by allowing the nonlinear constraints used to specify the factor loadings and variances for the latent interaction term to be freely estimated. The assumptions for the unconstrained model are the same as those for the K-J model except that the normality assumption is not imposed on the first-order latent variables. They proposed mean-centering all independent variable indicators and using the matching strategy

previously discussed to formulate the product indicators for the latent interaction term. The matching strategy comes from a simulation study conducted by Marsh and colleges (2004) who determined that the matched-pair products have better estimation precision than all-possible products proposed by Kenny and Judd (1984) and one-pair product proposed by Jöreskog and Yang (1996). In the current study, the three matched pairs are x_1x_4, x_2x_5, and x_3x_6. In the unconstrained model, the intercept term in the structural equation is omitted since the intercept terms τ_y are included in the measurement model. The LISREL specification for the unconstrained approach is as follows.

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \varsigma\,,$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_{y1} \\ \tau_{y2} \\ \tau_{y3} \end{pmatrix} + \begin{pmatrix} 1 \\ \lambda_{y2} \\ \lambda_{y3} \end{pmatrix}\eta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix},$$

$$\begin{pmatrix} x_1 - \tau_{x_1} \\ x_2 - \tau_{x_2} \\ x_3 - \tau_{x_3} \\ x_4 - \tau_{x_4} \\ x_5 - \tau_{x_5} \\ x_6 - \tau_{x_6} \\ (x_1 - \tau_{x_1})(x_4 - \tau_{x_4}) \\ (x_2 - \tau_{x_2})(x_5 - \tau_{x_5}) \\ (x_3 - \tau_{x_3})(x_6 - \tau_{x_6}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ \lambda_3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_8 \\ 0 & 0 & \lambda_9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{pmatrix},$$

$$\mathbf{K} = \mathrm{E}(\boldsymbol{\xi}) = \begin{pmatrix} 0 \\ 0 \\ \phi_{21} \end{pmatrix}, \qquad \boldsymbol{\Phi} = \mathrm{cov}(\boldsymbol{\xi}) = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix},$$

$$\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} = diag(\theta_{\varepsilon 1}, \theta_{\varepsilon 2}, \theta_{\varepsilon 3})\,,$$

$$\boldsymbol{\Theta}_{\boldsymbol{\delta}} = diag(\theta_{\delta 1}, \theta_{\delta 2}, \theta_{\delta 3}, \theta_{\delta 4}, \theta_{\delta 5}, \theta_{\delta 6}, \theta_{\delta 7}, \theta_{\delta 8}, \theta_{\delta 9}) .$$

The unconstrained approach is selected for the current study among the product indicator approaches due to the following advantages. First, this approach is comparable with the constrained approach in bias, type I error and power. Second, it is easy for researchers to implement in that the complicated nonlinear constraints are not required. Third, this approach can be implemented in all SEM packages (March et al., 2006).

The disadvantages of the product indicator approaches are no solid theoretical foundation for how to formulate the product indicators, inconsistent conclusions which may be derived from using different product indicators, existing problems on the robustness of the maximum likelihood (ML) estimation to violation of multivariate normality, and difficulties in extending to more general nonlinear models (Cudeck, Harring & du Toit, 2009; Lee & Song, 2007; Wall, 2009).

2.3 Mixture Models

2.3.1 Introduction. Finite mixture distributions have become broadly used in applied sciences such as neural networks, behavioral science, and social sciences (Lubke & Muthén, 2005; McLachlan & Peel, 2000). Part of their popularity stems from the mixture distribution's flexibility for accommodating unobserved population heterogeneity in observed data, which cannot be captured satisfactorily by a single

parametric family. In addition, mixture models can be used to approximate complex distributions through a proper selection of the components (groups, subpopulations, and components will be used interchangeably in this paper) (McLachlan & Peel, 2000). As stated in McLachlan and Peel (2000), one of the pioneering studies in mixture models was undertaken by Karl Pearson (1894). He fit a mixture of two univariate normal distributions with different means and variances for data that consisted of observations on the ratio of forehead to body length of 1000 crabs from the Bay of Naples. Pearson's supposition was that the crabs had evolved toward two new subspecies, which could be uncovered by using a mixture analytic approach.

Mixture models and multiple group analyses are similar because they separate the data into groups and allow parameters to vary across the groups or constrain them to be equal. However, there are several differences between them. First, the number of subpopulations is known in multi-group analyses whereas the number of subpopulations is unknown in mixture models. In other words, the probability of each observation belonging to a subpopulation is either 0 or 1 since the grouping variable is observed in multi-group analyses. On the other hand, the probability of each observation's group membership is estimated by the posterior probability which represents the contribution of each observation to the estimated parameters in each subpopulation. Second, multi-group analyses estimate the mean and covariance

matrix in one population independently without considering the other subpopulations. However, the posterior means and covariance matrices of each subpopulation are not independent in mixture models because the parameters in each subpopulation are estimated by the weighted composites of the posterior probability of each individual belonging to a subpopulation (Bauer & Curran, 2003; Muthén & Muthén, 1998, Appendix 8). This correlation will therefore influence the estimation of the parameters common to all subpopulations. Third, it is not necessary to impose a normality assumption in multi-group analyses since identification is considered only with respect to the parameterization of the mean vector and covariance matrix for each subpopulation. On the other hand, a distributional assumption is needed to identify membership in a subpopulation in mixture models (Arminger & Stein, 1997).

2.3.2 Theoretical backgrounds for mixture models.

2.3.2.1 Definition. A distribution consisting of two or more components where the component membership for each observation is unobserved and needs to be inferred from the data is called a finite mixture distribution. Typically, it is assumed that the components arise from the same parametric family with parameters that are allowed to vary across components or latent classes and have pre-determined distributions (McLachlan & Peel, 2000).

The density of a K-component finite mixture distribution is defined as

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i)$$

where $\mathbf{y}_1, \cdots, \mathbf{y}_n$ denotes an observed random sample of size n, $\mathbf{y}_i$ is a p-dimensional random vector with probability density function $f(\mathbf{y}_i)$ on $\mathbf{R}^p$ and π_i satisfying $0 \le \pi_k \le 1$ for all $k = 1, \cdots, K$ and $\sum_{k=1}^{K} \pi_k = 1$. The quantities $\pi_1, \cdots, \pi_K$ are called mixing proportions or weights. $f_k(\mathbf{y}_i)$ is the component density. The densities of mixtures of univariate and multivariate distributions as well as linear structural equation models are specified as follows.

2.3.2.2 Mixtures of univariate distributions. Let y_i denote the value of the observed variable for individual i in a sample of size n and K denote the number of mixture components. A function f is a mixture of K component distributions $f_1, \cdots, f_K$ if

$$f(y_i \mid \boldsymbol{\varphi}) = \sum_{k=1}^{K} \pi_k f_k(y_i \mid \boldsymbol{\theta}_k),$$

where $\boldsymbol{\varphi} = (\boldsymbol{\pi}', \boldsymbol{\theta}')'$ is the vector of all unknown parameters, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_K')'$ is the vector of the unknown parameters in all subpopulations, $\boldsymbol{\theta}_k$ is the vector of the unknown parameters for the k th subpopulation, and $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)'$ is the vector of probability of belonging to the subpopulation k with the elements satisfying $0 \le \pi_k \le 1$ for all $k = 1, \cdots, K$ and $\sum_{k=1}^{K} \pi_k = 1$. One of the most commonly used univariate mixture distributions is, for example, a mixture of two univariate normal distributions. Its density can be written as

$$f(y_i \mid \varphi) = \pi_1 \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{ -\frac{(y_i - \mu_1)^2}{2\sigma_1^2} \right\} + \pi_2 \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{ -\frac{(y_i - \mu_2)^2}{2\sigma_2^2} \right\}$$

where $\boldsymbol{\varphi} = (\pi_1, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)'$ and $\pi_2 = 1 - \pi_1$. The components of a mixture of univariate normal distributions may be homoscedastic (equal variance) or heteroscedastic (unequal variances). Figure 3 presents a mixture of two heteroscedastic normal distributions.

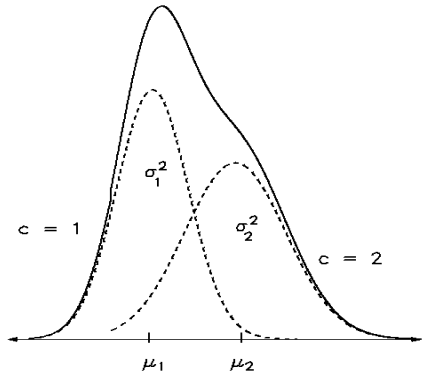


Figure 3. A mixture of two univariate normal distributions with unequal variances, where c denotes the class.

2.3.2.3 Mixtures of multivariate distributions. If $\mathbf{y}_i$ consists of multiple (p) (normal) random variables with unobserved subpopulation membership, then a mixture of multivariate (normal) distributions is applicable. Let $\mathbf{y}_i$ be a $(p \times 1)$ vector containing values for individual i in a sample of size n on a set of p observed continuous random variables and K be the number of mixture components.

A distribution f is a mixture of K component densities $f_1, \cdots, f_K$ if

$$f(\mathbf{y}_i \mid \boldsymbol{\varphi}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k), \quad i = 1, \cdots, n,$$

where $\boldsymbol{\varphi} = (\boldsymbol{\pi}', \boldsymbol{\theta}')'$ is the vector of all unknown parameters, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_K')'$ is the vector of the unknown parameters in all the subpopulations, $\boldsymbol{\theta}_k$ is the vector of the unknown parameters for the k-th subpopulation, and $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)'$ is the vector of probability of belonging to the k-th subpopulation. Specifically, the k-th component density of a mixture of multivariate normal distributions is given by

$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left| \boldsymbol{\Sigma}_k \right|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k) \right\},$$

where $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k', (vech\, \boldsymbol{\Sigma}_k)')'$ and vech($\boldsymbol{\Sigma}_k$) denotes a half-vectorization of a symmetric matrix $\boldsymbol{\Sigma}_k$ by stacking only the lower triangular part of $\boldsymbol{\Sigma}_k$. Figure 4 illustrates a mixture of four bivariate normal distributions with class-specific mean structures and covariance structures.

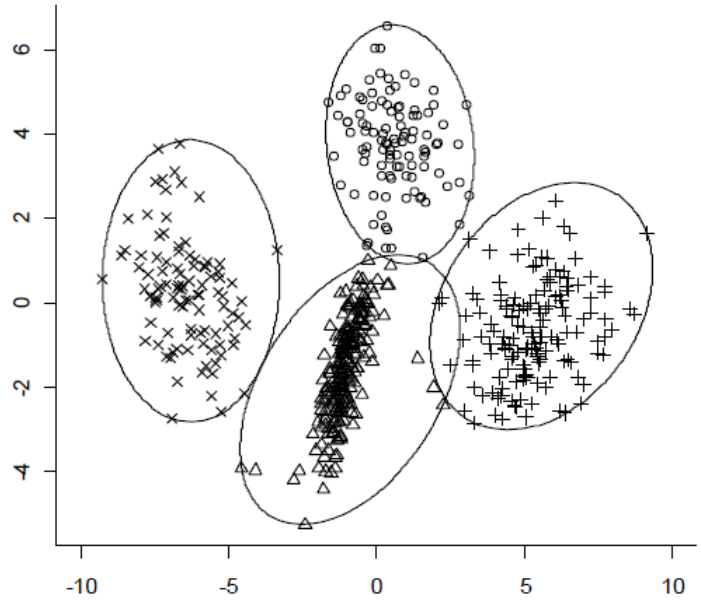


Figure 4. A mixture of four bivariate normal distribution with different mean structures and different covariance structures.

2.3.2.4 Mixture linear structural equation models. Mixture linear structural equation models (SEMs) are a combination of finite mixture modeling (or latent class models) and structural equation modeling. They were proposed by Arminger and Stein (1997), Dolan and van der Maas (1997), and Jedidi, Jagpal, and DeSarbo (1997). Following the concept and notations in Jedidi et al. (1997), a general mixture SEM given K components is reviewed.

Let $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)'$ be a vector of mixing proportions satisfying $0 \le \pi_k \le 1$ for all $k = 1, \cdots, K$ and $\sum_{k=1}^{K} \pi_k = 1$. Let $\mathbf{y}_i \mid k$ be a $p \times 1$ vector of observed variables measuring the endogenous latent variables $\boldsymbol{\eta}_{ik}$ $(m \times 1)$ for individual i in class k and $\mathbf{x}_i \mid k$ be a $q \times 1$ vector of observed variables measuring the exogenous latent variables $\boldsymbol{\xi}_{ik}(n \times 1)$ for individual i in class k. The measurement equations for the mixture structural equation model are

$$\mathbf{y}_i \mid k = \boldsymbol{\nu}_{yk} + \boldsymbol{\Lambda}_{yk} \boldsymbol{\eta}_{ik} + \boldsymbol{\varepsilon}_{ik},$$

$$\mathbf{x}_i \mid k = \boldsymbol{\nu}_{xk} + \boldsymbol{\Lambda}_{xk} \boldsymbol{\xi}_{ik} + \boldsymbol{\delta}_{ik},$$

where $\boldsymbol{\nu}_{yk}(p \times 1)$ and $\boldsymbol{\nu}_{xk}(q \times 1)$ are vectors of intercepts for class k, $\boldsymbol{\Lambda}_{yk}(p \times m)$ and $\boldsymbol{\Lambda}_{xk}(q \times n)$ are matrices with factor loadings for class k, $\boldsymbol{\varepsilon}_{ik}(p \times 1)$ and $\boldsymbol{\delta}_{ik}(q \times 1)$ are vectors of measurement errors in $\mathbf{y}_i \mid k$ and $\mathbf{x}_i \mid k$, respectively. $\boldsymbol{\varepsilon}_{ik}$ and $\boldsymbol{\delta}_{ik}$ are assumed to have zero mean vectors and have covariance matrices $\boldsymbol{\Theta}_{\varepsilon k}$ and $\boldsymbol{\Theta}_{\delta k}$, respectively, where $\boldsymbol{\Theta}_{\varepsilon k}$ and $\boldsymbol{\Theta}_{\delta k}$ are not necessarily diagonal. It is

also assumed that $\boldsymbol{\varepsilon}_{ik}$ and $\boldsymbol{\delta}_{ik}$ are uncorrelated with the latent variables. Define the mean vector and covariance matrix for $\boldsymbol{\xi}_{ik}$ to be $\boldsymbol{\kappa}_k$ and $\boldsymbol{\Phi}_k$, respectively. The structural equation of the linear structural equation is

$$\mathbf{B}_k \boldsymbol{\eta}_{ik} = \boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_{ik} + \boldsymbol{\varsigma}_{ik},$$

where $\mathbf{B}_k$ is an $m \times m$ nonsingular matrix of structural parameters specifying the relations among the endogenous latent variables for class k, $\boldsymbol{\alpha}_k$ is an $m \times 1$ vector of factor intercepts for class k, $\boldsymbol{\Gamma}_k$ is an $m \times n$ coefficient matrix for regressions of $\boldsymbol{\eta}_{ik}$ on $\boldsymbol{\xi}_{ik}$, and $\boldsymbol{\varsigma}_{ik}$ is an $m \times 1$ vector of random disturbances for class k with zero mean vector and covariance matrix $\boldsymbol{\Psi}_k$. It is assumed that vectors $\boldsymbol{\varsigma}_{ik}$ and $\boldsymbol{\xi}_{ik}$ are uncorrelated. Then, the conditional mean vectors $\boldsymbol{\mu}_k$ and conditional covariance matrices $\boldsymbol{\Sigma}_k$ are given by

$$\boldsymbol{\mu}_k = \begin{pmatrix} \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_{yk} \mathbf{B}_k^{-1} (\boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \boldsymbol{\kappa}_k) \\ \boldsymbol{\nu}_{xk} + \boldsymbol{\Lambda}_{xk} \boldsymbol{\tau}_k \end{pmatrix}, \tag{7}$$

and

$$\boldsymbol{\Sigma}_k = \begin{pmatrix} \boldsymbol{\Lambda}_{yk} \mathbf{B}_k^{-1} (\boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_k' + \boldsymbol{\Psi}_k) \mathbf{B}_k^{-1\prime} \boldsymbol{\Lambda}_{yk}' + \boldsymbol{\Theta}_{k\varepsilon} & \boldsymbol{\Lambda}_{yk} \mathbf{B}_k^{-1} \boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_{xk}' \\ \boldsymbol{\Lambda}_{xk} \boldsymbol{\Theta}_k \boldsymbol{\Gamma}_k' \mathbf{B}_k^{-1\prime} \boldsymbol{\Lambda}_{yk}' & \boldsymbol{\Lambda}_{xk} \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_{xk}' + \boldsymbol{\Theta}_{\delta k} \end{pmatrix}. \tag{8}$$

Let $\mathbf{z}_i \mid k = (\mathbf{y}_i' \mid k, \mathbf{x}_i' \mid k_i)'$. It is assumed that $\mathbf{z}_i \mid k$ has a conditional multivariate normal

distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. The density of $\mathbf{z}_i$ is given by

$$f(\mathbf{z}_i \mid \boldsymbol{\varphi}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k),\ i = 1, \cdots, n,$$
$$= \sum_{k=1}^{K} \pi_k \left[(2\pi)^{-(p+q)/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_k) \right\} \right], \quad (9)$$

where $\boldsymbol{\varphi} = (\boldsymbol{\pi}', \boldsymbol{\theta}')'$. Let $\mathbf{B} = (vec\mathbf{B}_1, \cdots, vec\mathbf{B}_K)$, $\boldsymbol{\Gamma} = (vec\,\boldsymbol{\Gamma}_1, \cdots, vec\,\boldsymbol{\Gamma}_K)$, $\boldsymbol{\Lambda}_x = (vec\boldsymbol{\Lambda}_{x1}, \cdots vec\boldsymbol{\Lambda}_{xK})$, $\boldsymbol{\Lambda}_y = (vec\boldsymbol{\Lambda}_{y1}, \cdots vec\boldsymbol{\Lambda}_{yK})$, $\boldsymbol{\nu}_x = (\boldsymbol{\nu}_{x1}', \cdots, \boldsymbol{\nu}_{xK}')'$, $\boldsymbol{\nu}_y = (\boldsymbol{\nu}_{y1}', \cdots, \boldsymbol{\nu}_{yK}')'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \cdots, \boldsymbol{\alpha}_K')'$, $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1', \cdots, \boldsymbol{\kappa}_K')'$, $\boldsymbol{\Phi} = (vech\,\boldsymbol{\Phi}_1, \cdots, vech\,\boldsymbol{\Phi}_K)$, $\boldsymbol{\Psi} = (vech\,\boldsymbol{\Psi}_1, \cdots, vech\,\boldsymbol{\Psi}_K)$, $\boldsymbol{\Theta}_\delta = (vec\,\boldsymbol{\Theta}_{\delta 1}, \cdots, vec\boldsymbol{\Theta}_{\delta K})$, $\boldsymbol{\Theta}_\varepsilon = (vec\,\boldsymbol{\Theta}_{\varepsilon 1}, \cdots, vec\,\boldsymbol{\Theta}_{\varepsilon K})$, where vec($\mathbf{A}$) denotes vectorization of a matrix $\mathbf{A}$ obtained by stacking the columns of $\mathbf{A}$ on the top of one another and vech($\mathbf{A}$) denotes a half-vectorization of a symmetric matrix $\mathbf{A}$ by stacking only the lower triangular part of $\mathbf{A}$.

$\boldsymbol{\theta} = (\mathbf{B}', \boldsymbol{\Gamma}', \boldsymbol{\Lambda}_x', \boldsymbol{\Lambda}_y', \boldsymbol{\nu}_x', \boldsymbol{\nu}_y', \boldsymbol{\alpha}', \boldsymbol{\kappa}', \boldsymbol{\Phi}', \boldsymbol{\Psi}', \boldsymbol{\Theta}_\delta', \boldsymbol{\Theta}_\varepsilon')'$ is a column vector containing all the unknown parameters. It is assumed that $\left|\boldsymbol{\Sigma}_k\right| > 0$. Therefore, the loglikelihood function for mixture linear structural equation models, which is often referred to as the *incomplete-data log likelihood function* due to the lack of population classification information, is

$$l(\boldsymbol{\varphi} \mid \mathbf{z}_i) = \log\left(\prod_{i=1}^{n} f(\mathbf{z}_i \mid \boldsymbol{\varphi}) \right)$$
$$= \log\left(\prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$
$$= \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k (2\pi)^{-(p+q)/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_k) \right\} \right). \quad (10)$$

A latent interaction model using the unconstrained approach can be viewed as a linear structural equation model since the latent interaction variable is formulated and inferred by the matched-pair product indicators of the linear factors. Therefore, the maximum of the loglikelihood function of the mixture latent variable interaction model has a closed form (Jedidi et al., 1997, Yung, 1997). However, maximizing the loglikelihood directly in Equation 10 is cumbersome. For example, calculating the first and second derivatives for the gradient and Hessian matrix, respectively, required for the Newton-Raphson method is tedious with the sum inside the logarithm. To circumvent the heavy direct computation, the expectation-maximization (EM) algorithm provides an indirect approach to obtain the maximum likelihood estimates (MLEs) for the parameters, which are the values most likely generating the given data (Wall, 2009).

2.3.3 Estimation.

2.3.3.1 The EM algorithm. The theoretical foundations and applications of the EM algorithm were not well established, formulated, and investigated until the seminal paper was published by Dempster, Laird, and Rubin (1977). The EM algorithm is an iterative procedure for finding MLEs for incomplete data or missing data. It is a broadly applicable approach since it simplifies ML estimation substantially by reformulating the given incomplete-data problem as a complete-data

problem (McLachlan & Krishnan, 2008).

2.3.3.2 Complete-data log likelihood. In mixture models, the population heterogeneity is captured by a latent categorical variable. Because of a lack of population classification information, the observed data $\mathbf{z}_i$ in the mixture model can be treated as incomplete – the missing components are the classification information. Adding the unobserved population indicator variables $\mathbf{c}_i$ to the observed data $\mathbf{z}_i$ makes the data complete and also facilitates the computation for MLE. Before the derivation of the complete-data likelihood function, define an unobservable zero-one indicator vector $\mathbf{c} = (\mathbf{c}_1', \cdots, \mathbf{c}_n')'$, where $\mathbf{c}_i$ is a $K \times 1$ vector and

$$c_{ik} = \begin{cases} 1, & \text{if the } i\text{th subject belongs to class } k \\ 0, & \text{otherwise} \end{cases}, \quad i = 1, \cdots, n,\ k = 1, \cdots, K .$$

This implies that $\Pr(c_{ik} = 1) = \mathrm{E}(c_{ik}) = \pi_k$, which is usually referred to as the prior probability of belonging to class *k*. In addition, $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n$ have an independent and identical multinomial distribution consisting of *K* categories with probabilities $\pi_1, \pi_2, \cdots, \pi_K$, respectively. That is, the density of $\mathbf{z}_i$ given $\mathbf{c}_i$ is observed is

$$f(\mathbf{z}_i \mid \mathbf{c}_i) = \prod_{k=1}^{K} \{\pi_k f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}^{c_{ik}} .$$

Therefore, the joint density of $\mathbf{z}_i$ given $\mathbf{c}_i$ can be factorized into the product of the conditional densities because $\mathbf{z}_i$'s are mutually independent , that is

$$\begin{aligned} f(\mathbf{z}_1, \cdots, \mathbf{z}_n \mid \mathbf{c}_1, \cdots, \mathbf{c}_n) &= \prod_{i=1}^{n} f(\mathbf{z}_i \mid \mathbf{c}_i, \boldsymbol{\varphi}) \\ &= \prod_{i=1}^{n} \prod_{k=1}^{K} \{\pi_k f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}^{c_{ik}} . \end{aligned}$$

The log likelihood function with "complete" data referred to as the *complete data loglikelihood function* can be written as

$$
\begin{aligned}
1_C(\boldsymbol{\varphi} \mid \mathbf{z}_i, \mathbf{c}_i) &= \log L_C(\boldsymbol{\varphi} \mid \mathbf{z}_i, \mathbf{c}_i) \\
&= \log \prod_{i=1}^{n} f(\mathbf{z}_i \mid \mathbf{c}_i, \boldsymbol{\varphi}) \\
&= \log \prod_{i=1}^{n} \prod_{k=1}^{K} \{\pi_k f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k)\}^{c_{ik}} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \left[\log \pi_k + \log f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right] \qquad (11) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \left\{\log \pi_k + \left[-\frac{p+q}{2} \log 2\pi - \frac{1}{2} \log|\boldsymbol{\Sigma}_k| - \frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_k)\right]\right\} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \log \pi_k - \frac{p+q}{2} \log 2\pi \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} \log|\boldsymbol{\Sigma}_k| \\
&\quad - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{K} c_{ik} (\mathbf{z}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_k). \qquad (12)
\end{aligned}
$$

2.3.3.3 ML estimation via the EM algorithm. In order to avoid a direct maximization approach to find the MLEs implied in Equation 10, the EM algorithm adds the unobservable zero-one indicator variables to form the complete-data likelihood function $l_C(\boldsymbol{\varphi} \mid \mathbf{z}, \mathbf{c})$ which can be broken down into two parts. The algorithm proceeds to then approximate the roots of $l_C(\boldsymbol{\varphi} \mid \mathbf{z}, \mathbf{c})$ iteratively in a two-step process. The first step, the E-step, computes the conditional expectation of the complete-data loglikelihood, $l_C(\boldsymbol{\varphi} \mid \mathbf{z}, \mathbf{c})$, given the observed data $\mathbf{z}$ and current value of $\boldsymbol{\varphi}$. That is,

$$
\begin{aligned}
Q(\boldsymbol{\varphi} \mid \boldsymbol{\varphi}^{(m)}) &= \mathrm{E}\left[l_C(\boldsymbol{\varphi} \mid \mathbf{z}, \mathbf{c}) \mid \mathbf{z}, \boldsymbol{\varphi}^{(m)}\right] \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} E\left(c_{ik} \mid \mathbf{z}, \boldsymbol{\varphi}^{(m)}\right) \{\log \pi_k + \log[f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k)]\} \qquad (13)
\end{aligned}
$$

where $\boldsymbol{\varphi}^{(m)}$ denotes the value of $\boldsymbol{\varphi}$ at iteration m,

$$f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-(p+q)/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z}_i - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_k)\right\},$$

and

$$E(c_{ik} \mid \mathbf{z}, \boldsymbol{\varphi}^{(m)}) = \Pr(c_{ik} = 1 \mid \mathbf{z}) = \frac{\pi_k^{(m)} f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k^{(m)}, \boldsymbol{\Sigma}_k^{(m)})}{\sum_{k=1}^{K} \pi_k^{(m)} f_k(\mathbf{z}_i \mid \boldsymbol{\mu}_k^{(m)}, \boldsymbol{\Sigma}_k^{(m)})} = \hat{P}_{ik}(\boldsymbol{\varphi}^{(m)})$$

is the posterior probability of the ith subject with the observed value $\mathbf{z}_i$ belonging to the kth class of the mixture by Bayes' theorem. The E-step is equivalent to replacing the unobserved data c_{ik} with the posterior probabilities P_{ik} based on the current values of the parameters. Therefore, the conditional expectation of $l_C(\boldsymbol{\varphi} \mid \mathbf{z}, \mathbf{c})$ can be expressed (ignoring the constant term without loss of generality) as

$$\begin{aligned} Q(\boldsymbol{\varphi} \mid \boldsymbol{\varphi}^{(m)}) &= Q_1(\boldsymbol{\pi} \mid \boldsymbol{\varphi}^{(m)}) + Q_2(\boldsymbol{\theta}_k \mid \boldsymbol{\varphi}^{(m)}) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log \pi_k + \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log\left[f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k)\right] \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log \pi_k - \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log\left|\boldsymbol{\Sigma}_k\right| + \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} T_k \right) \end{aligned}$$

where $Q_1(\boldsymbol{\pi} \mid \boldsymbol{\varphi}^{(m)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log \pi_k$, $Q_2(\boldsymbol{\theta}_k \mid \boldsymbol{\varphi}^{(m)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{P}_{ik} \log\left[f_k(\mathbf{z}_i \mid \boldsymbol{\theta}_k)\right]$, and $T_k = (\mathbf{z}_i - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_k)$.

The second step, the M-step, finds the MLEs for Q_1 and Q_2 by maximizing Q_1 and Q_2 with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\theta}_k$, respectively, and thereby deriving the updated parameter vector $\boldsymbol{\varphi}^{(m+1)}$. The MLE of $\boldsymbol{\pi}_k$ at iteration $(m+1)$ can be calculated by averaging the posterior probabilities of all the subjects, that is,

$$\hat{\pi}_k^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{P}_{ik}(\boldsymbol{\varphi}^{(m)})}{n}.$$

Q_2, which is independent of $\boldsymbol{\pi}$, can be maximized with respect to $\boldsymbol{\theta}$ by numerical optimization procedures (e.g., Newton-Raphson method, Fisher's scoring method or Powell's (1977) conjugate gradient iterative method). The updated estimate of $\hat{\boldsymbol{\theta}}$ (i.e. $\hat{\boldsymbol{\mu}}_k^{(m+1)}, \hat{\boldsymbol{\Sigma}}_k^{(m+1)}$) is used in the E-step for the next iteration to compute a new conditional expectation which will be again maximized. This procedure continues until the difference of $l_C(\boldsymbol{\varphi}^{(m+1)}) - l_C(\boldsymbol{\varphi}^{(m)})$ is within an arbitrarily small range. The details of the EM algorithm are included in Harring (2010), McLachlan and Peel (2000), and Wall (2009). If the EM sequence $\{\hat{\boldsymbol{\varphi}}^{(m)}\}$ converges to $\hat{\boldsymbol{\varphi}}$, then $\hat{\boldsymbol{\varphi}}$ has been proved to be the MLE for the incomplete data log likelihood function in Equation 10 under mild regularity conditions(McLachlan & Peel, 2000; Wu, 1983).

2.3.3.4. Computational considerations. It is known that using the EM algorithm for the fitting of mixture models has the following computational liabilities. First, the convergence of the EM algorithm is slow, which will be exacerbated by a poor selection of starting values. If the starting values are too close to the boundary of the parameter space, the sequence $\{\hat{\boldsymbol{\theta}}^{(m)}\}$ generated by the EM algorithm may diverge.

Second, multiple local maxima may occur in mixture problems. If the likelihood function is bounded over the parameter space, then the global maximum will exist. However, the likelihood function for a mixture model is often unbounded and has

multiple local maximizers. The global maximizer $\hat{\boldsymbol{\varphi}}_{ML}$ (i.e., the MLE) usually corresponds to a local maximizer. One way to identify $\hat{\boldsymbol{\varphi}}_{ML}$ is to locate all the roots of the likelihood equation and then choose the root with the largest local maximum. Therefore, it is recommended to have a wide variety of starting points to search for all local maxima and to avoid nonconvergence (McLachlan & Basford, 1988; McLachlan & Peel, 2000).

The EM algorithm is broadly used in optimization problems due to its convenient implementation and numerical stability (Heath, 2007). Although it ensures that the likelihood function is increasing from iteration to iteration (Dempster et al., 1977), the EM algorithm may not derive global optimal solutions to mixture models. Not many procedures modify the EM algorithm regarding to the global optimization qualities so far. Two methods in the operations research literature- Cross-Entropy (CE) method (De Boer, Kroese, Mannor, & Rubinstein, 2005) and Model Reference Adaptive Search (MRAS) have been proposed to achieve global solutions to mixture models which have a problem of simulating the positive definite covariance matrices efficiently. To resolve this issue, Heath (2007) modified these two methods based on the EM algorithm and Cholesky decomposition, which result in four new algorithms- CE-EM, MRAS-EM, CE-CD and MRAS-CD. He also proved that MRAS-CD achieves the global optimal solutions to normal mixture models. However, these

algorithms have not been implemented in SEM software.

Another option to examine whether the solutions are global or local is to use the OPTSEED syntax on the seed numbers from the best log-likelihood values in Mplus (Jung & Wickrama, 2007). Suppose there are two seeds that yield the best log-likelihood values. If the results of these two outputs are the same, then the solution is not a local solution. However, this may not be manageable in simulation studies since the seed numbers have to be manually typed in for each replication.

Lastly, label switching often occurs in mixture models. This problem occurs if the log-likelihood remains invariant by permuting the mixture components (Redner & Walker, 1984). The use of true parameter values as starting values, model constraints, or the examination of parameter estimates may prevent label-switch problems to some extent (Tueller, Drotar, & Lubke, 2011). Label-switching is a problem for mixture models using Bayesian estimation, but it is not a problem using ML estimation in Mplus if the way to identify the constraints across classes has been built in the programs (L. Muthen, personal communication, April, 22, 2011).

2.3.3.5 Implementation in Mplus. As the unconstrained approach linearizes the latent interaction factor, a latent interaction mixture model can be estimated using the mixture module in Mplus 6.2 (Muthén & Muthén, 2009). There are two stages in Mplus ML optimization. In the initial stage, an optimization is performed 10 iterations

for each of the 10 sets (default) of random generated starting values. In the final stage, two (default) solutions with the largest loglikelihoods are selected as the starting values to be iterated until they converge to the same (best) likelihood. If the best loglikelihood value is not reached by at least 2 final-stage optimizations, then a warning message of possible local maximum will be given in the output. The STARTS=20 5 option can be used to change the number of starting values in the initial stage from 10 (default) to 20 and the number of optimizations from 2 (default) to 5. The STITERATION option allows for setting the maximum number of iteration in the initial stage. The default is 10. The STCONVERGENCE option allows for specifying the value of derivative convergence criteria in the initial stage. The default is one (Muthén & Muthén, 1998-2010).

The literature review covers the previous work on latent interaction models and mixture models. However, incorporating latent classes into a latent interaction model has not been widely investigated. Therefore, this study is to estimate the mixture latent interaction model using an unconstrained product indicator approach, to assess the estimation performance consisting of parameter recovery and data classification, and to examine model selection for (mixture) latent interaction models. The methodology will be detailed in Chapter 3.

Chapter 3: Methodology

In this chapter, the first research question will be answered in Section 3.1. The design of the simulation study, including the population model, both manipulated factors and invariant factors, and data generation will be discussed in Section 3.2. Criteria measures will be defined in Section 3.3 and preliminary analyses will also be included in Section 3.4. An analysis plan will be articulated in Section 3.5 to dissect which factors or combinations of factors have the greatest impact on parameter recovery and classification results.

3.1 Estimation Method

The current study uses the unconstrained approach (Marsh et al., 2004) to estimate the mixture latent interaction models in that it considers the latent interaction factor to be linear which is formulated by the product of the observed variables of the linear factors and the mixture module in existing SEM software (e.g., Mplus) can be used to provide a workable solution. Furthermore, it has the advantages of its comparable results to the constrained approach, ease of specification (i.e., no nonlinear constraints needed), and ease of implementation in SEM packages.

3.2 Design of the Study

Monte Carlo simulation is used to derive empirical sampling distributions by generating random samples from populations with known parameter values (Bandalos,

2006; Mooney, 1997). Among its many uses, a simulation study helps methodologists investigate how different design conditions influence outcome measures of interest, particularly for statistical problems which have no analytic solutions.

3.2.1 Population model. All variables are simulated according to the mixture of latent variable interaction model which consists of two latent classes: one with interaction whereas the other without interaction. The path diagram of this model is depicted in Figure 5. The model is written as follows.

$$\eta_i^{(k)} = \gamma_0^{(k)} + \gamma_1^{(k)}\xi_{1i}^{(k)} + \gamma_2^{(k)}\xi_{2i}^{(k)} + \gamma_3^{(k)}\xi_{1i}^{(k)}\xi_{2i}^{(k)} + \zeta_i^{(k)}, \quad i = 1, \cdots, n,\ k = 1, 2, \tag{14}$$

where k =1, 2 is the latent class, $\xi = \begin{pmatrix} \xi_{1i} \\ \xi_{2i} \end{pmatrix} = \begin{pmatrix} \pi_1\xi_{1i}^{(1)} + (1-\pi_1)\xi_{1i}^{(2)} \\ \pi_1\xi_{2i}^{(1)} + (1-\pi_1)\xi_{2i}^{(2)} \end{pmatrix} \sim (\boldsymbol{\mu}, \Sigma)$,

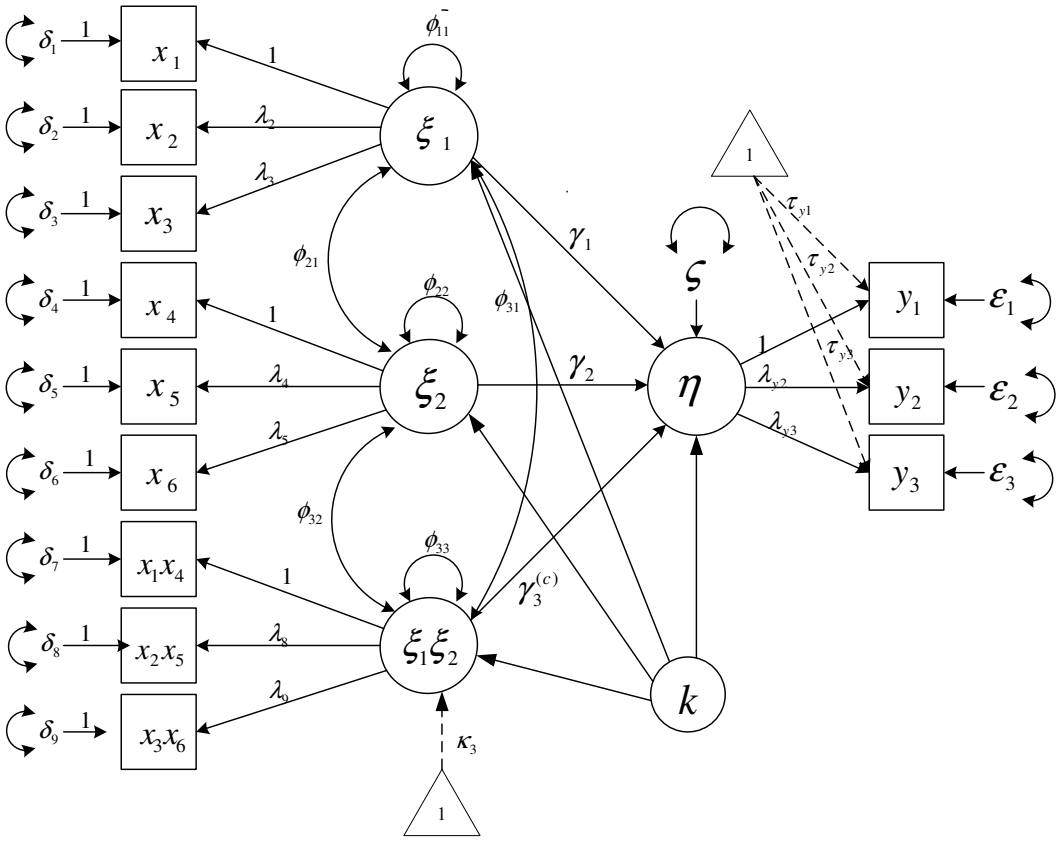


Figure 5. Path diagram of the mixture latent variable interaction model.

$\xi_{pi}^{(k)} \sim N(\mu_p^{(k)},1),\ p=1,2,$ and $\Sigma = \begin{pmatrix} 1 & \\ \phi_{12} & 1 \end{pmatrix}$ is assumed to be invariance across classes. Therefore, $\boldsymbol{\mu} = \pi_1 \boldsymbol{\mu}^{(1)} + (1-\pi_1)\boldsymbol{\mu}^{(2)} = \pi_1 \begin{pmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \end{pmatrix} + (1-\pi_1) \begin{pmatrix} \mu_1^{(2)} \\ \mu_2^{(2)} \end{pmatrix}$. In addition, it is assumed that $\zeta^{(k)} \sim N(0, \theta_\zeta^{(k)})$, with $\theta_\zeta^{(k)}$ set to be 10 for $k=1,2$. Without loss of generality, the exogenous factor mean vector of class one, $\boldsymbol{\mu}^{(1)} = \begin{pmatrix} \mu_1^{(1)} & \mu_2^{(1)} \end{pmatrix}'$, is fixed to be a zero vector. $\gamma_3^{(1)}$ is always set to 0 in the current study and $\gamma_3^{(2)}$ varies depending on the magnitude of the effect size of the interaction effects (i.e., $R^2_{\gamma_3^{(2)}}$). That is, class 1 is normally distributed whereas class 2 is non-normally distributed. The current study sets the paths of first-order factors as well as the intercepts in the data to be equal across classes (i.e., $\gamma_0^{(1)} = \gamma_0^{(2)} = 10,\ \gamma_1^{(1)} = \gamma_1^{(2)} = 2,$ and $\gamma_2^{(1)} = \gamma_2^{(2)} = 3$), focusing on the examination of class-specific latent interaction effects. The measurement models for both classes are assumed to be invariant across classes and are the same as those in Equation 5 and 6. The reliabilities and the loadings are set to be 0.9 and 0.7 for all indicators in the model, respectively. The errors, $\varepsilon_1, \varepsilon_2, \varepsilon_3$ and $\delta_1, \cdots, \delta_9$ will be generated as zero-mean normal variables. The error variances will be selected so that the reliability of each indicator follows the levels stated above.

3.2.2 Manipulated factors. The performance of the mixture of latent variable interaction models is influenced by many factors. The manipulated factors are chosen

partly from the latent variable interaction literature, such as the effect size and the covariance between two exogenous factors, and partly from the mixture models, such as class separation between two exogenous factor means, the covariance between two exogenous factors, and class proportions. Table 1 presents the levels of the factors manipulated in the current study. The factors that are held constant across classes are loadings, reliabilities, error variances, and exogenous factor variances.

Table 1. *Conditions of the manipulated factors*.

Manipulated factors	Conditions
1. Class separation (CS)	MD=0, 1.5, 3
2. $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})$ (ES)	(0, 0.05), (0, 0.15), (0, 0.25)
3. Covariance ϕ_{12} (COV)	0.2, 0.6
4. Sample sizes (SS)	500, 1000, 2000
5. Class proportion $(\pi_1, 1-\pi_1)$ (CP)	$0.5/0.5$, $0.2/0.8$

Class separation has been recognized as an important factor in class assignment and class comparison in mixture models (e.g., Lubke & Muthén, 2007; Tofighi & Enders, 2008). Larger class separation increases correct class assignment, but it is difficult to quantify the distance between classes, which is needed to yield a satisfactory probability of assigning subjects to their true class or to recover the true model parameters (Lubke & Muthén, 2007). Referring to Lubke and Neale (2006) and Lubke and Muthén (2007), the current study is carried out at three different levels of class separation due to the exogenous factor means: 0, 1.5, and 3, which are measured

by Mahalanobis distance (MD) for the factors. MD is defined as follows:

$$MD = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}),$$

where $\boldsymbol{\mu}^{(1)} = \left(\mu_1^{(1)} \quad \mu_2^{(1)})\right)'$, $\boldsymbol{\mu}^{(2)} = \left(\mu_1^{(2)} \quad \mu_2^{(2)}\right)'$, and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \phi_{12} \\ \phi_{12} & 1 \end{pmatrix}$ is the covariance matrix assumed to be equal across classes. The MD incorporates the correlations of the data set and is scale-invariant. When $\phi_{12} = 0$, the covariance matrix reduces to the identity matrix, and the MD reduces to the squared Euclidean distance. The population parameter vectors of $\boldsymbol{\mu}^{(2)}$ for different conditions are given in Appendix B.

The covariance between the two first-order latent variable ξ_1 and ξ_2 is often manipulated in past simulation studies that investigate latent interaction effects. The values commonly used ranged from 0.2 to 0.4 (e.g., Jaccard & Wan, 1995; Klein & Muthén, 2007; Marsh et al., 2004; Moulder & Algina, 2002). Jaccard and Wan (1995) reviewed social science literature and reported that typically the correlation values were 0.2 and 0.4. However, there is not much difference in the factor means for both classes when ϕ_{12} is 0.2 and ϕ_{12} is 0.4. Therefore, the current study selects a more extreme level of ϕ_{12} and varies the values of ϕ_{12} to be 0.2 or 0.6.

The effect size refers to the increase of variance in η explained by the interaction term in addition to the first-order exogenous factors. The proportion of variance in η accounted for by the interaction effect of two normally distributed

random variables can be derived from the result of Bohrnstedt and Golderger (1967) and can be written as

$$R^2_{\gamma_3} = \frac{\text{var(int)}}{\text{var}(\eta)} = \frac{\gamma_3^2\ \text{var}(\xi_1\xi_2) + 2\,\text{cov}(\xi_1, \xi_1\xi_2) + 2\,\text{cov}(\xi_2, \xi_1\xi_2)}{\text{var}(linear) + \text{var(int)} + \text{var}(\zeta)},$$

where $\text{var}(\xi_1\xi_2) = (\phi_{11} + \mu_1^2)(\phi_{22} + \mu_2^2) + (\phi_{12} + \mu_1\mu_2)^2 - 2\mu_1^2\mu_2^2$,

$\text{cov}(\xi_1, \xi_1\xi_2) = \mu_1\phi_{12} + \phi_{11}\mu_2$,

$\text{cov}(\xi_2, \xi_1\xi_2) = \mu_2\phi_{12} + \phi_{22}\mu_1$,

and $\text{var}(linear) = \gamma_1^2\phi_{11} + \gamma_2^2\phi_{22} + 2\gamma_1\gamma_2\phi_{12}$.

It is clear that $R^2_{\gamma_3}$ is a function of μ_1, μ_2, γ_1, γ_2, γ_3 and ϕ_{12}. If $\mu_1 = \mu_2 = 0$, then $R^2_{\gamma_3} = \frac{\gamma_3^2(\phi_{11} + \phi_{22} + \phi_{12}^2)}{\text{var}(\eta)}$ is the same as that in Marsh et al. (2004). Given $R^2_{\gamma_3}$ and θ_ς^c, the corresponding γ_3 value can be calculated (see Appendix C). Jaccard and Wan (1995) reviewed social science literature and discovered that the effect sizes of the latent interaction effect generally ranged from 5% to 10%. The commonly used effect sizes in the latent interaction literature were 5% and 10% (e.g., Jaccard & Wan, 1995; Moulder & Algina, 2002; Weiss, 2010). However, large interaction effects of the empirical examples in nonlinear SEM articles can be found. For example, the effect size of the interaction effect in the empirical example in Cudeck, Harring, and du Toit (2009) was 27.73%. Three sets of effect sizes of latent interactions for class 1 and class 2 are selected in the current study: $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})$ = (0, 0.05), (0, 0.15), (0, 0.25). The parameter values of $\gamma_3^{(2)}$ for different conditions are given in Appendix B.

In simulation studies examining latent variable interactions, three commonly adopted sample sizes related to small, medium, and large sample sizes are 100, 250 (or 200), and 500. Several studies used larger sample sizes. For example, Wall and Amemiya (2003) used 200, 500, and 1000 to compare different methods' performance. Little, Bovaird, and Widaman (2006) chose a large sample size of 1500 to evaluate different methods. Therefore, the total samples are approximately twice the size of those used in Wall and Amemiya (2003) as the true model has two classes. The results of pilot work, which is included in 3.1.5, showed that the largest biases in γ_3 for both classes occurred when sample size is 200, MD is 1.5, and the other factors are fixed at the specific levels. Therefore, three different total sample sizes are manipulated in the current study: 500, 1000, and 2000.

Class mixing proportions have been frequently selected in the simulation studies of mixture models due to their influence on classification results. In growth mixture models, the more balanced class mixing proportions are, the more likely it is to identify the correct number of classes (Tofighi & Enders, 2008). The current study selects two combinations of the proportions for class 1 and class $2-$ one with balanced proportions 0.5/0.5 and one with more unbalanced proportions $0.2/0.8-$ to investigate how they impact the classification results.

3.2.3 Data generation. Data were generated with SAS 9.2 (SAS Institute, 2008).

Each of the conditions was replicated 500 times, which is adequate for SEM Monte Carlo studies to obtain stable standard error estimates even when data were simulated from a non-normal distribution (Bandalos, 2006).

3.3 Criteria Measures

To determine the number of latent classes in the data, the most commonly used are Akaike's Information Criterion (AIC; Akaike, 1987), Bayesian Information Criteria (BIC; Schwartz, 1978) and the adjust BIC (ABIC; Sclove, 1987). The AIC, BIC, and ABIC are defined as

$$AIC = -2\log L + 2p\,,$$

$$BIC = -2\log L + p\log(n)\,,$$

$$ABIC = -2\log L + p\log(n^*)\,,$$

where $\log L$ is the loglikelihood of the fitted model, p is the number of free model parameters, n is the sample size, and $n^* = (n+2)/24$. AIC has been reported to overestimate the number of latent classes in finite mixture models (Celeux & Soromenho, 1996; Soromenho, 1993), whereas BIC has been considered a consistent indicator for class enumeration in finite linear mixture structural equation model (Jedidi, Jagpal, & DeSargo, 1997). In other words, BIC is more likely to select the correct model as sample size gets large (Haughton, 1988; Tofighi & Enders, 2008). Furthermore, BIC is the best indicator among the ICs in latent class models, factor

mixture models, and growth mixture models (Nylund, Asparouhov, & Muthén, 2007). ABIC is to include the sample size adjustment which lowers the sample size penalty. Yang (2006) conducted a simulation study on latent class models and found the ABIC had the best performance among the information criteria statistics. Previous research chose these three information criteria to decide the number of latent classes in data (e.g., Lubke & Neale, 2006; Tofighi & Enders, 2008). In addition to the ICs, an alternative model-selection technique is based on log-likelihood ratio test (LRT) which is to compare the fit of two nested models (e.g., Bollen, 1989). In other words, it is to test if there is significant improvement in model fit between k-1 and k-class models. McLachlan and Peel (2000) proposed a bootstrap log-likelihood ratio test (BLRT) to approximate the empirical distribution of the log-likelihood difference test statistic using bootstrap samples. BLRT is very consistent fit criteria of class enumeration for latent class analysis, factor mixture models, and growth mixture models (Nylund, et al., 2007). However, skewed data which are modeled as normal may lead to incorrect BLRT p values. In addition, BLRT is time-consuming for each iteration and BIC has comparable performance with BLRT (Nylund, et al., 2007; Melnykov & Maitra, 2010). Therefore, the current study use AIC, BIC, and ABIC rather than BLRT to select models from one-, two-, and three-class latent interaction models and then calculate the proportions of correct model selection for all

conditions.

The performance of the proposed estimation method in parameter recovery is assessed with respect to bias (or relative bias) and precision computed by the parameter estimates $\hat{\theta}_i$ $(i = 1, \cdots, 500)$ obtained from 500 replications. Bias is computed as the difference between the average of the parameter estimates and the true value of the parameter being estimated. That is,

$$\text{Bias of } \hat{\theta} = \frac{\sum_1^{500} \hat{\theta}_i}{500} - \theta_0,$$

where θ_0 is the population value for θ. If population value is zero, then relative bias is used to measure bias. It is defined as follows.

$$\text{Relative bias of } \hat{\theta} = \frac{Bias\ of\ \hat{\theta}}{\theta_0}.$$

The precision of the standard error estimates will be evaluated by

$$\text{Precision of } \hat{\theta} = \frac{SE(\hat{\theta})}{SD_1(\hat{\theta})},$$

where $SE(\hat{\theta})$ is the square root of the mean variance of $\hat{\theta}$ derived from the 500 replications (i.e., $SE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{500} \left[SE_i(\hat{\theta})\right]^2}{500}}$), $SE_i(\hat{\theta})$ is the standard error estimates of θ for replication i, and $SD_1(\hat{\theta}) = SD(\hat{\theta}) \cdot \sqrt{\frac{499}{500}}$ is the corrected sample standard deviation of 500 parameter estimates in a given cell. If the estimated standard errors computed based on an approach are accurate, $SE(\hat{\theta})$ is close to $SD_1(\hat{\theta})$ and the ratio is close to 1 (Lee, Song, & Poon, 2004). $SE(\hat{\theta})$ is an unbiased estimate of the

true sampling variability, whereas the mean of the standard error estimates (i.e., $\frac{\sum_{i=1}^{500} SE_i(\hat{\theta})}{500} = \frac{\sum_{i=1}^{500} \sqrt{\text{var}_i(\hat{\theta})}}{500}$) is not. This is because taking a square root, a nonlinear transformation, does not result in an unbiased estimator of the population standard error although the variance of an estimator is unbiased. Values greater than 1 indicate that the standard errors are underestimated by the model whereas values less than 1 indicate that the standard errors are overestimated by the model.

The performance of the estimation method in classification results is evaluated by entropy and correct class assignment. Entropy is an indicator of classification quality, which is defined as (e.g., Ramaswamy, Desarbo, Reibstein, & Robinson, 1993)

$$E_K = 1 - \frac{\sum_i \sum_k (-\hat{p}_{ik} \ln \hat{p}_{ik})}{n \ln K},$$

where $\hat{p}_{ik}$ is the estimated conditional probability for individual i in class k, n is the sample size and K is the number of latent classes. Entropy measures the extent to which the latent classes are separated. Its values range from zero to one with higher values indicating better class separation (Muthén, 1998-2004). Entropy values indicate the degree the latent classes overlap. However, lower entropy values do not imply that the mixture patterns are not discernible.

Another criterion measure is related to correct class assignment. The model assigns the observations to their most likely class according to the greatest posterior

probability. Correct assignment probability is to calculate the proportion of subjects whose true class is consistent with the class he/she is assigned based on the posterior probability. In real data, entropy values can be derived whereas correct assignment probability is not possible to calculate.

Finally, convergence rates will also be included. A properly convergent replication in this study refers to the one that is not only convergent but also has a consistent maximum reached by multiple random start values. The divergent replications, including problems of multiple local maxima, will be ignored, and the simulation process will continue until 500 convergent replications are achieved to decrease the impact of divergent solutions.

3.4 Preliminary Analyses

The impact of the manipulated factors on parameter recovery is examined by the pilot work of 100 replications. Table 2 presents the bias in γ_3 and the SE/SD ratios by varying sample sizes and holding the other four factors constant. The sample size (SS) of 200 has the largest biases in γ_3 for both classes. There is a decreasing tendency when it is less than 1000. But the biases increase when SS is 1000 and level off when SS is larger than 1000. Tables 3 through 5 show that parameter bias in γ_3 decreases with class separation for both classes when SS is 500 and other three factors are fixed at the specified levels and parameter bias in γ_3 increases with the

Table 2. *Parameter Biases and SE/SD Ratios for Different Sample Sizes under MD=1.5,* $\phi_{12}=0.2$, $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})=(0, 0.15)$, *and* $(\pi_1, 1-\pi_1)=(0.5, 0.5)$.

	$\gamma_3^{(1)}=0$		$\gamma_3^{(2)}=0.4$	
n	bias	SE/SD	bias	SE/SD
200	0.282	1.212	0.150	1.039
500	0.082	1.116	0.075	1.036
1000	0.046	1.09	0.074	1.28
2000	0.118	1.094	0.095	1.116
5000	0.105	0.996	0.088	1.099
10000	0.09	1.099	0.069	1.01

Table 3. *Parameter Biases and SE/SD Ratios for Different Class Separations under* $n=500$, $\phi_{12}=0.2$, $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})=(0, 0.05)$, *and* $(\pi_1, 1-\pi_1)=(0.5, 0.5)$.

MD	$\gamma_3^{(2)}$	Bias	SE/SD	$\gamma_3^{(1)}$	Bias	SE/SD
0	1.13	0.617	0.869	0	0.562	1.195
1.5	0.13	0.049	1.062	0	0.051	1.125
3	0.09	0.029	1.167	0	0.01	1.159

Table 4. *Parameter Biases and SE/SD Ratios for Different Class Separations under* $n=500$, $\phi_{12}=0.2$, $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})=(0, 0.15)$, *and* $(\pi_1, 1-\pi_1)=(0.5, 0.5)$.

MD	$\gamma_3^{(2)}$	Bias	SE/SD	$\gamma_3^{(1)}$	Bias	SE/SD
0	2.08	1.088	0.925	0	1.031	1.405
1.5	0.4	0.074	1.036	0	0.081	1.066
3	0.29	0.037	1.149	0	0.03	1.163

Table 5. *Parameter Biases and SE/SD Ratios for Different Class Separations under* $n=500$, $\phi_{12}=0.2$, $(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})=(0, 0.25)$, *and* $(\pi_1, 1-\pi_1)=(0.5, 0.5)$.

MD	$\gamma_3^{(2)}$	Bias	SE/SD	$\gamma_3^{(1)}$	Bias	SE/SD
0	2.85	1.421	0.981	0	1.411	1.033
1.5	0.7	0.107	0.967	0	0.125	1.084
3	0.51	0.046	1.128	0	0.051	1.161

interaction effects in class 2 when SS is 500 and the other three factors are fixed at the specified levels.

In order to avoid local maxima in mixture models, it is recommended to increase random start values. The results of the pilot work showed that the convergence rate of some conditions for 100 replications are at least 80% with STARTS=25 3. In addition, the unperturbed solution can be obtained by using STARTS=25 25 in each replication to examine whether it is identical to the maximum converged by other start values. It is usually the case that only part of the first stage start values are used for the second stage start values since it is extremely time consuming to run STARTS=25 25 (e.g., about 31 hours for 100 replications of a cell when MD=0) compared with STARTS=25 3 (about 2.5 hours for 100 replications of the same cell) and there is no difference or slight difference in the convergence rate between STARTS=25 25 and STARTS=25 3 for 100 replications in different class separation in factor means. For example, two cells with medium (MD=1.5) and large (MD=3) class separation in factor means have identical convergence rates using STARTS=25 25 and STARTS=25 3. There is 6% difference in the convergence rate between STARTS=25 25 (86%) and STARTS=25 3 (80%) for a cell with MD=0. The current study selects STARTS=25 3 because it is more manageable and the convergence rates are slightly inconsistent when MD=0.

The feasibility of the fourth research question about model selection is assessed by conducting pilot work on a cell with MD 3 in order to examine whether the fitted one-, two-, and three-class latent interaction models underextract, correctly extract, or overextract the number of latent classes from the data generated by a true two-class model. The model estimation went well except for the three-class models, which were very computationally intensive and had very low convergence rate. For example, ten replications took four hours with only one converged replication. This suggests that the estimation of the three-class models is unmanageable. However, the last research question can be investigated instead by fitting one- and two-class latent interaction models. Using these two fitted models to assess how likely the underextraction, correct extraction, or overextraction of the latent classes would occur if data come from the populations with one-class and two-class latent interaction effects, respectively. This research question can be illustrated by Table 6.

Table 6. *Illustration of the Fourth Research Question*.

		True model	
		1-class	2-class
Selected model	1-class	correctly extract	underextract
	2-class	overextract	correctly extract

For data generated from a two-class model, the population parameter vlaues and manipulated conditions are the same as described in Section 3.1.1 and 3.1.2. For data

generated from a one-class latent interaction model, the population model is as follows.

$$\eta_i = \gamma_0 + \gamma_1\xi_{1i} + \gamma_2\xi_{2i} + \gamma_3\xi_{1i}\xi_{2i} + \zeta_i, \quad i = 1,\cdots,n.,$$

where $\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \sim N(\begin{pmatrix}0.6\\ 0.8\end{pmatrix},\begin{pmatrix}1 & \\ \phi_{12} & 1\end{pmatrix}$, the other parameter values are the same as those in Section 3.1.1 and γ_3 values can be calculated according to the manipulated effect sizes. The manipulated conditions are reduced 18 conditions (2 (covariance between linear factors ϕ_{12} =0.2, 0.6)×3 (effect size of the latent interaction $R^2_{\gamma_3}$ = 0.05, 0.15, 0.25)×3 (total sample sizes= 500, 1000, 2000)).

3.5 Data Analysis

Data were analyzed by Mplus 6.2 (Muthén & Muthén, 2010) using the mixture module to obtain parameter estimates, classify observations to their most likely latent class, and identify the optimal number of latent classes in the data. The impact of the manipulated factors can be evaluated by factorial ANOVA. The current study will assess the significance of the manipulated factors and their interactions on parameter recovery and classification results by eta-squared (denoted by η^2) using the results from factorial ANOVA. η^2, which is equivalent to R^2 in regression, is the proportion of variance in the dependent variable explained by each of the main effects and interactions in ANOVA. In spite of its upward bias, η^2 has the advantages of additivity to 1 and straightforward interpretation.

Full factorial ANOVA is used to examine the influence of the manipulated factors and their combinations on criteria measures. Due to limited space in this chapter, only the effects of the factors and their interactions are reported if they are both statistically significant (p-value ≤ 0.05) and have effect sizes greater or equal than 0.05 ($\eta^2 \geq 0.05$). $\eta^2 \geq 0.05$ is selected to identify the effects (especially interactions) since their η^2 values are mostly less than 0.1. In the current study, the effect size is classified to be small if $0.05 \leq \eta^2 < 0.1$, to be medium if $0.1 \leq \eta^2 < 0.2$, and large if $\eta^2 \geq 0.2$.

In summary, the research questions in the current study are as follow.

1. Can a mixture of latent interaction models be estimated using an unconstrained product indicator approach?
2. How well does the estimation method for two-class latent variable interaction models perform in recovering the parameters of interest under various conditions?
3. How well does this estimation method classify the observations from a two-class latent variable interaction model?
4. What is the probability to underextract, correctly extract, or overextract the number of latent classes in data, generated from populations with one- or two-class latent interaction effects, when fitted with one- and two-class

latent interaction models using model fit statistics?

The current study has two simulation studies. The first simulation study examines the estimation performance of the two-class latent interaction models which answers the first three research questions whereas the second one examines model selection for one- and two-class latent interaction models which answers the fourth research question. In particular, the first simulation study evaluates parameter recovery and classification quality of the estimated two-class latent interaction model. The manipulated factors are: the class separation of the factor means (MD=0, 1.5, 3), the covariance between linear factors (ϕ_{12}=0.2, 0.6), the effect size of latent interaction across classes ($(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})$ = ((0, 0.05), (0, 0.15), (0, 0.25), the total sample size (500, 1000, 2000), and the class proportion ($(\pi_1, 1-\pi_1)$ =(0.5, 0.5), (0.2, 0.8)), which constitutes a $3\times2\times3\times3\times2$ (=108) factorial design. Parameter recovery and class assignment will be evaluated by relative bias (or bias), SE/SD ratio, entropy, and proportion of assigning subjects to their true class. The impact of the manipulated factors will be evaluated by factorial ANOVA. Convergence rates are also discussed. The second simulation study examines class enumeration of the data generated from populations with one- and two-class interaction effects using one- and two-class latent interaction models. The manipulated factors are covariance between linear factors (ϕ_{12}=0.2, 0.6), the effect size of latent interaction across classes ($(R^2_{\gamma_3^{(1)}}, R^2_{\gamma_3^{(2)}})$ = ((0,

0.05), (0, 0.15), (0, 0.25), and the total sample size (500, 1000, 2000)), which constitutes a $2\times3\times3$ (=18) factorial design. Model selection is based on three information criteria- AIC, BIC, and ABIC.

Chapter 4: Results

This chapter presents the results from the simulation work. Convergence and results of model selection are discussed in Section 4.1 and Section 4.2, respectively. Section 4.3, provides the results regarding parameter recovery using the unconstrained method for latent interaction models within the EM algorithmic framework. In Section 4.4, the results regarding the estimation method's classification quality are presented. Although the parameter recovery of $\gamma_3^{(1)}$ and $\gamma_3^{(2)}$ are of primary interest, parameter recovery is also investigated for other structural parameters. The results of each condition represent the average values of the outcome measures over 500 properly converged replications.

4.1 Convergence

A properly converged replication refers to the one that is not only converged but also has a consistent maximum reached by multiple random start values. The convergence rates are calculated using the previous 500 replications for each cell. From Table 7, convergence rates with satisfactory class separation between the exogenous means (i.e., MD=1.5 and 3) have noticeable improvements compared to those with zero class separation. In particular, the convergence rates for the conditions with larger class separation range from 0.982 to 1 whereas those for zero class separation range from 0.71 to 0.854. It is found that most of the non-converged cases have the problem of multiple local maxima when examining the results. Only one cell with MD=0 and one cell with MD=1.5 have one replication with the problem of not being able converge. Although it is not the focus of the current study, further examination can be conducted to compare parameter estimates of the non-converged cases with those from the converged cases.

An arcsine transformation is applied to the convergence rates before conducting a full factorial ANOVA since they are measured in proportions. Class separation is the only significant factor which explains 96.1% (large effect) of total variation in the transformed convergence rate.

Table 7 . *Convergence rates.*

π_1	MD	$R^2_{\chi_3^{(2)}}$	n=500		n=1000		n=2000	
			$\phi_{12}=0.2$	$\phi_{12}=0.6$	$\phi_{12}=0.2$	$\phi_{12}=0.6$	$\phi_{12}=0.2$	$\phi_{12}=0.6$
0.5	0	0.05	0.826	0.816	0.852	0.764	0.854	0.830
		0.15	0.810	0.776	0.806	0.710	0.836	0.752
		0.25	0.814	0.758	0.778	0.710	0.756	0.712
	1.5	0.05	0.982	0.990	1.000	0.998	0.998	1.000
		0.15	0.990	0.984	0.998	1.000	1.000	0.998
		0.25	0.990	0.988	0.998	0.996	1.000	1.000
	3	0.05	0.990	1.000	1.000	0.998	1.000	1.000
		0.15	0.996	1.000	0.998	1.000	1.000	1.000
		0.25	0.996	0.998	1.000	1.000	1.000	1.000
0.2	0	0.05	0.822	0.758	0.834	0.756	0.852	0.794
		0.15	0.820	0.754	0.814	0.722	0.842	0.730
		0.25	0.820	0.756	0.816	0.708	0.808	0.774
	1.5	0.05	0.994	0.988	1.000	0.998	1.000	1.000
		0.15	0.992	0.992	0.998	0.998	1.000	1.000
		0.25	0.990	1.000	1.000	1.000	1.000	1.000
	3	0.05	1.000	1.000	0.998	1.000	1.000	1.000
		0.15	0.998	1.000	1.000	1.000	1.000	1.000
		0.25	0.998	1.000	1.000	1.000	1.000	1.000

4.2 Identification of Latent Classes

One research question in the current study is to examine whether the fitted one- and two-class latent interaction models correctly identify the number of latent classes in the data generated by populations with one- and two-class interactions based on AIC, BIC, and ABIC. The results (see Table 8) show that when data were generated by a true two-class interaction model, the fitted models selected by model fit statistics correctly extract the number of latent classes 100% of the time. However, when data were generated by a true one-class interaction model, the fitted models selected by the fit statistics overestimate the number of latent classes in the data 100% of the time. In other words, every model fit statistic used in this study indicates that two-class models provide better fit than one-class models regardless of whether the data generated from populations with one- or two-class interactions. This is an important discovery, which will be further discussed in Chapter 5.

Table 8. *Identification of Latent Classes Using One- and Two-class Interaction Models.*

		True model	
		1-class	2-class
Selected model	1-class	correctly extract	underextract
	2-class	overextract (100%)	correctly extract (100%)

4.3 Parameter Recovery

Results of factorial ANOVA on relative bias (or bias) and standard error precision are discussed in Section 4.3.1.1 and Section 4.3.1.2, respectively, to investigate the effects of the manipulated factors and their combinations on the outcome measures. Precision is evaluated by SE/SD ratios. Section 4.3.2 and Section 4.3.3 present the results of parameter recovery including relative bias (or bias) for the structural parameter estimates and the class proportion estimates and SE/SD estimates

for structural parameters, respectively. Graphical representations are provided if clear patterns are observed. Only graphical representations for small sample size (SS=500) are presented if the patterns for outcome measures are similar across different sample sizes.

Bias is the difference between the average of the estimates derived from all replications and the true parameter value whereas relative bias is the ratio of bias to the true parameter value. Relative bias provides a common scale for researchers to compare the deviations of the estimates from different nonzero population parameter values. If the absolute values of relative bias are less than 0.1, they are considered to be acceptable. In the conditions where relative bias is either impossible to calculate (e.g., population parameter value, the denominator of the relative bias, is zero) or difficult to interpret (e.g., population parameter is measured in proportion), bias rather than relative bias is reported.

4.3.1 Factorial ANOVA on results of parameter recovery.

4.3.1.1Factorial ANOVA on relative bias (or bias). Table 9 summarizes the significant factors of their combinations for relative bias (or bias) of all the parameter estimates using the criteria stated previously. The relative bias (or bias) values of the structural parameter estimates and the class proportion estimates according to different sample sizes are reported in Tables 10 through 12.

It is expected that class separation (CS) is an important factor which significantly affects the relative bias (or bias) for all the estimated parameters. Class proportion (CP) and effect size (ES) significantly influence the relative bias (or bias) of three and four estimated parameters, respectively. Total sample size (SS) does not significantly affect the relative bias (or bias) of any parameter. The interaction between CS and CP

as well as CS and ES is significant for the relative bias (or bias) of some parameter estimates.

Table 9. *Factors and Their Combinations Which Explain More Than 5% of the Variance in the Relative Bias (or Bias) of the Estimated Parameters.*

	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)*}$	$\gamma_3^{(2)}$	π_1^*
SS							
COV	7.4%						
CP	24.2%		27.3%		8.5%		86.8%
CS	14.6%	89.5%	35.4%	14.5%	65.8%	8.4%	11.4%
ES	9.8%				8.9%	37.1%	
SS*CS	9.0%						
COV*CP	6.4%						
CP*CS			15.0%	5.5%		30.4%	
CS*ES					9.0%	16.8%	
SS*CS*ES	5.2%						
SS*COV*CS*ES				5.3%			
SS*COV*CP*CS*ES				5.4%			

* reported in the form of bias.

Table 10. *Relative Bias (or Bias) of Parameter Estimates When SS=500.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$							$\phi_{12}=0.6$						
			$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *	$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *
0.5	0	0.05	0.004	-0.002	0.043	0.019	0.569	-0.534	0.041	0.004	-0.002	0.028	0.018	0.599	-0.518	0.040
		0.15	0.001	0.003	0.007	0.007	1.118	-0.509	0.041	-0.001	0.004	0.009	0.006	1.133	-0.503	0.060
		0.25	0.003	0.004	0.018	0.026	1.471	-0.498	0.050	0.001	0.002	0.017	0.031	1.458	-0.483	0.066
	1.5	0.05	0.002	-0.029	0.052	0.033	0.078	-0.409	-0.015	0.003	-0.033	0.064	0.035	0.104	-0.456	-0.029
		0.15	0.001	-0.032	0.048	0.036	0.114	-0.202	-0.016	0.003	-0.032	0.046	0.026	0.105	-0.190	-0.031
		0.25	0.000	-0.034	-0.037	0.033	0.135	-0.148	-0.015	0.000	-0.035	0.030	0.019	0.121	-0.144	-0.028
	3	0.05	0.002	-0.028	0.045	0.029	0.049	-0.400	-0.034	0.002	-0.030	0.053	0.029	0.067	-0.394	-0.042
		0.15	0.001	-0.029	0.040	0.027	0.031	-0.122	-0.036	0.002	-0.031	0.043	0.019	0.033	-0.115	-0.044
		0.25	-0.003	-0.034	0.039	0.028	0.068	-0.104	-0.034	-0.003	-0.036	0.035	0.016	0.057	-0.085	-0.042
0.2	0	0.05	0.007	-0.003	0.037	0.013	0.744	-0.250	0.343	0.002	0.000	0.034	0.015	1.102	-0.229	0.345
		0.15	0.005	0.002	0.016	0.013	1.738	-0.218	0.342	0.002	0.004	0.012	0.009	1.802	-0.212	0.353
		0.25	0.003	0.009	0.020	0.022	2.604	-0.215	0.346	0.002	0.006	0.015	0.023	2.578	-0.213	0.356
	1.5	0.05	0.007	-0.044	0.090	0.057	0.212	-0.559	0.229	0.004	-0.039	0.080	0.044	0.165	-0.406	0.232
		0.15	0.005	-0.042	0.088	0.058	0.331	-0.222	0.227	0.004	-0.040	0.081	0.047	0.300	-0.171	0.230
		0.25	0.005	-0.049	0.096	0.071	0.506	-0.163	0.231	0.003	-0.045	0.080	0.055	0.451	-0.124	0.232
	3	0.05	0.008	-0.049	0.101	0.063	0.345	-0.856	0.177	0.003	-0.047	0.094	0.053	0.233	-0.626	0.192
		0.15	0.007	-0.045	0.094	0.059	0.330	-0.280	0.175	0.003	-0.048	0.098	0.056	0.302	-0.240	0.193
		0.25	0.004	-0.057	0.119	0.042	0.554	-0.224	0.175	0.000	-0.058	0.113	0.067	0.432	-0.171	0.191

* reported in the form of bias.

Table 11. *Relative Bias (or Bias) of Parameter Estimates When SS=1000.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$							$\phi_{12}=0.6$						
			$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *	$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *
0.5	0	0.05	0.005	-0.004	0.038	0.019	0.585	-0.547	0.021	0.004	-0.001	0.051	0.023	0.568	-0.530	0.021
		0.15	0.005	0.001	0.027	0.013	1.060	-0.513	0.022	0.004	0.002	0.051	0.023	1.047	-0.520	0.029
		0.25	0.001	0.004	0.021	0.019	1.926	-0.511	0.029	0.003	0.004	0.026	0.017	1.413	-0.516	0.027
	1.5	0.05	0.001	-0.028	0.050	0.029	0.062	-0.339	-0.019	0.002	-0.030	0.055	0.026	0.077	-0.362	-0.031
		0.15	0.001	-0.031	0.046	0.029	0.098	-0.174	-0.022	0.001	-0.031	0.045	0.021	0.096	-0.174	-0.031
		0.25	0.000	-0.037	0.052	0.035	0.159	-0.159	-0.021	0.000	-0.035	0.034	0.016	0.120	-0.136	-0.032
	3	0.05	0.001	-0.027	0.046	0.027	0.029	-0.389	-0.040	0.002	-0.028	0.053	0.026	0.046	-0.384	-0.048
		0.15	0.001	-0.034	0.049	0.035	0.070	-0.161	-0.039	0.001	-0.036	0.053	0.028	0.073	-0.157	-0.047
		0.25	0.001	-0.039	0.056	0.033	0.082	-0.112	-0.037	0.000	-0.038	0.046	0.016	0.053	-0.085	-0.045
0.2	0	0.05	0.006	-0.002	0.038	0.019	0.933	-0.256	0.321	0.006	0.000	0.056	0.026	0.865	-0.238	0.324
		0.15	0.005	0.003	0.029	0.015	1.859	-0.228	0.323	0.006	0.002	0.052	0.022	1.696	-0.215	0.327
		0.25	0.004	0.007	0.015	0.012	2.500	-0.223	0.328	0.004	0.007	0.027	0.014	2.343	-0.219	0.340
	1.5	0.05	0.005	-0.035	0.074	0.044	0.158	-0.472	0.226	0.002	-0.032	0.068	0.034	0.134	-0.355	0.230
		0.15	0.004	-0.037	0.073	0.047	0.286	-0.169	0.224	0.002	-0.037	0.070	0.039	0.270	-0.141	0.229
		0.25	0.003	-0.050	0.103	0.068	0.520	-0.149	0.224	0.002	-0.047	0.086	0.051	0.450	-0.108	0.226
	3	0.05	0.006	-0.039	0.085	0.051	0.233	-0.778	0.172	0.002	-0.039	0.082	0.046	0.171	-0.606	0.188
		0.15	0.005	-0.048	0.102	0.068	0.369	-0.332	0.172	0.001	-0.049	0.100	0.060	0.296	-0.266	0.188
		0.25	0.004	-0.055	0.119	0.074	0.480	-0.209	0.172	0.000	-0.054	0.110	0.061	0.370	-0.157	0.189

* reported in the form of bias.

Table12. *Relative Bias (or Bias) of Parameter Estimates When SS=2000.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$							$\phi_{12}=0.6$						
			$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *	$\hat{\gamma}_0^{(1)}$	$\hat{\gamma}_0^{(2)}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3^{(1)}$ *	$\hat{\gamma}_3^{(2)}$	$\hat{\pi}_1$ *
0.5	0	0.05	0.004	-0.003	0.033	0.019	0.608	-0.548	0.008	0.005	-0.001	0.053	0.028	0.552	-0.535	0.008
		0.15	0.006	-0.001	0.029	0.021	1.046	-0.525	0.007	0.004	0.001	0.047	0.029	1.000	-0.510	0.019
		0.25	0.007	-0.001	0.040	0.022	1.507	-0.526	0.004	0.004	0.001	0.059	0.028	1.512	-0.511	0.041
	1.5	0.05	0.006	-0.030	0.056	0.038	0.079	-0.425	-0.022	0.002	-0.030	0.059	0.034	0.087	-0.420	-0.034
		0.15	0.001	-0.032	0.050	0.319	0.102	-0.186	-0.021	0.000	-0.029	0.043	0.020	0.094	-0.176	-0.033
		0.25	0.000	-0.035	0.044	0.032	0.139	-0.152	-0.022	0.000	-0.032	0.028	0.014	0.113	-0.136	-0.033
	3	0.05	0.003	-0.031	0.054	0.034	0.053	-0.422	-0.040	0.003	-0.032	0.060	0.032	0.067	-0.424	-0.048
		0.15	0.002	-0.034	0.053	0.034	0.071	-0.168	-0.040	0.002	-0.035	0.056	0.026	0.073	-0.154	-0.047
		0.25	0.000	-0.036	0.047	0.028	0.074	-0.096	-0.038	0.000	-0.038	0.046	0.016	0.062	-0.083	-0.046
0.2	0	0.05	0.004	-0.001	0.033	0.019	0.949	-0.246	0.309	0.003	0.001	0.055	0.030	0.868	-0.229	0.312
		0.15	0.006	0.002	0.026	0.021	1.672	-0.229	0.309	0.006	0.002	0.041	0.025	1.631	-0.222	0.318
		0.25	0.008	0.002	0.038	0.018	2.408	-0.229	0.315	0.006	0.003	0.053	0.025	2.439	-0.226	0.337
	1.5	0.05	0.004	-0.038	0.077	0.051	0.166	-0.504	0.224	0.002	-0.036	0.072	0.043	0.144	-0.399	0.228
		0.15	0.004	-0.042	0.087	0.055	0.303	-0.204	0.222	0.002	-0.040	0.079	0.044	0.277	-0.162	0.227
		0.25	0.003	-0.045	0.091	0.060	0.457	-0.142	0.223	0.002	-0.041	0.076	0.044	0.413	-0.104	0.227
	3	0.05	0.008	-0.041	0.087	0.055	0.239	-0.789	0.171	0.004	-0.042	0.088	0.051	0.187	-0.646	0.188
		0.15	0.007	-0.048	0.101	0.064	0.346	-0.297	0.172	0.002	-0.048	0.099	0.056	0.278	-0.237	0.189
		0.25	0.005	-0.054	0.111	0.069	0.444	-0.193	0.171	0.015	-0.055	0.107	0.060	0.358	-0.148	0.188

* reported in the form of bias.

.

4.3.1.2 Factorial ANOVA on precision estimates. Table 13 summarizes the factors and their combinations that are significant for the SE/SD estimates. The estimated SE/SD ratios for different samples sizes are presented in Tables 14 through 16. Class separation is the only main effect which significant affects the precision of three estimated parameters. The effects that account for 5-10% of the total variation in precision are the interactions among the factors.

Table 13. *Factors and Their Combinations Which Explain More Than 5% of the Variance in the Precision of Estimated Parameters.*

Factors	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$
SS						
COV						
CP						
CS	11.9%				21.3%	12.1%
ES						
SS*CS			6.0%			
SS*ES					5.0%	6.5%
COV*CS			5.9%	6.5%		
CP*CS	9.9%				8.5%	
CP*ES						5.4%
SS*COV*CS			7.7%			
SS*CP*CS		5.3%		8.8%		
SS*CS*ES		5.8%			10.0%	8.5%
COV*CP*CS			5.2%			
SS*CP*CS*ES	8.8%	5.4%	5.3%		7.5%	
SS*COV*CS*ES		7.9%				
SS*COV*CP*CS			5.3%	8.8%		
SS*COV*CP*CS*ES		7.7%		7.1%		

Table 14. *Estimated SE/SD Ratios When SS=500.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$						$\phi_{12}=0.6$					
			$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$
0.5	0	0.05	1.052	0.931	0.992	1.015	1.343	0.931	1.156	1.073	1.022	1.060	2.125	0.993
		0.15	1.072	1.106	1.017	1.006	1.150	0.897	1.447	1.110	1.051	0.962	3.710	0.909
		0.25	1.008	1.469	1.032	1.570	1.96	0.980	1.022	0.998	0.923	0.879	1.670	0.835
	1.5	0.05	1.060	1.098	1.048	1.060	1.064	0.994	1.105	1.131	1.093	1.103	1.070	1.054
		0.15	1.057	1.108	1.069	1.041	1.011	1.010	1.111	1.103	1.093	1.030	0.996	1.022
		0.25	1.039	1.035	0.998	0.985	0.971	0.977	1.072	1.054	1.024	1.030	0.986	1.007
	3	0.05	1.223	1.239	1.177	1.169	1.128	1.117	1.056	1.091	1.051	1.052	1.043	1.019
		0.15	1.031	1.043	1.029	1.028	0.998	1.041	1.064	1.092	1.055	1.059	1.005	1.031
		0.25	1.136	1.140	1.098	1.095	1.081	1.033	1.111	1.161	1.102	1.100	1.083	1.034
0.2	0	0.05	3.989	1.059	0.996	0.998	9.057	0.937	2.910	1.078	1.031	1.044	10.663	0.934
		0.15	1.103	1.122	1.018	0.996	1.070	0.899	1.129	1.111	1.017	0.943	2.054	0.880
		0.25	1.287	1.167	1.010	0.946	6.077	0.945	5.005	1.018	1.075	1.073	6.190	1.068
	1.5	0.05	1.064	1.131	1.060	1.052	1.100	1.039	1.129	1.164	1.101	1.115	1.101	1.068
		0.15	1.024	1.128	1.114	1.071	1.072	1.079	1.083	1.173	1.118	1.108	1.086	1.073
T		0.25	1.077	1.092	1.082	1.076	1.073	1.075	1.107	1.130	1.054	1.088	0.991	1.036
	3	0.05	0.978	1.109	1.069	1.052	1.170	1.036	1.003	1.124	1.084	1.088	1.154	1.061
		0.15	0.985	1.088	1.100	1.058	1.125	1.097	1.084	1.113	1.120	1.061	1.092	1.080
		0.25	1.236	1.127	1.142	1.175	1.222	1.140	1.036	1.076	1.055	1.046	0.996	1.016

Table 15. *Estimated SE/SD Ratios When SS=1000.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$						$\phi_{12}=0.6$					
			$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$
0.5	0	0.05	1.017	1.044	1.078	1.036	1.309	0.947	1.044	0.985	1.024	1.011	2.663	0.942
		0.15	1.066	1.067	0.996	1.000	1.641	0.931	1.008	0.985	0.942	0.924	1.333	0.890
		0.25	0.822	1.075	1.023	1.078	0.532	0.953	1.129	1.269	1.133	1.280	1.923	0.995
	1.5	0.05	1.072	1.014	1.035	1.059	1.002	1.051	1.071	1.028	1.011	1.041	0.999	1.022
		0.15	1.071	1.108	1.028	1.056	1.013	0.972	1.092	1.092	1.018	1.064	1.015	0.987
		0.25	1.084	1.213	1.116	1.204	1.373	1.102	1.039	1.075	1.023	0.999	0.999	1.009
	3	0.05	0.952	1.025	1.005	1.015	0.995	1.014	0.973	1.047	1.008	1.009	0.995	0.995
		0.15	1.107	1.066	1.068	1.038	0.995	1.032	1.051	1.011	1.028	0.980	0.989	0.992
		0.25	1.028	1.046	1.038	1.022	1.027	1.052	1.019	1.081	1.063	1.055	1.047	1.074
0.2	0	0.05	1.033	1.035	1.090	1.039	1.615	0.957	1.030	0.991	1.036	0.995	1.648	0.901
		0.15	3.756	1.193	1.441	1.533	8.790	1.420	1.102	1.032	1.015	1.020	1.562	0.949
		0.25	2.546	1.073	1.042	1.069	6.558	0.995	0.960	1.003	0.901	0.954	1.318	0.935
	1.5	0.05	1.084	1.117	1.037	1.057	1.024	1.025	1.060	1.067	1.023	1.024	1.027	1.016
		0.15	1.052	1.061	0.998	1.015	0.999	0.955	1.064	1.060	1.013	1.028	1.002	0.964
		0.25	1.071	1.061	1.053	1.031	1.043	1.101	1.067	1.031	1.013	1.001	0.967	0.994
	3	0.05	0.972	1.118	1.091	1.067	1.047	1.051	0.988	1.079	1.057	1.022	1.049	1.013
		0.15	1.017	1.050	1.053	1.035	1.035	1.018	0.992	1.052	1.064	1.025	1.024	1.042
		0.25	1.044	1.087	1.103	1.053	1.109	1.060	1.059	1.216	1.158	1.103	1.151	1.082

Table 16. *Estimated SE/SD Ratios When SS=2000.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12}=0.2$						$\phi_{12}=0.6$					
			$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$	$\gamma_0^{(1)}$	$\gamma_0^{(2)}$	γ_1	γ_2	$\gamma_3^{(1)}$	$\gamma_3^{(2)}$
0.5	0	0.05	1.029	1.055	1.001	1.021	1.188	0.929	1.030	0.992	0.987	1.005	0.851	0.867
		0.15	1.059	0.998	1.013	1.006	1.143	0.946	0.987	0.970	0.962	0.977	1.375	0.911
		0.25	1.000	0.999	0.991	0.973	0.996	0.921	1.670	1.006	0.972	0.977	5.601	0.905
	1.5	0.05	1.045	0.989	0.959	1.009	0.996	0.982	1.061	1.000	0.980	1.027	1.019	1.009
		0.15	1.025	0.970	1.055	1.030	1.022	1.059	1.031	0.997	1.055	1.015	1.010	1.023
		0.25	0.997	1.046	0.995	1.028	0.963	1.004	1.001	1.070	1.138	1.083	1.169	1.036
	3	0.05	1.024	1.054	1.035	1.058	1.025	1.044	1.054	1.055	1.005	1.056	1.022	1.037
		0.15	1.017	1.084	1.091	1.106	1.101	1.067	1.011	1.024	1.029	1.007	0.989	0.991
		0.25	0.989	1.052	1.061	1.135	1.116	1.231	0.994	1.045	1.035	1.058	0.991	1.036
0.2	0	0.05	1.005	1.079	1.003	1.023	1.267	0.918	1.047	0.984	0.986	1.017	0.996	0.867
		0.15	2.044	2.933	2.003	2.183	1.336	1.481	1.007	0.986	0.960	1.017	1.609	0.943
		0.25	1.414	1.081	2.571	2.199	14.602	0.956	1.102	1.013	0.960	0.915	2.784	0.860
	1.5	0.05	1.068	0.960	0.946	0.993	1.003	0.973	1.069	0.978	0.982	1.006	1.000	1.004
		0.15	1.017	1.055	1.152	1.118	1.128	1.174	1.034	1.088	1.179	1.056	1.075	1.799
		0.25	1.026	1.023	1.004	1.029	0.984	1.038	1.038	1.039	1.231	1.083	1.157	1.075
	3	0.05	0.979	1.035	1.031	1.030	1.014	1.022	0.982	1.000	0.994	1.020	1.034	1.007
		0.15	0.958	1.041	1.016	1.018	1.024	0.986	0.992	1.031	0.994	1.012	0.992	0.974
		0.25	0.966	1.039	1.015	1.026	1.019	0.996	0.955	1.030	0.998	1.037	1.015	0.986

4.3.2 Relative bias (or bias).

This section illustrate the main effects and interaction effects for relative bias (or bias) for $\hat{\gamma}_0^{(1)}$, $\hat{\gamma}_3^{(1)}$, $\hat{\gamma}_3^{(2)}$, and $\hat{\pi}_1$ with graphs. It will also examine if the relative bias (or bias) for all estimated parameters are accetable using quantiles.

Figures 6a-6f compare the relative bias values of $\hat{\gamma}_0^{(1)}$ for different combinations of class proportions and class separation when sample size, covariance between the linear factors, and effect size of the latent interaction term are fixed at the specified levels. The graphs show that the conditions with unequal class proportion (CP=0.2) have greater relative bias values of $\hat{\gamma}_0^{(1)}$ in magnitude than those with equal class proportion (CP=0.5). These figures help to explain why class proportion is a significant factor accounting for the largest proportion of variation (24.3%, large effect) in relative bias of $\hat{\gamma}_0^{(1)}$ in Table 9.

Figure 6a. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.05, and COV=0.2.

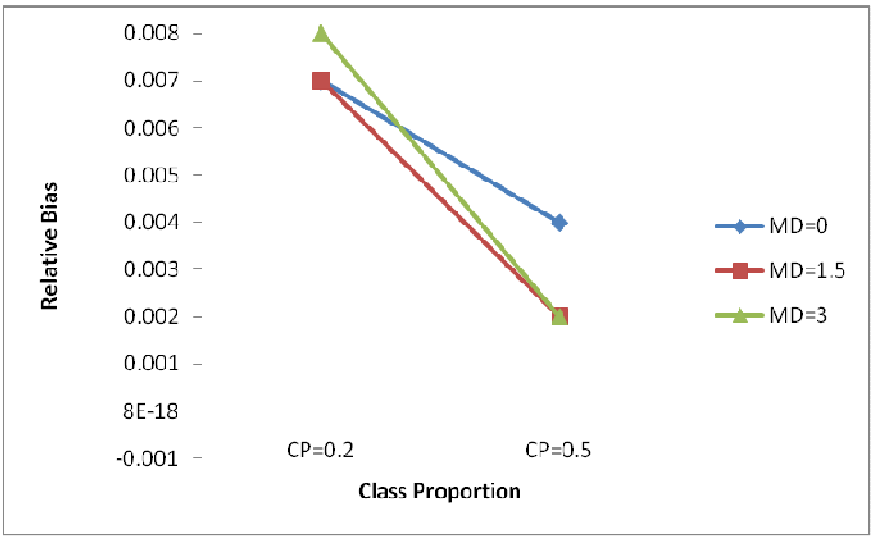


Figure 6b. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.15, and COV=0.2.

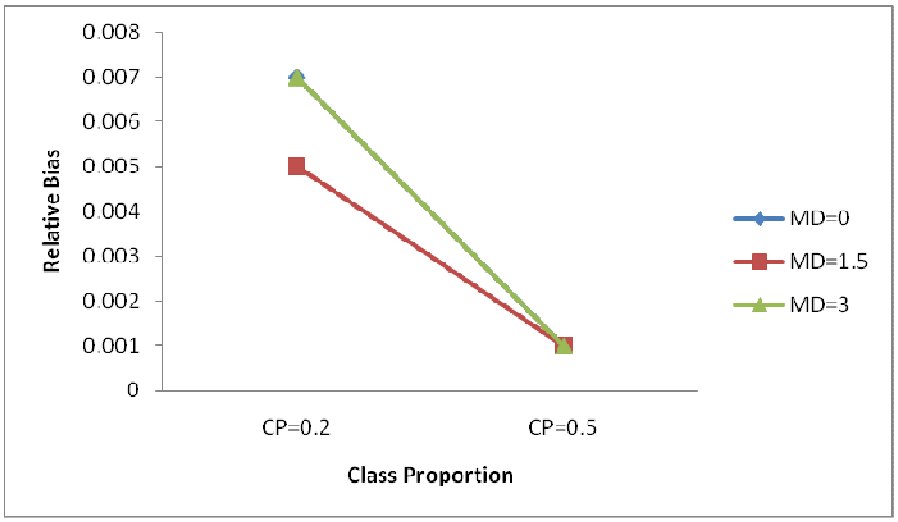


Figure 6c. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.25, and COV=0.2.

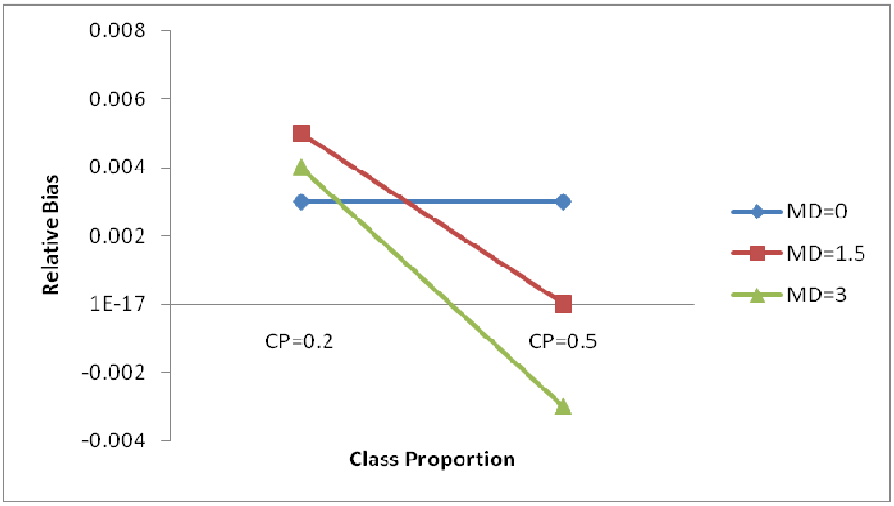


Figure 6d. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.05, and COV=0.6.

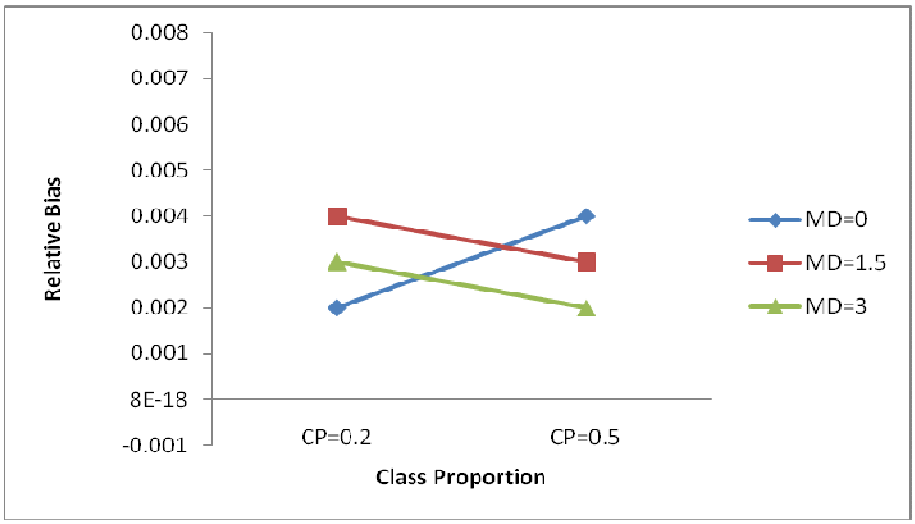


Figure 6e. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.15, and COV=0.6.

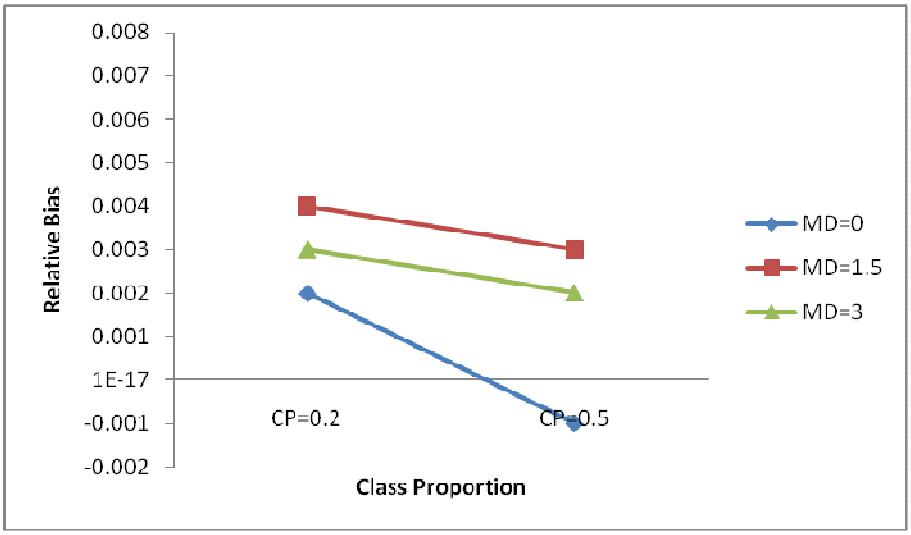
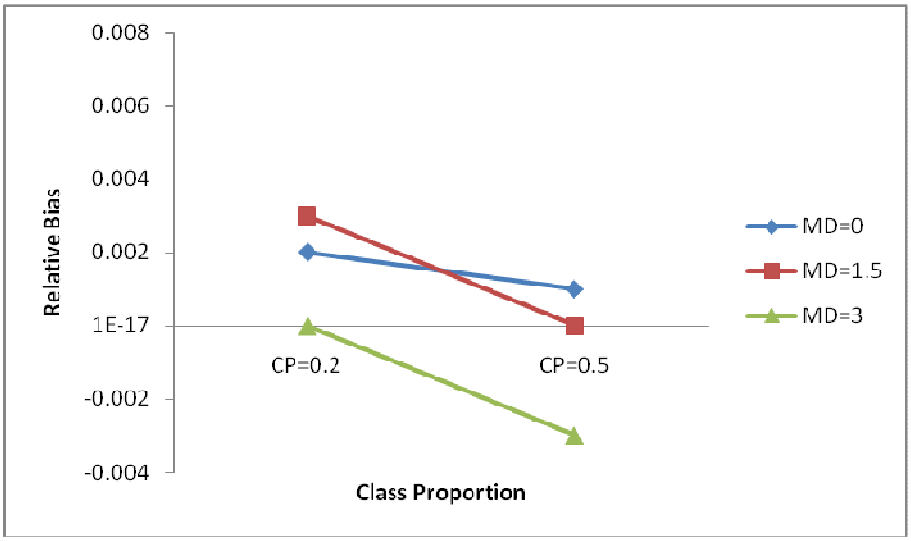


Figure 6f. Relative Bias of $\hat{\gamma}_0^{(1)}$ for SS=500, ES=0.25, and COV=0.6.



Figures 7a through 7d present the bias values of $\hat{\gamma}_3^{(1)}$ for different combinations of interaction effect sizes and class separation when sample size, covariance between the linear factors, and class proportions are fixed at the specified levels. The bias values of $\hat{\gamma}_3^{(1)}$ increase as the interaction effect size gets large with a class proportion 0.5. It can be clearly seen from the graphs that the bias values for zero class separation ranging from 0.569 to 2.604 are much larger when compared with those ranging from 0.033 to 0.554 for larger class separation. These patterns suggest that class separation (65.8%, large effect) and interaction effect size (8.9%, small effect) both are significant in explaining the variation in the bias of $\hat{\gamma}_3^{(1)}$.

Figure 7a. Bias of $\hat{\gamma}_3^{(1)}$ for SS=500, COV=0.2, CP=0.5.

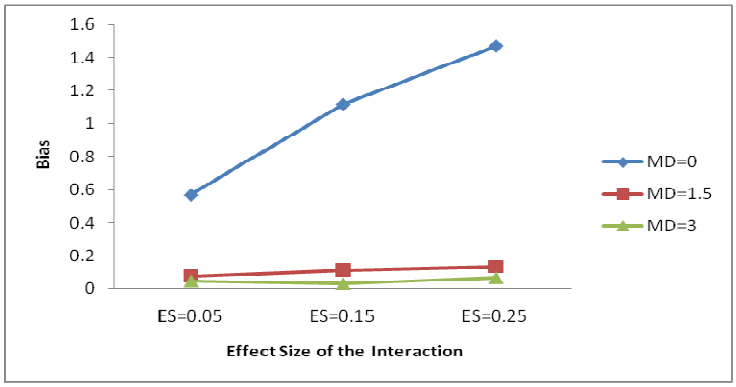


Figure 7b. Bias of $\hat{\gamma}_3^{(1)}$ for SS=500, COV=0.6, CP=0.5.

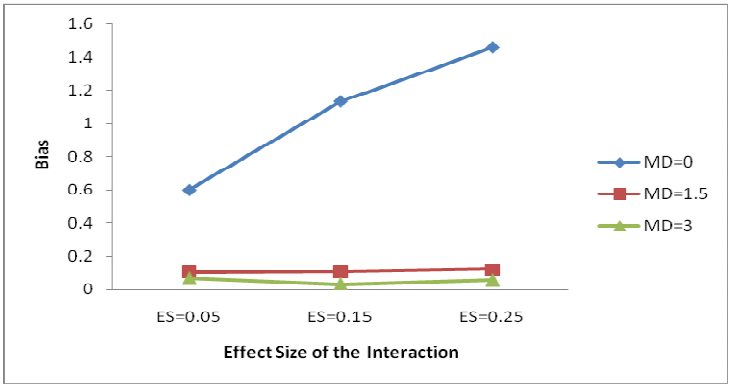


Figure 7c. Bias of $\hat{\gamma}_3^{(1)}$ for SS=500, COV=0.2, CP=0.2.

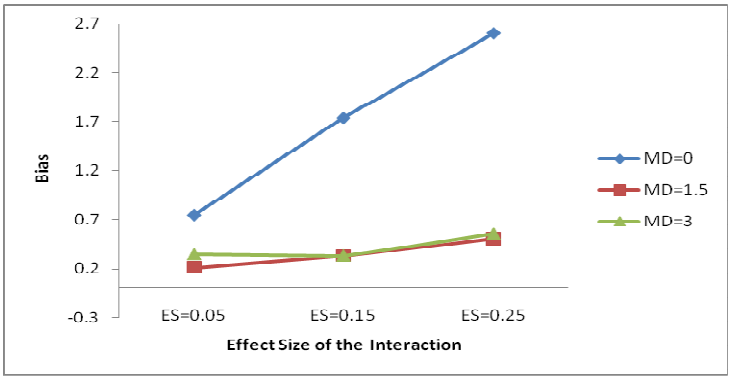
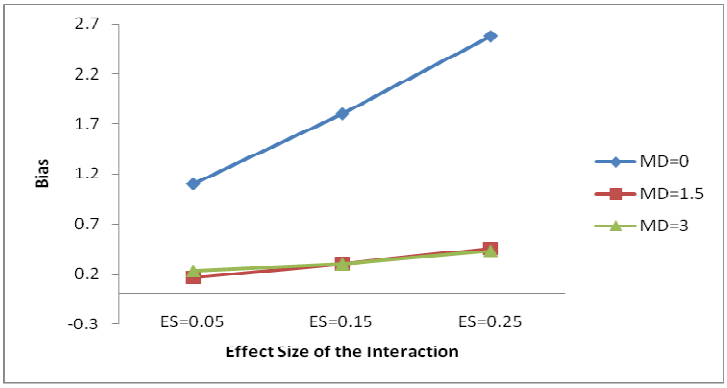


Figure 7d. Bias of $\hat{\gamma}_3^{(1)}$ for SS=500, COV=0.6, CP=0.2.



Figures 8a through 8d present the magnitude of the relative bias of $\hat{\gamma}_3^{(2)}$ for different combinations of effect size and class separation when sample size, covariance between the linear factors, and class proportion are fixed at the specified levels. From the graphs, the magnitude of the relative bias values of $\hat{\gamma}_3^{(2)}$ decrease with the effect size of the latent interaction term under the conditions with larger class separation (MD=1.5 and 3). The patterns help to understand that class separation (8.4%, small effect) and interaction effect size (37.1%, large effect) both have significant impacts on explaining the total variation in relative bias of $\hat{\gamma}_3^{(2)}$.

It is noted that some of the relative bias of $\hat{\gamma}_3^{(2)}$ for the conditions with small effect size (ES=0.05) are much larger in magnitude (ranging from 0.229 to 0.856) than that those (ranging from .082 to 0.526) for the conditions with larger effect sizes (ES=0.15 or 0.25). This may because the random variability of the parameter estimates for ES=0.05 is relatively large compared with those of larger effect sizes.

Figure 8a. Relative Bias of $\hat{\gamma}_3^{(2)}$ for SS=500, COV=0.2, CP=0.5.

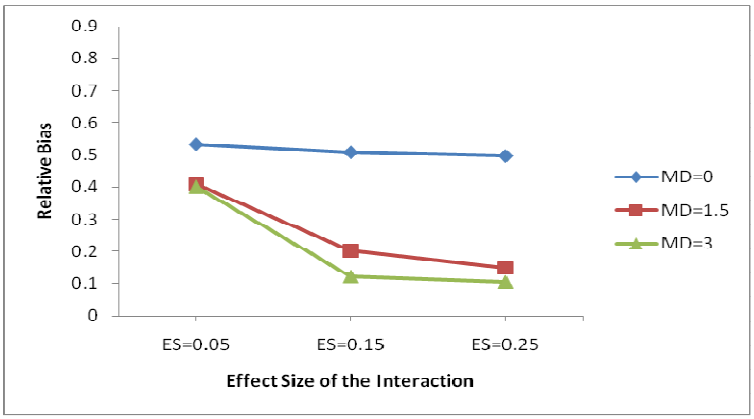


Figure 8b. Relative Bias of $\hat{\gamma}_3^{(2)}$ for SS=500, COV=0.2, CP=0.2.

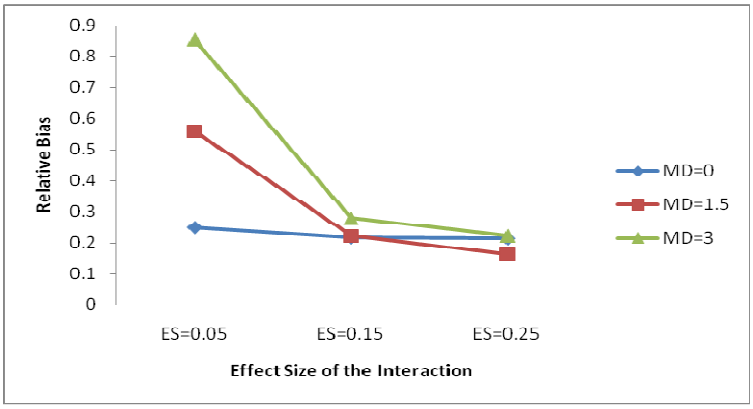


Figure 8c. Relative Bias of $\hat{\gamma}_3^{(2)}$ for SS=500, COV=0.6, CP=0.5.

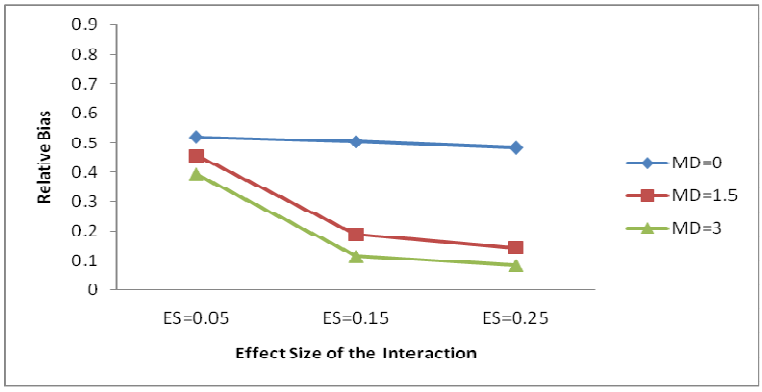


Figure 8d. Relative Bias of $\hat{\gamma}_3^{(2)}$ for SS=500, COV=0.6, CP=0.2.

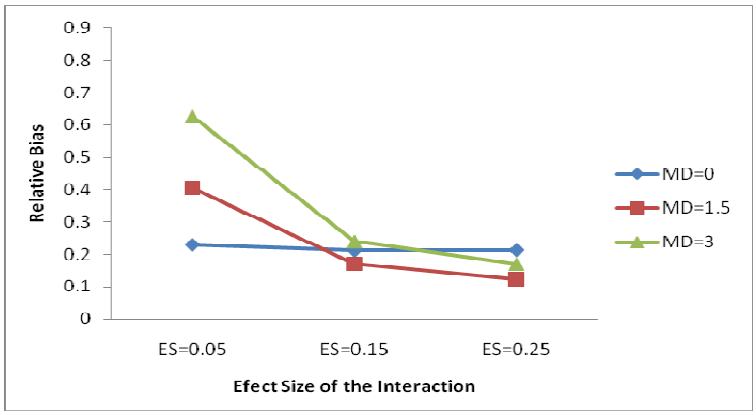


Figure 8e illustrates the magnitude of the relative bias values of $\hat{\gamma}_3^{(2)}$ for different combinations of effect size and class separation average across covariance and class separation when sample size is 500. It suggests that there is an interaction between effect size and class separation because the relation of the magnitude of relative bias of $\hat{\gamma}_3^{(2)}$ on effect size depends on different levels of class separation. Figure 8e is an aggregate graph for Figures 8a through 8d, which demonstrate the interaction patterns at the four specific levels of covariance and class proportion.

Figure 8e. Interaction between Effect Size and Class Separation Marginalized over Covariance and Class Proportion.

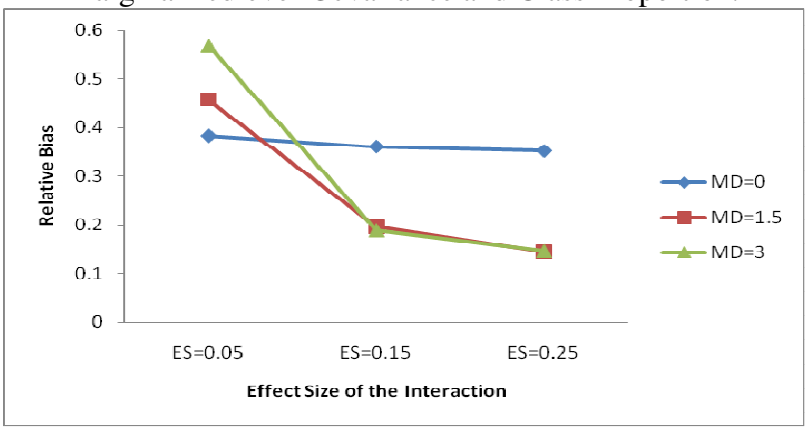


Figure 9a through 9f present the bias values of $\hat{\pi}_1$ for different combinations of class separation and class proportion when sample size, covariance between the linear factors, and effect size are fixed. Larger class separation tends to have lower bias in estimating π_1. Conditions with unequal class proportions have larger bias of $\hat{\pi}_1$ than those with equal class proportions. Specifically, the bias values for $\hat{\pi}_1$ range from -0.048 to 0.066 when class proportion is 0.5 whereas the bias values for $\hat{\pi}_1$ range from 0.171 to 0.356 when class proportion is 0.2. This also explains that class proportion (86.8%) and class separation (11.4%) have significant effects on the bias values of $\hat{\pi}_1$. Both the main effects account for almost all the variation in the bias of $\hat{\pi}_1$. Because the lines of bias on class separation are parallel for different values of class proportions, it suggests that there is no interaction between class separation and class proportion (see Table 9).

Figure 9a. Bias of $\hat{\pi}_1$ for SS=500, ES=0.05, COV=0.2.

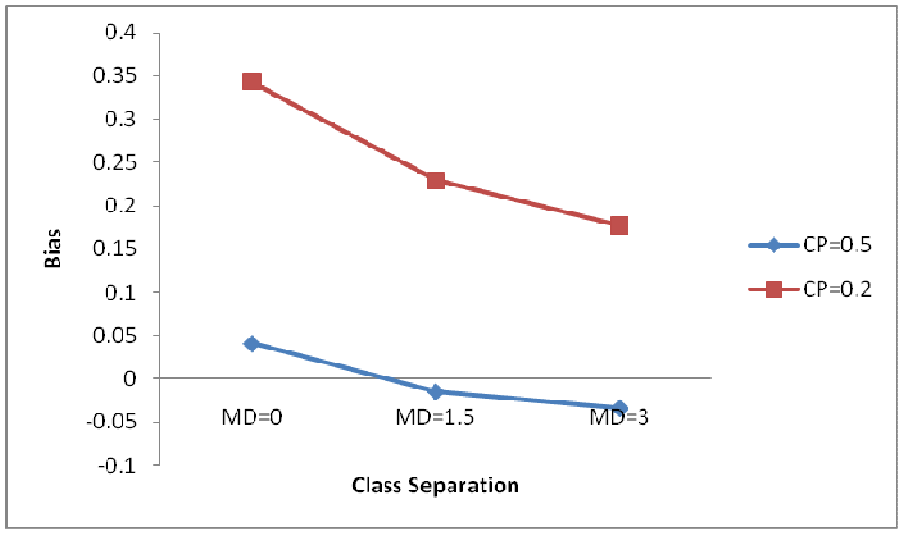


Figure 9b. Bias of $\hat{\pi}_1$ for SS=500, ES=0.15, COV=0.2.

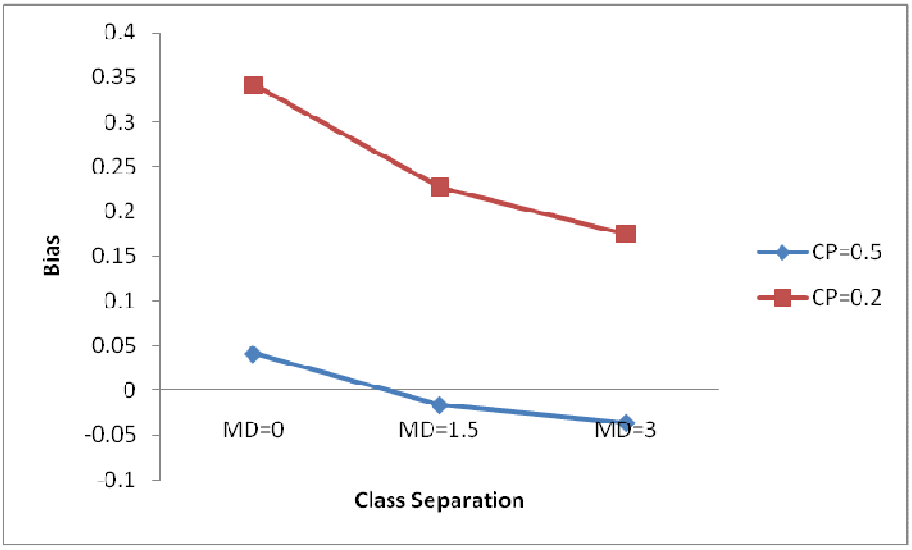


Figure 9c. Bias of $\hat{\pi}_1$ for SS=500, ES=0.25, COV=0.2.

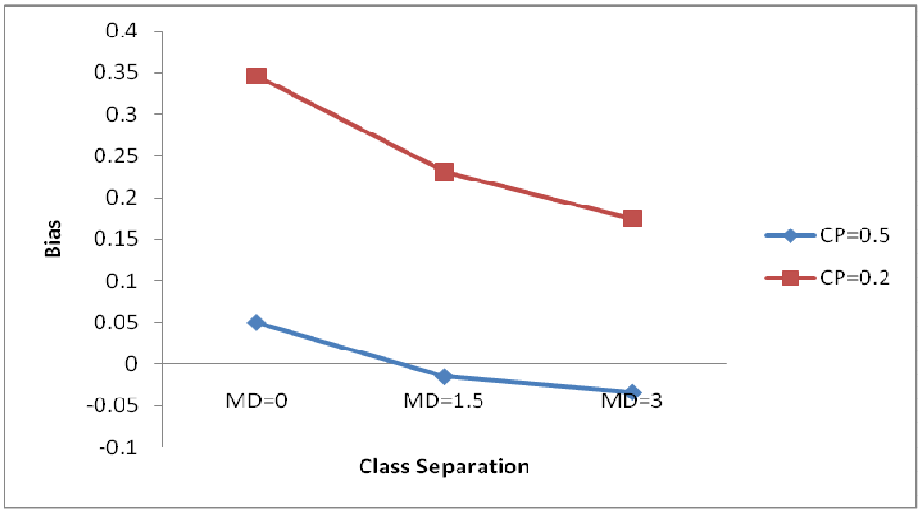


Figure 9d. Bias of $\hat{\pi}_1$ for SS=500, ES=0.05, COV=0.6.

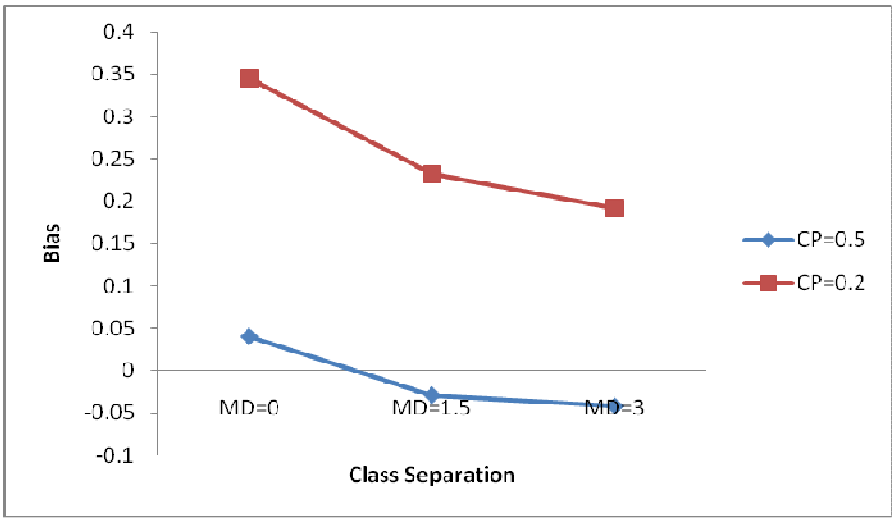


Figure 9e. Bias of $\hat{\pi}_1$ for SS=500, ES=0.15, COV=0.6.

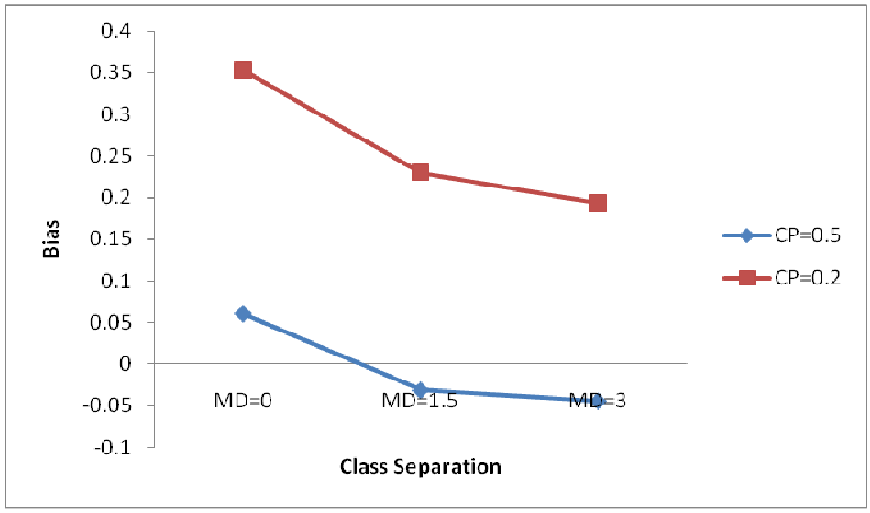


Figure 9f. Bias of $\hat{\pi}_1$ for SS=500, ES=0.25, COV=0.6.

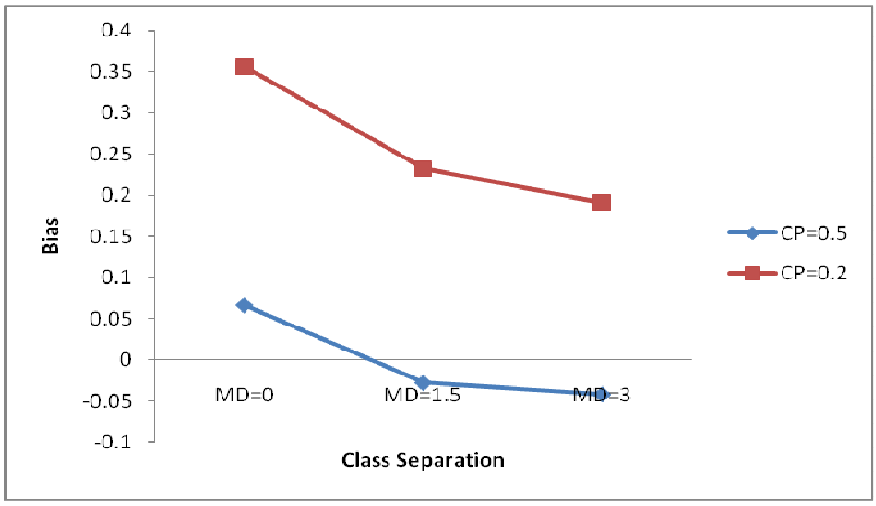


Table 17 presents the intervals from 5th percentile to 95th percentile of relative bias (or bias) values for all parameter estimates under conditions with all class separation levels (MD=0, 1.5, and 3) and conditions with larger class separation levels (MD=1.5 and 3). The overall empirical performances in terms of relatvie bias (or bias) are acceptable (<0.1) for all structural paramter estimates except $\hat{\gamma}_3^{(1)}$, $\hat{\gamma}_3^{(2)}$, and $\hat{\pi}_1$. $\hat{\gamma}_3^{(1)}$ and $\hat{\pi}_1$ have relatively smaller intervals if conditions of zero class separation are excluded.

Table 17. *5th to 95th Percentile of Relative Bias (or Bias) Values.*

Bias	MD=0, 1.5, 3	MD=1.5, 3
$\hat{\gamma}_0^{(1)}$	(0.000, 0.007)	(0.000, 0.007)
$\hat{\gamma}_0^{(2)}$	(-0.054, 0.004)	(-0.055, -0.028)
$\hat{\gamma}_1$	(0.015, 0.107)	(0.034, 0.111)
$\hat{\gamma}_2$	(0.013, 0.068)	(0.016, 0.071)
$\hat{\gamma}_3^{(1)}$	(0.053, 2.343)	(0.046, 0.48)
$\hat{\gamma}_3^{(2)}$	(-0.606,- 0.104)	(-0.646, -0.096)
$\hat{\pi}_1$	(-0.045, 0.342)	(-0.047, 0.230)

4.3.3 Precision.

The SE/SD ratio is used to evaluate the precision of the standard error estimates. If the ratio is close to one, it indicates that standard errors estimated by the model reflect the variation in the population. Table 18 presents the intervals of 5th percentile to 95th percentile of precision estimates for all structural parameters under conditions with all class separation levels (MD=0, 1.5, and 3) and with larger class separation levels (MD=1.5 and 3). It is clear that the range of precision estimates for all structural parameters become smaller if the conditions of zero class separation are excluded, meaning that all the structural parameters generally have acceptable range for precision estimates. 90 percent of the SE/SD estimates for all the structural parameters are within the range of 0.9 and 1.2 under the conditions with larger class separation. This suggests that the overall empirical performances in terms of precision are satisfactory for all structural parameters when class separation in the exogenous factor means is large.

Table 18. *5th to 95th Percentile of SE/SD Estimates.*

SE/SD	MD=0, 1.5, 3	MD=1.5, 3
$\hat{\gamma}_0^{(1)}$	(0.964, 2.044)	(0.964, 1.129)
$\hat{\gamma}_0^{(2)}$	(0.984, 1.213)	(0.989, 1.173)
$\hat{\gamma}_1$	(0.960, 1.177)	(0.982, 1.158)
$\hat{\gamma}_2$	(0.954, 1.204)	(0.999, 1.135)
$\hat{\gamma}_3^{(1)}$	(0.984, 6.190)	(0.984, 1.170)
$\hat{\gamma}_3^{(2)}$	(0.890, 1.140)	(0.973, 1.140)

4.4 Classification Results

Section 4.4.1.1 and Section 4.4.1.2 provide the results of factorial ANOVA to assess the impact of the simulated factors and their combinations on entropy and correct assignment probability, respectively. Results of entropy (Section 4.4.2) and correct assignment probability (Section 4.4.3) will be discussed. Table 19 presents the entropy and correct assignment probability under different conditions for all sample sizes.

Table 19. *Classification Results for All Sample Sizes.*

π_1	MD	$R^2_{\gamma_3^{(2)}}$	n=500				n=1000				n=2000			
			ϕ_{12} = 0.2		ϕ_{12} = 0.6		ϕ_{12} = 0.2		ϕ_{12} = 0.6		ϕ_{12} = 0.2		ϕ_{12} = 0.6	
			entropy	assign	entropy	assign	entropy	assign	entropy	assign	entropy	assign	entropy	assign
0.5	0	0.05	0.788	0.501	0.817	0.501	0.780	0.499	0.823	0.499	0.775	0.500	0.832	0.520
		0.15	0.782	0.503	0.811	0.500	0.773	0.501	0.813	0.501	0.774	0.501	0.819	0.500
		0.25	0.782	0.502	0.803	0.502	0.768	0.501	0.802	0.501	0.769	0.501	0.797	0.501
	1.5	0.05	0.862	0.676	0.892	0.657	0.852	0.677	0.882	0.657	0.848	0.676	0.878	0.658
		0.15	0.862	0.676	0.891	0.657	0.852	0.676	0.882	0.657	0.848	0.676	0.879	0.657
		0.25	0.861	0.677	0.890	0.659	0.853	0.677	0.884	0.660	0.848	0.678	0.879	0.659
	3	0.05	0.881	0.740	0.902	0.716	0.870	0.740	0.892	0.716	0.869	0.740	0.891	0.717
		0.15	0.880	0.740	0.901	0.716	0.871	0.741	0.893	0.717	0.868	0.742	0.890	0.717
		0.25	0.879	0.742	0.900	0.718	0.872	0.742	0.894	0.719	0.868	0.742	0.890	0.718
0.2	0	0.05	0.785	0.475	0.813	0.474	0.777	0.487	0.820	0.485	0.773	0.495	0.825	0.493
		0.15	0.780	0.476	0.809	0.469	0.769	0.487	0.812	0.483	0.772	0.495	0.818	0.490
		0.25	0.777	0.472	0.800	0.467	0.765	0.484	0.805	0.476	0.763	0.492	0.795	0.478
	1.5	0.05	0.860	0.655	0.890	0.642	0.849	0.658	0.881	0.644	0.846	0.658	0.877	0.644
		0.15	0.860	0.655	0.890	0.642	0.850	0.658	0.881	0.643	0.846	0.660	0.877	0.645
		0.25	0.858	0.655	0.889	0.644	0.851	0.660	0.881	0.647	0.846	0.660	0.877	0.647
	3	0.05	0.878	0.729	0.900	0.704	0.869	0.732	0.891	0.706	0.867	0.733	0.889	0.708
		0.15	0.878	0.731	0.900	0.704	0.870	0.732	0.892	0.708	0.866	0.733	0.888	0.708
		0.25	0.878	0.729	0.900	0.705	0.870	0.731	0.892	0.707	0.866	0.733	0.889	0.708

4.4.1 Factorial ANOVA on classification results

4.4.1.1 Factorial ANOVA on Entropy. The factors that are significant in explaining the variation in entropy are reported in Table 20. Covariance between the linear factors had a significant effect on entropy and explained 12.2% of the total variance. In addition, class separation affected entropy significantly, accounting for most (85%) of the total variance.

Table 20. *Proportion of Variance Explained in Entropy Exceeds 5%*

Factors	entropy
SS	
COV	12.2%
CP	
CS	85.0%
ES	

4.4.1.2 Factorial ANOVA on Correct assignment. Table 21 presents the factors that account for more than 5% of the total variation in correct assignment probability for all manipulated factors and their combinations. The only significant main effect is class separation, which explained most (98.4%) of the variation in correct assignment.

Table 21. *Proportion of Variance Explained in Correct Assignment Exceeds 5%*

Factors	Assignment
SS	
COV	
CP	
CS	98.4%
ES	

4.4.2 Entropy.

Entropy is an indicator of classification quality. Values close to one indicates good classification and values greater than or equal to 0.8 are usually considered acceptable (Muthén et al., 2002). The results show that the entropy values range from 0.763 to 0.902 across all simulation conditions. Compared to the zero separation models (ranging from 0.763 to 0.832), larger separation models have increasing and acceptable entropy values (ranging from 0.846 to 0.902). This pattern is illustrated by Figures 10a through 10f, which present the entropy values for different combinations of class separation and covariance when sample size, effect size, and class proportion are fixed at specified levels. Furthermore, the entropy values increase with the covariance between the linear factors. Class separation and covariance between linear factors are both significant in explaining the variation in entropy (see Table 20). Because the lines of entropy on class separation are parallel for different levels of covariance, this suggests that there is no interaction between these two factors.

Figure 10a. Entropy for SS=500, ES=0.05, and CP=0.5.

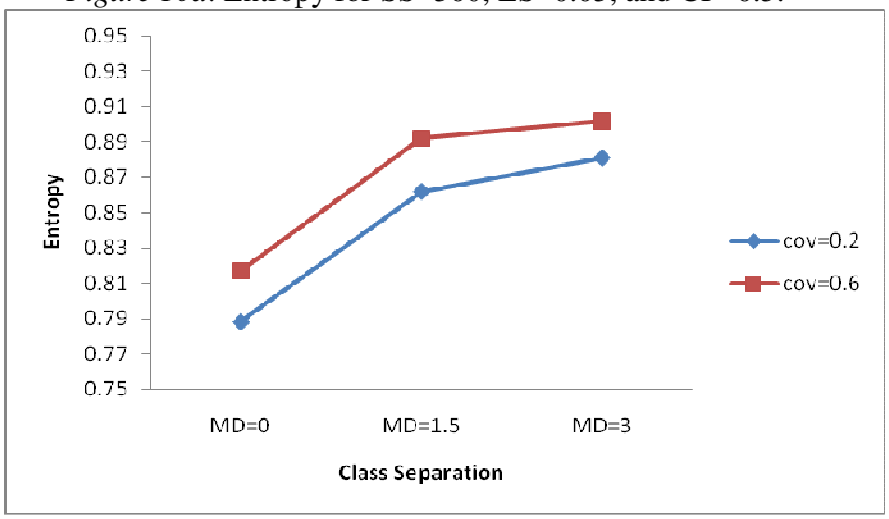


Figure 10b. Entropy for SS=500, ES=0.15, and CP=0.5.

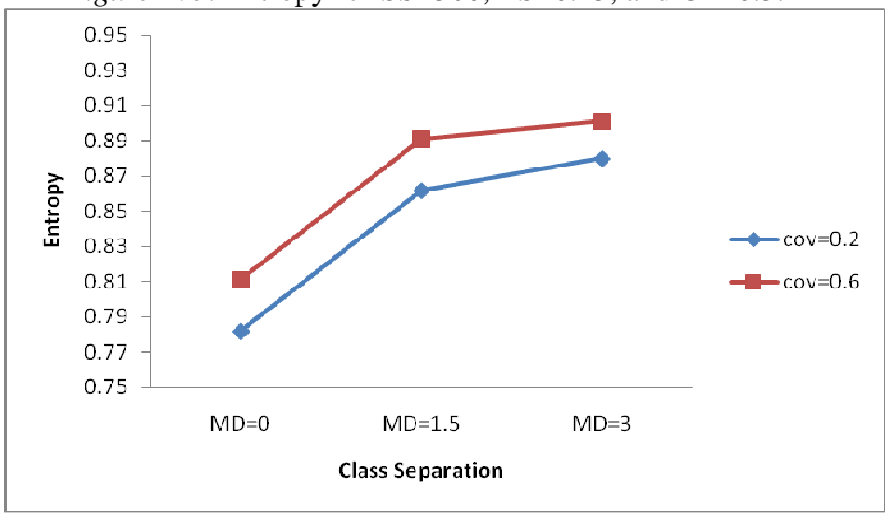


Figure 10c. Entropy for SS=500, ES=0.25, and CP=0.5.

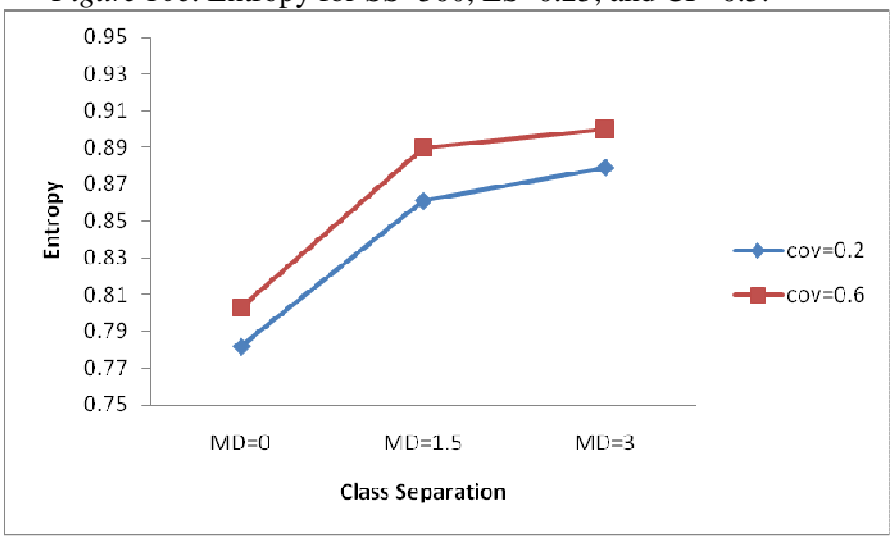


Figure 10d. Entropy for SS=500, ES=0.05, and CP=0.2.

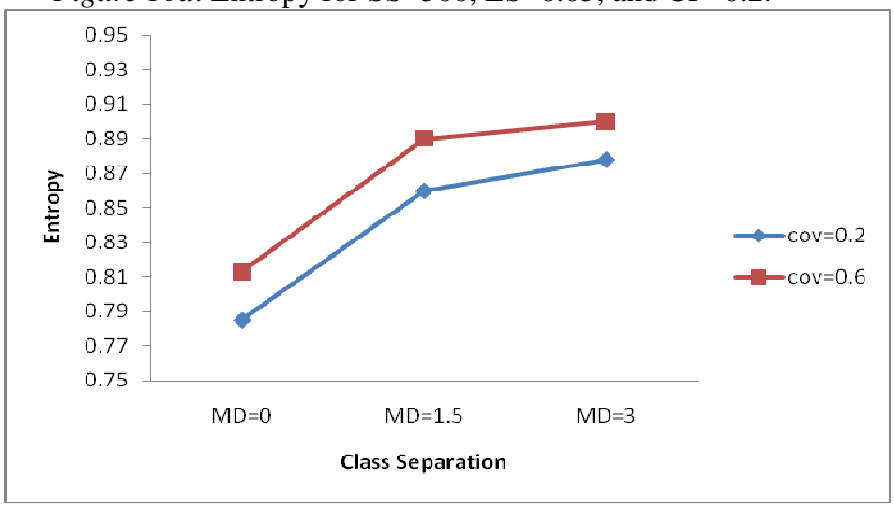


Figure 10e. Entropy for SS=500, ES=0.15, and CP=0.2.

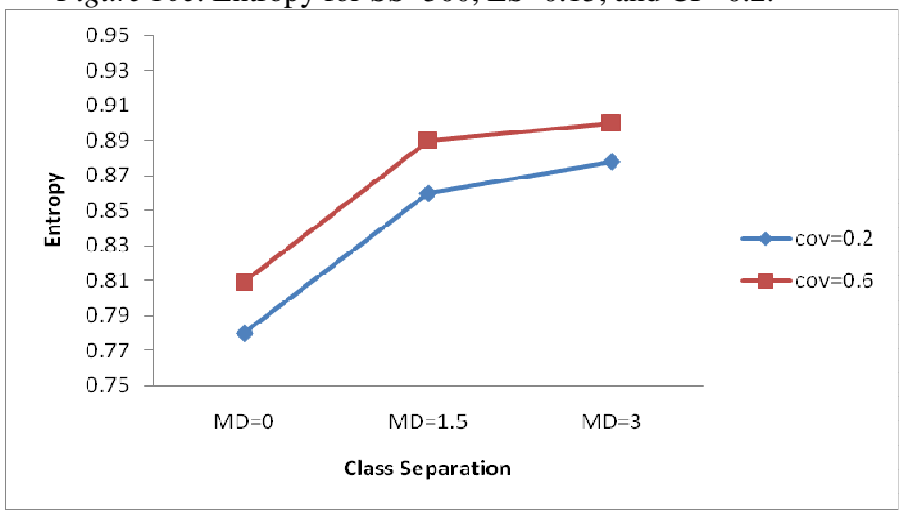
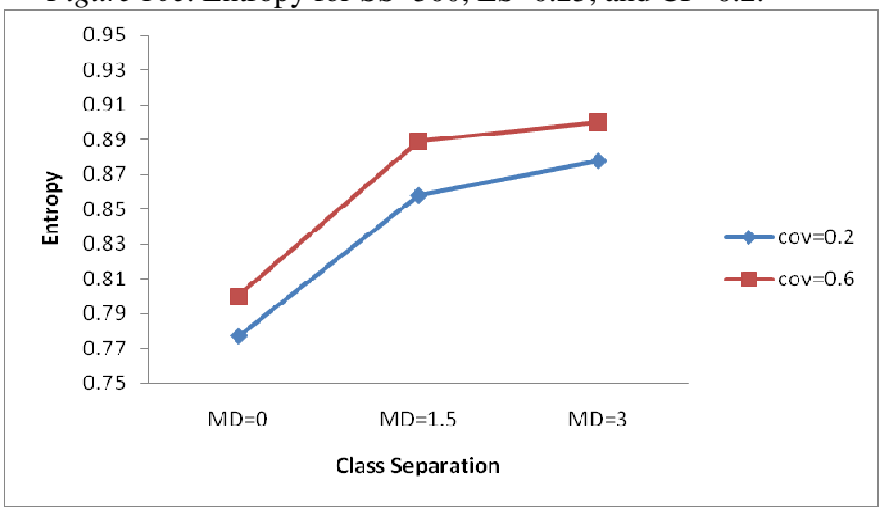


Figure 10e. Entropy for SS=500, ES=0.25, and CP=0.2.



4.4.3 Correct class assignment.

Correct class assignment refers to the proportion of subjects whose true class is consistent with the class he/she is assigned based on the posterior probability. This probability ranges from 0.467 to 0.742 across all simulation conditions. Compared to the conditions of zero separation (ranging form 0.467 to 0.403), conditions with larger class separation have noticeable improvement in correct assignment (ranging from 0.644 to 0.742). This pattern can be observed in Figures 11a through 11f, which Indicate that class separation has a substantially significant impact on correct assignment. This main effect accounts for almost all the variation (98.4%) in correct assignment. In addition, conditions with balanced class proportions have slightly increased correct assignment probability (about 1%-4%) compared with those with unbalanced class proportions. This may explain why class proportion is not a significant factor for correct assignment. The lines in all the figures appear to be parallel, indicating that there is no interaction between class separation, and class proportion.

Figure 11a. Correct Assignment for SS=500, ES=0.05, and COV=0.2.

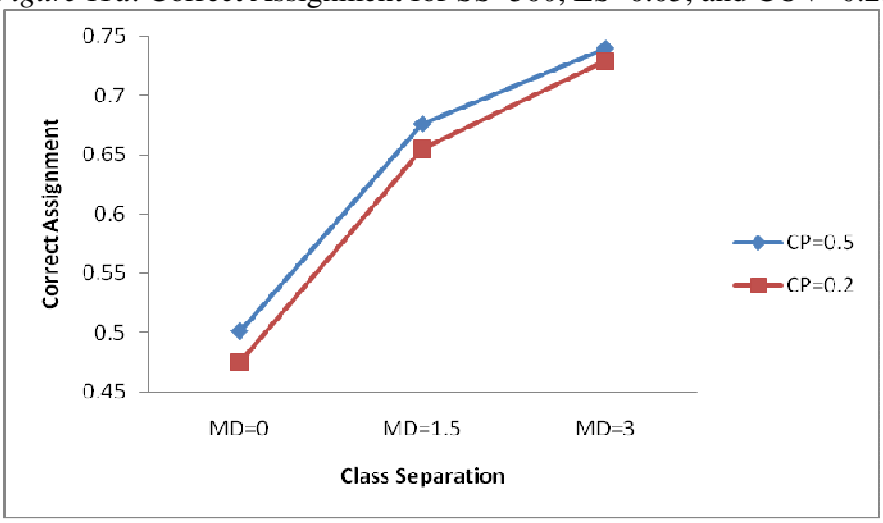


Figure 11b. Correct Assignment for SS=500, ES=0.15, and COV=0.2.

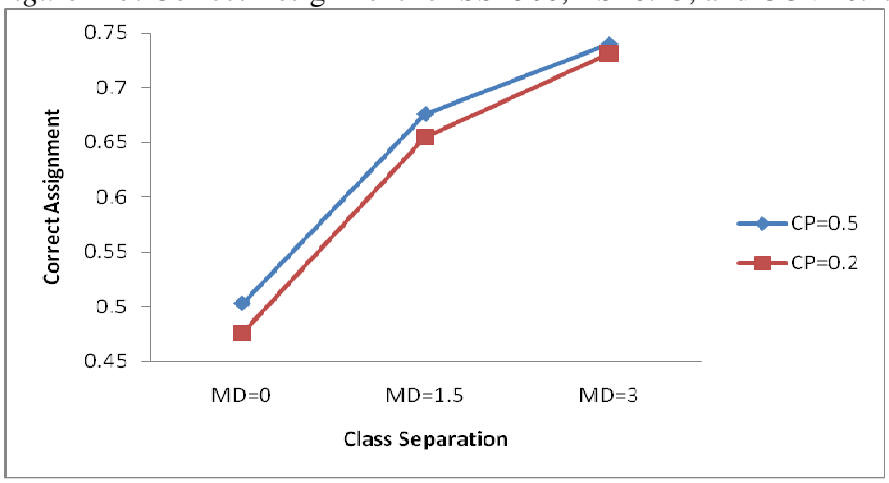


Figure 11c. Correct Assignment for SS=500, ES=0.25, and COV=0.2

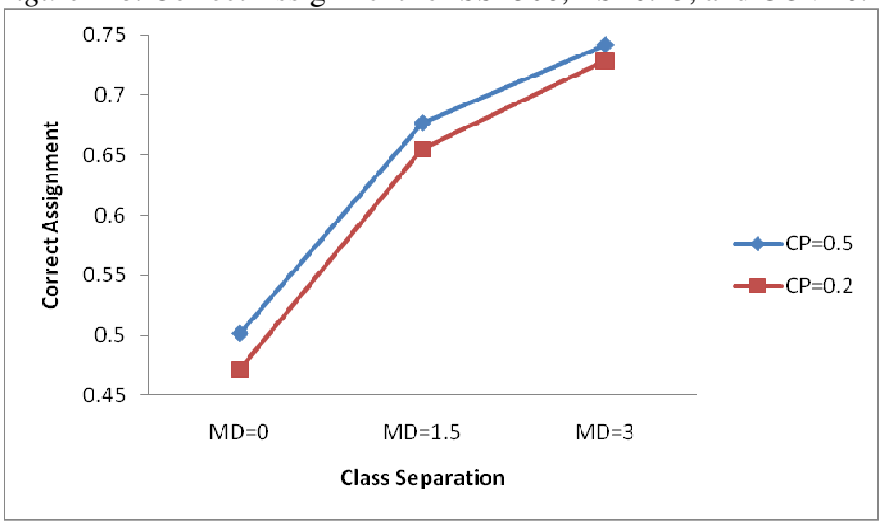


Figure 11d. Correct Assignment for SS=500, ES=0.05, and COV=0.6.

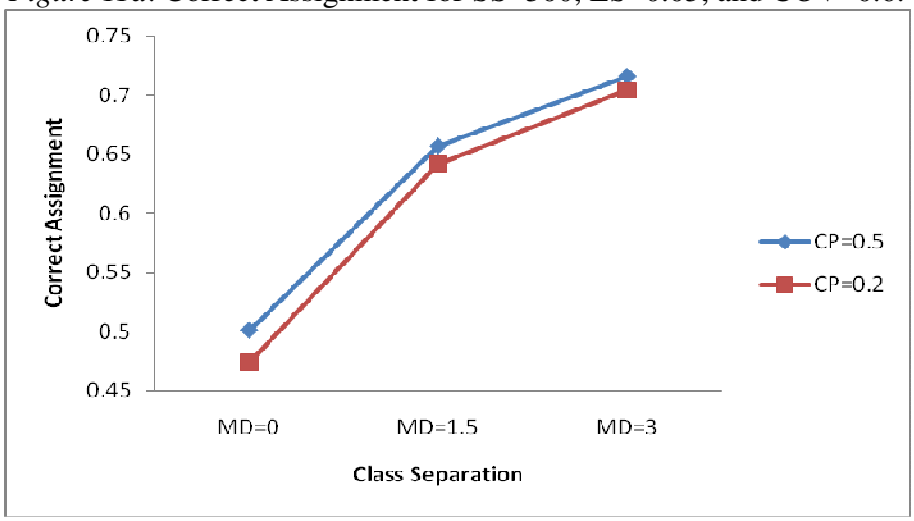


Figure 11e. Correct Assignment for SS=500, ES=0.15, and COV=0.6.

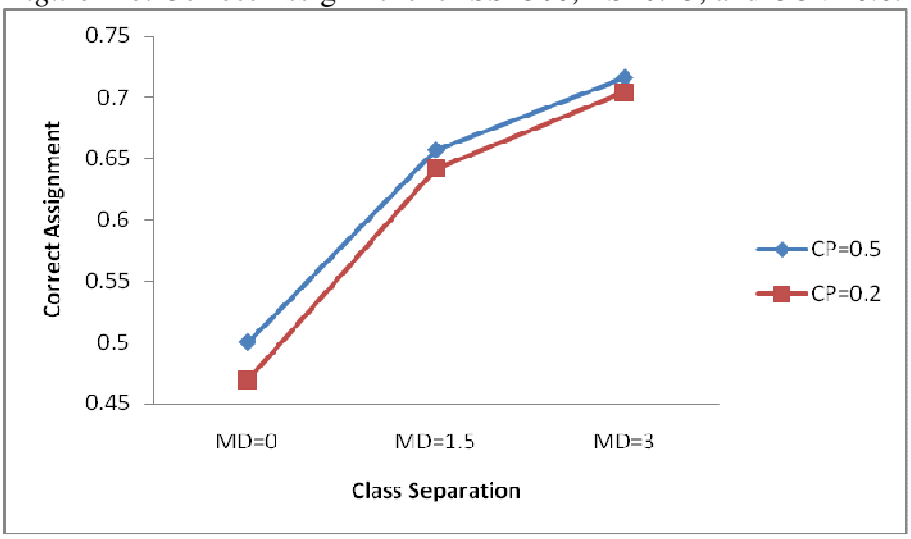
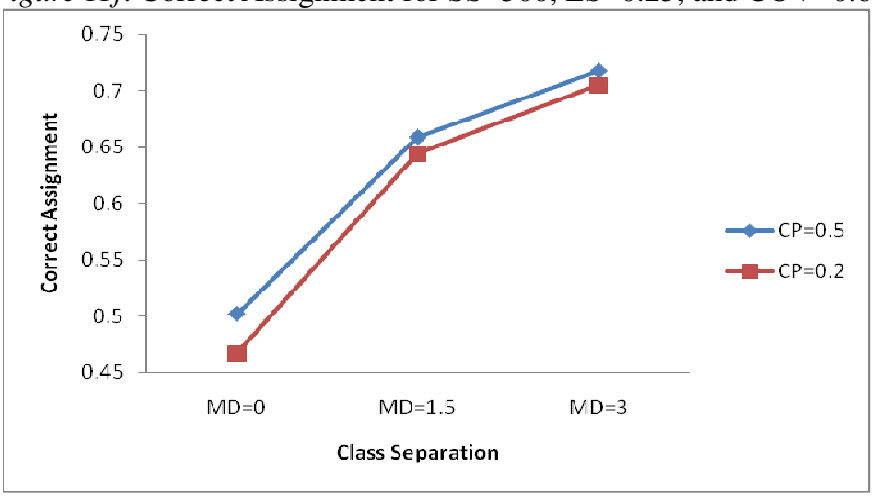


Figure 11f. Correct Assignment for SS=500, ES=0.25, and COV=0.6.



Chapter 5: Discussion

The simulation work in this study investigates the performance of the unconstrained approach for a mixture latent variable interaction model in terms of parameter recovery and classification quality. It also examines the identification of latent classes in the data generated from populations with one- and two-class interaction effects based on model fit statistics. In this chapter, major findings are outlined in Section 5.1. Discussion is provided in Section 5.2. Recommendations for researchers are addressed in Section 5.3. Finally, limitations and suggestions for future study are presented in Section 5.4. The conclusions of the current study are limited to the circumstances where convergence occurred.

5.1 Summary of Findings

5.1.1 Parameter recovery.

5.1.1.1 Relative bias (or bias). The relative bias (or bias) values of structural parameter estimates are acceptable except for those of the class-specific interaction estimates and class proportion estimates. Table 22 summarizes the factors which significantly affect the criteria measures in this study. It is clear that class separation between the exogenous factor means has a significant effect on the relative bias (or bias) values of all structural parameter estimates. Covariance between the linear factors and class proportion significantly affect the relative bias (or bias) of some parameter estimates. Surprisingly, total sample size has no significant impact on the relative bias (or bias) of any parameter estimates. It is worth noting that the relative bias (or bias) of the estimates of both class-specific interaction coefficients are significantly affected by the

latent interaction's effect size.

The results of preliminary analyses indicated that bias and precision of the class-specific interaction estimates did not improve much when sample sizes increased to 5000 or 10000. Weiss (2010) also found in a fairly comprehensive study of one-class latent interaction models that sample size did not significantly affect bias and precision of the parameter estimates. This non-intuitive outcome could in a later study be further investigated. Based on the results of the current study along with those reported in Weiss, both mixture latent interaction models and single class models should be investigated.

5.1.1.2 Precision. Class separation between the exogenous factor means is the only factor significantly influencing the precision estimates of $\gamma_0^{(1)}$, $\gamma_3^{(1)}$, and $\gamma_3^{(2)}$ (see Table 22). The SE/SD ratios of all the structural parameters are fairly close to 1 for larger separation models (i.e., MD=1.5 and 3).

Table 22. *Summary of Significant Factors on Outcome Measures.*

Factor	**Criteria Measure**				
	RB (Bias)	**Precision**	**Entropy**	**Assign**	**Converge**
Sample Size	NS	NS	NS	NS	NS
Covariance	some parameters S (1 out of 7)	NS	S	NS	NS
Class Proportion	some parameters S (4 out of 7)	NS	NS	NS	NS
Class Separation	all parameters S (7 out of 7)	some parameters S (3 out of 7)	S	S	S
Interaction Effect Size	some parameters S (3 out of 7)	NS	NS	NS	NS

S=Significant, NS=Nonsignificant.

5.1.2 Classification results.

5.1.2.1 Entropy. Class separation and covariance between the linear factors significantly affect the entropy values (Table 22). With the increase of the covariance between the linear factors, the entropy values increases. Entropy improves with increasing class separation. In addition, larger separation models have acceptable entropy values (ranging from 0.846 to 0.902).

5.1.2.2 Correct* class *assignment. Class separation has a significant effect on correct class assignment. Correct assignment improves dramatically with increasing class separation. Models with larger class separation have larger correct class assignment ranging from 0.644 to 0.742, than those with zero class separation ranging from 0.467 to 0.503.

5.1.2.3 Convergence. Class separation is the only factor significantly influencing convergence rates. Conditions with larger class separation have higher convergence rate, which ranges from 0.982 to 1, than those with zero class separation.

5.1.3 Identification of latent classes. All model fit statistics used in this study (i.e., AIC, BIC, and ABIC) select two-class interaction models 100% of the time in fitting the data generated from populations with either one- or two-class latent interaction effects.

5.2 Discussion

It is expected that the performances of the models with no separation between the exogenous factor means in parameter recovery and classification results are not as good as those with larger separation. The population structural equation (see Equation 16) for generating the data consists of two submodels- one with two linear effects only and the

other with an additional latent interaction effect. The difference in the endogenous factor means between two classes is

$$\begin{aligned}
\mathrm{E}(\eta^{(2)}) - \mathrm{E}(\eta^{(1)}) &= \left[\gamma_0 + \gamma_1 \mathrm{E}(\xi_1^{(2)}) + \gamma_2 \mathrm{E}(\xi_2^{(2)}) + \gamma_3^{(2)} \mathrm{E}(\xi_1^{(2)}\xi_2^{(2)}) + \mathrm{E}(\zeta^{(2)})\right] \\
&\quad - \left[\gamma_0 + \gamma_1 \mathrm{E}(\xi_1^{(1)}) + \gamma_2 \mathrm{E}(\xi_2^{(1)}) + \gamma_3^{(1)} \mathrm{E}(\xi_1^{(1)}\xi_2^{(1)}) + \mathrm{E}(\zeta^{(1)})\right] \\
&= \gamma_3^{(2)} \mathrm{E}(\xi_1^{(2)}\xi_2^{(2)}) \qquad \text{since} \quad \gamma_3^{(1)} = 0, \quad \mathrm{E}(\xi_1^{(1)}) = \mathrm{E}(\xi_2^{(1)}) = 0, \\
&\quad \mathrm{E}(\xi_1^{(2)}) - \mathrm{E}(\xi_1^{(1)}) = 0, \text{ and } \quad \mathrm{E}(\xi_2^{(2)}) - \mathrm{E}(\xi_2^{(1)}) = 0. \\
&= \gamma_3^{(2)} \left[\mathrm{E}(\xi_1^{(2)}) \mathrm{E}(\xi_2^{(2)}) + \phi_{12}\right] \\
&= \gamma_3^{(2)} \phi_{12}.
\end{aligned}$$

If the exogenous factors of both classes share the same mean, then the difference between the two class endogenous factor means depends on $\gamma_3^{(2)}$ and ϕ_{12}. This difference is small ($-1 \le \phi_{12} \le 1$ since variances of ξ_1 and ξ_1 are both assumed to be 1) unless $\gamma_3^{(2)}$ is very large. In other words, no separation between the exogenous factor means would cause two class distributions of the endogenous factors to largely overlap. It will be difficult to differentiate whether the observation is drawn from the class with linear effects or the class with an additional cross-product effect. Misclassifying the observations to their true group leads to a biased estimation of the class proportion, which consequently affects the estimation of the parameters within the class and results in biased estimates. Class-specific endogenous factor means and class-specific standard deviations corresponding to different MDs, interaction effects, and covariances are provided in Appendix D.

The occurrence of large relative bias (or bias) values in $\hat{\gamma}_3^{(1)}$ and $\hat{\gamma}_3^{(2)}$ for large class separation may result from the bias in estimating class proportion. Another possible

reason is that the unconstrained method used maximum likelihood estimation, which assumes that the observed variables are distributed as multivariate normal. However, this distributional assumption is violated since the product indicators are non-normal and as a result so are the observed variables for the endogenous factor as well (Jöreskog & Yang, 1996). This violation will consequently contribute to the bias of parameter estimates.

The discovery that two-class interaction models are superior to one-class interaction models in fitting the observed data via fit criteria used in this study is consistent with the previous research (e.g., Bauer & Curran, 2003; 2004). It is clear that the data simulated from models with one- or two-class latent interaction effects are both non-normally distributed. The inclusion of additional latent classes provides a greater resemblance to non-normality in the observed data which leads to an increased model fit even when the population only consists of one group. In other words, the estimated multiple latent classes allow for modeling the violation of distributional assumptions such as multivariate normality for the observed variables or linear relationship between variables. Therefore, it is difficult to distinguish whether the estimated latent classes reflect the true population structure or are only used to better fit a complex but homogenous distribution (Bauer & Curran, 2003; 2004). This dilemma has existed for over a century, as described by Pearson (1895): "The question may be raised, how are we to discriminate between a true curve of skew type and a compound curve, supposing we have no reason to suspect our statistics or priori mixture (p. 394)."

Another possible reason for overestimating the number of latent classes is the mismatch between data and model assumption. The ML estimation for the unconstrained approach typically assumes that the latent variables are normally distributed although this

approach does not impose any constraints based on normality assumption. However, the data for the mixture latent interaction models are non-normally distributed. This lack of fit which is caused by model misspecification can be compensated by estimating additional latent classes (Bauer & Curran, 2004).

5.3 Recommendations

The current study uses the unconstrained approach (Marsh et. al., 2004) due to its ease of implementation in SEM software. The estimation provides a workable solution to the two-class latent variable interaction model which includes obtaining estimates of model parameters and classifying observations to their most likely latent class. Despite large relative bias (or bias) values in estimating class interaction effects, the precision estimates for all structural parameters and classification results are satisfactory when class separation between the exogenous factor means is sufficient. As long as sample size is moderately large (e.g., 500 in this study), increasing sample size does not seem to improve the performance of the outcome measures. Consistent with prior research (e.g., Lubke & Muthén, 2007), class separation between the exogenous factor means is a critical factor due to its significant effect on all the outcome measures in this study.

As suggested in Bauer and Curran (2003; 2004), a multiclass model would better approximate the data drawn from a non-normal distribution even with only one group in the population than a single-class model. Therefore, to conclude that population heterogeneity exists on the basis of model fit statistics is not sufficient unless it is also justified by a substantive theory. That is, identifying latent classes in the observed data from a (mixture) latent interaction model via fit criteria may be used in a more

confirmatory mode.

The mixture latent interaction models allow for accommodating unobserved population heterogeneity in the interaction patterns since it is likely that participants in the sample have different sets of parameter values. For instance, the relationship between latent achievement in college and latent ability may depend on different levels of latent motivation for one subpopulation whereas the relationship between latent achievement in college and latent ability remains the same irrespective of the levels of latent motivation (i.e., linear) for another subpopulation. The heterogeneity here is not able to be defined based on an observed variable. Hence, the membership of the observations belonging to a subpopulation needs to be inferred from the data and the model. If theory indicates that the population is likely to have different interaction patterns, then it is recommended to use theory to guide the estimation and interpretation of the model.

5.4 Limitations and Suggestions for Future Study

Other factors rather than the five used in the current study, various levels of the manipulated factors, or different constraint specifications can be examined for the mixture latent variable interaction models. For example, additional levels for class separation such as 1MD and 2MD can be included for further investigation. In addition, the invariance of within-class covariance matrix can be assumed to be class-specific.

In terms of estimation methods, the current study uses the unconstrained approach (Marsh et. al., 2004), which creates product indicators for the latent interaction variables to estimate a two-class latent interaction model. Other estimation methods, such as marginal maximum likelihood estimation via the EM algorithm or Bayesian estimation

via MCMC methods, can be used to estimate the mixture latent variable interaction model.

Identifying latent classes in the observed data according to fit indices is limited to being used in a more confirmatory base. Comparing the traditional fit statistics among models is not sufficient to make inferences about the number and the characteristics of the latent classes. Although more procedures have been developed to deal with this issue such as bootstrap likelihood ratio test (Aiken, Anderson, & Hinde, 1981; Arminger et al., 1999; McLachlan, 1987) and the adjusted likelihood ratio test (Lo, Mendell, & Rubin, 2001), these procedures may have similar limitations as other fit criteria do since they all depend on the same pieces of information (Bauer & Curran, 2003). Thus, new fit criteria or model-checking processes are needed to better inform researchers whether the non-normally distributed data come from a true mixture or a single-class population.

Adding covariates such as age and gender to the nonlinear structural equation is helpful in better predicting the endogenous latent variables (Song & Lee, 2006). In addition, it also improves the correct class membership assignment in factor mixture models (Lubke & Muthén, 2007). Therefore, future studies can incorporate covariates to the mixture latent variable interaction model to examine whether the performance of the model can be improved.

Appendix A. Covariance between errors

$$x_1 = \tau_{x_1} + \xi_1 + \delta_1$$

$$x_4 = \tau_{x_4} + \xi_2 + \delta_4$$

$$x_1 x_4 = (\tau_{x_1} + \xi_1 + \delta_1)(\tau_{x_4} + \xi_2 + \delta_4)$$

$$= \tau_{x_1}\tau_{x_4} + \tau_{x_1}\xi_2 + \tau_{x_1}\delta_4 + \tau_{x_4}\xi_1 + \xi_1\xi_2 + \xi_1\delta_4 + \tau_{x_4}\delta_1 + \xi_2\delta_1 + \delta_1\delta_4$$

$$= \tau_{x_1}\tau_{x_4} + (\tau_{x_1}\xi_2 + \tau_{x_4}\xi_1 + \xi_1\xi_2) + (\xi_1\delta_4 + \xi_2\delta_1 + \tau_{x_4}\delta_1 + \tau_{x_1}\delta_4 + \delta_1\delta_4)$$

$$Cov(\delta_1, (\xi_1\delta_4 + \xi_2\delta_1 + \tau_{x_4}\delta_1 + \tau_{x_1}\delta_4 + \delta_1\delta_4))$$

$$= \mathrm{E}\left[\delta_1(\xi_1\delta_4 + \xi_2\delta_1 + \tau_{x_4}\delta_1 + \tau_{x_1}\delta_4 + \delta_1\delta_4)\right] - \mathrm{E}(\delta_1)\mathrm{E}(\xi_1\delta_4 + \xi_2\delta_1 + \tau_{x_4}\delta_1 + \tau_{x_1}\delta_4 + \delta_1\delta_4)$$

$$= \left[\mathrm{E}(\delta_1\xi_1\delta_4) + \mathrm{E}(\xi_2\delta_1^2) + \tau_{x_4}\,\mathrm{E}(\delta_1^2) + \tau_{x_1}\,\mathrm{E}(\delta_1\delta_4) + \mathrm{E}(\delta_1^2\delta_4)\right]$$
$$- \mathrm{E}(\delta_1)\mathrm{E}(\xi_1\delta_4 + \xi_2\delta_1 + \tau_{x_4}\delta_1 + \tau_{x_1}\delta_4 + \delta_1\delta_4)$$

$$= \mathrm{E}(\delta_1)E(\xi_1)\mathrm{E}(\delta_4) + \mathrm{E}(\xi_2)\mathrm{E}(\delta_1^2) + \tau_{x_4}\,\mathrm{E}(\delta_1^2) + \tau_{x_1}\,\mathrm{E}(\delta_1)\mathrm{E}(\delta_4) + \mathrm{E}(\delta_1^2)\mathrm{E}(\delta_4)$$

since $\mathrm{E}(\delta_1) = 0$, ξ_i and δ_i are mutually uncorrelated, and δ_i and δ_j are mutually uncorrelated.

$$= \tau_{x_4}\,\mathrm{E}(\delta_1^2) \quad \text{since} \quad E(\delta_i) = 0, \forall i = 1, \cdots, 9 \quad \text{and} \quad E(\xi_i) = 0.$$

$$= \tau_{x_4}\theta_{\delta 1}$$

Appendix B: Derivation of γ_3 given $R^2_{\gamma_3}$ and $\text{var}(\zeta)$

$$\eta = \gamma_0 + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$$

$$\text{var}(\xi_1\xi_2) = (\phi_{11} + \mu_1^2)(\phi_{22} + \mu_2^2) + (\phi_{21} + \mu_1\mu_2)^2 - 2\mu_1^2\mu_2^2$$

$$\text{cov}(\xi_1, \xi_1\xi_2) = \mu_1\phi_{12} + \phi_{11}\mu_2$$

$$\text{cov}(f_2, f_1f_2) = \mu_2\phi_{12} + \phi_{22}\mu_1$$

$$\begin{aligned}\text{var}(\eta_3) &= \gamma_1^2\phi_{11} + \gamma_2^2\phi_{22} + \gamma_3^2\,\text{var}(\xi_1\xi_2) + 2\gamma_1\gamma_2\phi_{12} \\ &\quad + 2\gamma_1\gamma_3\,\text{cov}(\xi_1, \xi_1\xi_2) + 2\gamma_2\gamma_3\,\text{cov}(\xi_2, \xi_1\xi_2) + \text{var}(\zeta)\end{aligned}$$

$$\begin{aligned}&= \gamma_1^2\phi_{11} + \gamma_2^2\phi_{22} + \gamma_3^2\,\text{var}(\xi_1\xi_2) + 2\gamma_1\gamma_2\phi_{12} \\ &\quad + 2\gamma_1\gamma_3(\mu_1\phi_{12} + \phi_{11}\mu_2) + 2\gamma_2\gamma_3(\mu_2\phi_{12} + \phi_{22}\mu_1) + \text{var}(\zeta)\end{aligned}$$

$$\begin{aligned}&= [\gamma_1^2\phi_{11} + \gamma_2^2\phi_{22} + 2\gamma_1\gamma_2\phi_{12}] \\ &\quad + [\gamma_3^2\,\text{var}(\xi_1\xi_2) + 2\gamma_1\gamma_3(\mu_1\phi_{12} + \phi_{11}\mu_2) + 2\gamma_2\gamma_3(\mu_2\phi_{21} + \phi_{22}\mu_1)] + \text{var}(\xi)\end{aligned}$$

$$=\text{var(linear)+var(int)+var(d)}$$

$$=\text{var(reg)+var(d)}$$

$$R^2_{\gamma_3} = \frac{\text{var(int)}}{\text{var}(\eta)}$$

$$\Rightarrow \text{var}(\eta)\cdot R^2_{\gamma_3} = \text{var(int)}$$

$$\Rightarrow \text{var}(\eta)\cdot R^2_{\gamma_3} - \text{var(int)} = 0$$

$$\Rightarrow [\text{var}(\textit{linear}) + \text{var(int)} + \text{var}(\zeta)]\cdot R^2_{\gamma_3} - \text{var(int)} = 0$$

$$\Rightarrow [\text{var}(\textit{linear}) + \text{var(int)} + \text{var}(\zeta)] - \frac{\text{var(int)}}{R^2_{\gamma_3}} = 0$$

$$\Rightarrow \text{var(int)}\left[1 - \frac{1}{R^2_{\gamma_3}}\right] + \left(\text{var}(\textit{linear}) + \text{var}(\xi)\right) = 0$$

$$\Rightarrow \left[\gamma_3^2\,\text{var}(\xi_1\xi_2) + 2\gamma_1\gamma_3\,\text{cov}(\xi_1, \xi_1\xi_2) + 2\gamma_2\gamma_3\,\text{cov}(\xi_2, \xi_1\xi_2)\right]\cdot \frac{R^2_{\gamma_3} - 1}{R^2_{\gamma_3}}$$

$$+ \left(\text{var}(\textit{linear}) + \text{var}(\zeta)\right) = 0$$

$$\Rightarrow \gamma_3^2 \operatorname{var}(\xi_1\xi_2) + \gamma_3[2\gamma_1 \operatorname{cov}(\xi_1, \xi_1\xi_2) + 2\gamma_2 \operatorname{cov}(\xi_2, \xi_1\xi_2)]$$

$$+ \left[\operatorname{var}(linear) + \operatorname{var}(\zeta)\right] \cdot \frac{R_{\gamma_3}^2}{R_{\gamma_3}^2 - 1} = 0$$

$$\Rightarrow a\gamma_3^2 + b\gamma_3 + c = 0$$

where $a = \operatorname{var}(\xi_1\xi_2)$,

$b = [2\gamma_1 \operatorname{cov}(\xi_1, \xi_1\xi_2) + 2\gamma_2 \operatorname{cov}(\xi_2, \xi_1\xi_2)]$,

and $c = \left[\operatorname{var}(linear) + \operatorname{var}(\zeta)\right] \cdot \frac{R_{\gamma_3}^2}{R_{\gamma_3}^2 - 1}$.

Therefore, $\gamma_3 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$,

where $a = \operatorname{var}(\xi_1\xi_2)$,

$b = [2\gamma_1 \operatorname{cov}(\xi_1, \xi_1\xi_2) + 2\gamma_2 \operatorname{cov}(\xi_2, \xi_1\xi_2)]$,

and $c = \left[\operatorname{var}(linear) + \operatorname{var}(\zeta)\right] \cdot \frac{R_{\gamma_3}^2}{R_{\gamma_3}^2 - 1}$.

Appendix C: $\boldsymbol{\mu}^{(2)}$ and $\gamma_3^{(2)}$ for different conditions.

MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12} = 0.2$		$\phi_{12} = 0.6$	
		$\boldsymbol{\mu}^{(2)}$	$\gamma_3^{(2)}$	$\boldsymbol{\mu}^{(2)}$	$\gamma_3^{(2)}$
0	0.05	$\begin{pmatrix}0\\0\end{pmatrix}$	1.134	$\begin{pmatrix}0\\0\end{pmatrix}$	1.081
0	0.15	$\begin{pmatrix}0\\0\end{pmatrix}$	2.076	$\begin{pmatrix}0\\0\end{pmatrix}$	1.980
0	0.25	$\begin{pmatrix}0\\0\end{pmatrix}$	2.853	$\begin{pmatrix}0\\0\end{pmatrix}$	2.720
1.5	0.05	$\begin{pmatrix}0.85\\0.85\end{pmatrix}$	0.127	$\begin{pmatrix}0.693\\0.693\end{pmatrix}$	0.138
1.5	0.15	$\begin{pmatrix}0.85\\0.85\end{pmatrix}$	0.397	$\begin{pmatrix}0.693\\0.693\end{pmatrix}$	0.432
1.5	0.25	$\begin{pmatrix}0.85\\0.85\end{pmatrix}$	0.698	$\begin{pmatrix}0.693\\0.693\end{pmatrix}$	0.758
3	0.05	$\begin{pmatrix}1.2\\1.2\end{pmatrix}$	0.090	$\begin{pmatrix}0.980\\0.980\end{pmatrix}$	0.099
3	0.15	$\begin{pmatrix}1.2\\1.2\end{pmatrix}$	0.286	$\begin{pmatrix}0.980\\0.980\end{pmatrix}$	0.312
3	0.25	$\begin{pmatrix}1.2\\1.2\end{pmatrix}$	0.508	$\begin{pmatrix}0.980\\0.980\end{pmatrix}$	0.554

Appendix D: Class-specific means and standard deviations
of the endogenous latent variables for different conditions

MD	$R^2_{\gamma_3^{(2)}}$	$\phi_{12} = 0.2$				$\phi_{12} = 0.6$			
		$E(\eta^{(1)})$	$E(\eta^{(2)})$	$SD(\eta^{(1)})$	$SD(\eta^{(2)})$	$E(\eta^{(1)})$	$E(\eta^{(2)})$	$SD(\eta^{(1)})$	$SD(\eta^{(2)})$
0	0.05	10	10.227	3.924	4.091	10	10.680	4.494	4.685
0	0.15	10	10.415	3.924	4.459	10	11.246	4.494	5.105
0	0.25	10	10.571	3.924	4.885	10	11.712	4.494	5.592
1.5	0.05	10	14.367	3.924	4.091	10	14.418	4.494	4.689
1.5	0.15	10	14.616	3.924	4.459	10	14.775	4.494	5.116
1.5	0.25	10	14.894	3.924	4.886	10	15.173	4.494	5.611
3	0.05	10	16.148	3.924	4.091	10	16.184	4.494	4.688
3	0.15	10	16.469	3.924	4.459	10	16.853	4.494	5.117
3	0.25	10	16.833	3.924	4.886	10	17.036	4.494	5.612

References

Aiken, L. S., & West, S. G. (1996). *Multiple regression: Testing and interpreting interactions*. London: Sage.

Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society Series A (General), 144*, 419-461.

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

Algina, J., & Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using LISREL 8.3. *Structural Equation Modeling, 8*, 40-52.

Arminger, G., & Stein, P. (1997). Finite mixtures covariance structure models with regressors. *Sociological Methods & Research*, *26*, 148-182.

Arminger, G., & Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika, 64*, 475-494.

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hasting algorithm. *Psychometrika, 63*, 271-300.

Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age Publishing.

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

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

Methods, 9, 3-29.

Bohrnstedt, G. W., & Golderger, A. S. (1967). On the exact covariance products of random variables, *Journal of the American Statistical Association*, *64*, 1439-1442.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression /correlation analysis for the behavioral sciences. Mahwah, NJ: Erlbaum.

Cudeck, R., Harring, J. R., & du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interaction. *Journal of Educational and Behavioral Statistics*, *34*, 131-144.

De Boer, P., Droese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research, 134*, 19-67.

Dolan, C. V., & Van der Mass, H. L. J. (1997). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227-253.

Hancock, G. R., & Muller, R. O. (2006). Introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age Publishing.

Harring, J. R. (2010). *A note on centering in nonlinear latent variable structural models*. Unpublished manuscripts.

Harring, J. R. (in review). Finite mixtures of nonlinear mixed effects models. In G. R. Hancock & J. R. Harring (Eds.), *Advances in Longitudinal Methods in Education and the Social and Behavioral Sciences*. Greenwich, CT: Information Age Publishing.

Haughton, D. (1988). On the choice of a model to fit data from an exponential family. *The Annuals of Statistics*, *16*, 342-355.

Hayduck, L. A. (1987). Structural equation modeling with LISREL. Baltimore, MD: Johns Hopkins University Press.

Heath, J. W., Fu, M., & Jank, W. (2007). Global optimization of finite mixture models. Unpublished doctoral dissertation, University of Maryland, College Park.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research*, *26*, 329-367.

Hu, J., Fu, M. C., & Marcus, S. I. (2007). A model reference adaptive search algorithm for global optimization. *Operations Research, 55*, 549-568.

Jaccard, J., & Wan, D. k. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicators and structural equation approaches. *Psychological Bulletin, 117*, 348-357.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). STEMM: A general finite mixture structural equation model, *Journal of Classification*, *14*, 23-50.

Jöreskog, K. G., & Sörbom, D. (1996). LISREL 8: *Structural Equation Modeling with the SIMPLIS command Language*. Scientific Software International.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57-88). Mahwah, NJ: Erlbaum.

Jung, T., & Wickrama, K. A. S. (2007). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 1*, 302-317.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects

of latent variables. *Psychological Bulletin, 96*, 201-201.

Kelava, A., Moosbrugger, H., Dimitruk, P., & Schermelleh-Engel, K. (2008). Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. *Methodology, 4*, 51-66.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent Interaction models. *Psychometrika*, 65, 457-474.

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42*, 647-673.

Lee, S. Y., & Zhu, H. T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*, 189-210.

Lee, S. Y., Song, X. Y., & Poon, W. Y. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research*, *39*, 37-67.

Lee, S. Y., Song, X. Y., & Tang, N. S. (2007). Bayesian methods for analyzing structural equation models with covariate, interaction, and quadratic latent variables. *Structural Equation Modeling, 14*, 404-434.

Little, T. D., James, A. B., & Widaman, K. F. (2006). On the merits of orthogonalizing Powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling, 13,* 497-519.

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

Lubke, G. H., & Muthen, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21-39.

Lubke, G. H., & Muthen, B. (2007). Performance of factor mixture models as a

function of model size, covariate effects, and class-specific parameters, *Structural Equation Modeling, 14,* 26-47.

Lubke, G. H., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, *41*, 499-532.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*, 275-300.

Marsh, H. W., Wen, Z., & Hau, K. T. (2006). Structural equation models of latent Interactions and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 225-265). Greenwich, CT: Information Age Publishing.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36*, 318-324.

McLachlan, G. J., & Basford, K. E. (1998). *Mixture models: Inference and applications to clustering*. New York: Marcel Dekker.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions*. New York: Wiley.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Melnykov, V., & Maitra, R. (2010). Finite mixture models and model-based clustering. *Statistics Surveys, 4*, 80-116.

Mooney, C. Z. (1997). *Monte carlo simulation*. Thousand Oaks, CA: Sage.

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimation and testing latent variable interactions. *Structural Equation Modeling, 9*, 1-19.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations.

Psychometrika, *54*, 557-585.

Muthén, B. O. (1998-2004). Mplus Technical Appendices. Los Angeles, Muthén and Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide*. Los Angeles, Muthén and Muthén.

Muthén, L. K., & Muthén, B. O. (2010). Mplus version 6.1 [Computer software]. Los Angeles, Muthén and Muthén.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C. et al (2002). General growth mixture modeling in randomized preventative interventions, *Biostatistics*, *3*, 459-475.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models*. Homewood, IL: Irwin.

Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew Variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A, 186*, 343-414.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming, 12,* 241-254.

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

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

SAS Institute, Inc. (2008). *SAS Version 9.2*. Cary, NC: SAS Institute, Inc.

Schumacker, R. E. (2002). Latent variable interaction modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 40-54.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annuals of Statistics*, *6*, 461-464.

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

Song, X. Y., & Lee, S. Y. (2006). Bayesian analysis of structural equation models with nonlinear covariates and latent variables. *Multivariate Behavioral Research, 41*, 337-365.

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

Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modelling, 18*, 110-131.

Vermunt, J. K., & Magidson, J. (2005). Structural equation models: Mixture models. In B. Everitt & D. Howell, (Eds.), *Encyclopedia of statistics in behavioral science*, (pp. 1922-1927). Chichester, UK: John Wiley & Sons.

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*, 1-29.

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*, 47-63.

Wall, M. M. (2009). Maximum likelihood and Bayesian estimation for nonlinear structural equation models. In R. E. Millsap and A. Maydeu-Olivares (Eds.), *The*

Sage handbook of quantitative methods in psychology (pp. 540-547). London: Sage.

Weiss, B. A. (2010). A comparison of methods for testing for interaction effects in structural equation modeling. Unpublished doctoral dissertation, University of Maryland, College Park.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annuals of Statistics, 11*, 95-103.

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype Identification. *Computational Statistics & Data Analysis*, *50*, 1090-1104.

Yung, Y. F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika*, *62*, 297-330.